



Christian Dustmann
Thomas Lemieux

Handbook of Labor Economics

VOLUME 5

NORTH-HOLLAND

Handbook of Labor Economics

Volume 5

This book belongs to Lorenzo Lagos
(lorenzolagos@gmail.com)

INTRODUCTION TO THE SERIES

The aim of the *Handbooks in Economics* series is to produce Handbooks for various branches of economics, each of which is a definitive source, reference, and teaching supplement for use by professional researchers and advanced graduate students. Each Handbook provides self-contained surveys of the current state of a branch of economics in the form of chapters prepared by leading specialists on various aspects of this branch of economics. These surveys summarize not only received results but also newer developments, from recent journal articles and discussion papers. Some original material is also included, but the main goal is to provide comprehensive and accessible surveys.

The Handbooks are intended to provide not only useful reference volumes for professional collections but also possible supplementary readings for advanced courses for graduate students in economics.

KENNETH J. ARROW and MICHAEL D. INTRILIGATOR

Handbook of Labor Economics

Volume 5

Editors

Christian Dustmann

UCL
London
United Kingdom

Thomas Lemieux

Vancouver School of Economics,
University of British Columbia,
Vancouver,
British Columbia,
Canada



North-Holland

An imprint of Elsevier

North-Holland is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands

First edition 2024

Copyright © 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Publisher's note: Elsevier takes a neutral position with respect to territorial disputes or jurisdictional claims in its published content, including in maps and institutional affiliations.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-443-29764-9

ISSN: 3050-7561

For information on all North-Holland publications
visit our website at <https://www.elsevier.com/books-and-journals>

Publisher: Zoe Kruze

Acquisitions Editor: Jason Mitchell

Editorial Project Manager: Sneha Apar

Production Project Manager: Abdulla Sait

Cover Designer: Arumugam Kothandan

Typeset by MPS Limited, India



Working together
to grow libraries in
developing countries

www.elsevier.com • www.bookaid.org

Contents

List of Contributors	xiii
Preface	xv
1. Instrumental variables with unobserved heterogeneity in treatment effects	1
<i>Magne Mogstad and Alexander Torgovitsky</i>	
1 Introduction	2
2 Background	4
2.1 IV in a nutshell	4
2.2 Why is there unobserved heterogeneity in treatment effects?	4
2.3 From the classical linear IV model to potential outcomes	5
2.4 Selection models	7
2.5 Full exogeneity	10
2.6 Target parameters	11
2.7 Testability	14
3 Reverse engineering: interpreting linear estimators	14
3.1 Estimators, estimands, and weak causality	15
3.2 Binary treatment, binary instrument, no covariates	16
3.3 Multivalued instruments	18
3.4 Violations of monotonicity	21
3.5 Multiple instruments	26
3.6 Ordered, cardinal treatments	27
3.7 Unordered or non-cardinal treatments	29
3.8 Covariates	34
3.9 Summary of reverse engineering	39
4 Forward engineering: estimating target parameters	44
4.1 Assuming away the problem	45
4.2 Estimating LATEs and ACRs in the presence of covariates	46
4.3 Marginal treatment effects	53
4.4 Binary treatments when monotonicity is violated	70
4.5 Ordered treatments	72
4.6 Unordered treatments	78

4.7 No selection model	81
4.8 Summary of forward engineering	86
5 Recommendations for practice	87
5.1 Step 1: Assess the likely role of UHTE	87
5.2 Step 2: Reverse engineer with caution	88
5.3 Step 3: Forward engineer estimates of interpretable target parameters	90
6 Conclusion	91
Appendix A Potential outcomes or latent variables? It's just notation ...	92
Appendix B Definition of a weakly causal estimand	92
Appendix C Deriving the average causal response and an alternative decomposition	94
Appendix D Estimating the average causal response with covariates	96
Appendix E Derivations for marginal treatment effects	96
E.1 Derivations of weighting expressions	96
E.2 The normal selection model	98
E.3 Saturated MTR specifications reproduce the LATE	98
References	99
2. Firm wage effects	115
<i>Patrick Kline</i>	
1 Background	119
2 What sorts of firms pay high wages?	120
2.1 Productivity, worker flows, and firm size	121
2.2 Entry, reallocation, and dynamics	122
2.3 Sorting, outsourcing, and displacement	123
2.4 Industry structure and amenities	124
3 The AKM model	125
3.1 An edgy interpretation of firm effects	127
3.2 Evaluating the AKM restrictions	135
3.3 Causality	140
4 Variance decomposition	146
4.1 Limited mobility bias	147
4.2 Cross-fitting and bias correction	148
4.3 Clustering approaches	156
4.4 How variable are worker and firm effects?	159
5 Regressing firm effects on observables	161
5.1 One step vs two	162
5.2 Variance estimation	163
5.3 Revisiting the firm size wage premium	164
6 Hiring origins and state dependence	166
6.1 Structural interpretation	167
6.2 Testable restrictions	168

6.3 It ain't where you're from, it's where you're at	169
6.4 Information and conduct	170
7 Conclusion	171
Appendix: Covariance between person and firm effects	172
References	176
3. Empirical Bayes methods in labor economics	183
<i>Christopher Walters</i>	
1 Introduction	184
2 Empirical Bayes basics	186
2.1 An empirical Bayes recipe	186
2.2 Gains from shrinkage	192
2.3 Practical shrinkage issues	196
2.4 Generalizations of linear shrinkage	200
2.5 EB decision rules	204
2.6 Precision-dependence	206
2.7 Connections to machine learning	211
2.8 Linear shrinkage application: school value-added in Boston	213
3 Non-parametric empirical Bayes	216
3.1 Bias-corrected variance estimation	216
3.2 Non-parametric priors and posteriors	221
3.3 Partial identification	224
3.4 EB for multiple testing: large-scale inference	227
3.5 Ranking problems	230
3.6 Compound decisions and shrinkage strategies	233
3.7 Non-parametric EB application: firm-level labor market discrimination	235
4 Conclusion	253
References	255
4. Minimum wages in the 21st century	261
<i>Arindrajit Dube, and Attila Lindner</i>	
1 Introduction and overview	261
2 A brief history of minimum wages	264
2.1 The rationale for minimum wage policies	264
2.2 The minimum wage debate	268
3 The wage and employment effects of minimum wages	271
3.1 Wages, employment and labor demand	271
3.2 Empirical methods to study the impact of minimum wage policies	274
3.3 Review of the evidence on employment effects	302
3.4 Effect on total hours	313
4 Margins of adjustment	314
4.1 Review the evidence on various margins of adjustment	314

4.2 Summary of evidence on margins of adjustment	335
4.3 Modeling implications and open questions	337
5 Inequality, distributional implications, and downstream effects	341
6 Minimum wages in developing countries	348
7 Conclusion and future directions	350
Appendix A Additional results	352
Appendix B Bias from heterogeneous pre-existing trends: a simulation study	361
Appendix C Data sources for cross-country Kaitz indices	364
Appendix D Constructing historical QCEW restaurant data	366
Appendix E Construction of 60 state-level minimum wage events	368
Appendix F Construction of probability groups using demographic predictors	370
References	370
Further reading	383
5. The micro and macro economics of short-time work	385
<i>Pierre Cahuc</i>	
1 Introduction	386
2 Overview of STW schemes	388
2.1 The spread of STW since the 1920s	388
2.2 STW and other workforce retention measures	389
2.3 The design of STW schemes	389
3 STW take-up	393
3.1 Take-up by type of workers and firms	393
3.2 Take-up in large recessions	401
3.3 Take-up outside of large recessions	401
3.4 Take-up and labor market regulation	403
3.5 STW design and administrative capacity	404
4 The theoretical models of STW	406
4.1 Normative approach	406
4.2 Positive approach	407
5 The efficiency of STW	412
5.1 The social willingness to pay for STW	414
5.2 The Net public cost of STW	416
6 Effects of STW at the macroeconomic level	419
6.1 STW before the Great recession of 2008-2009	419
6.2 STW in the Great recession of 2008-2009	420
6.3 STW in the Covid-19 crisis	421
6.4 The role of the timing of STW regulation and eligibility criteria	422
7 Effects on firms	423
7.1 Employment and hours of work	423
7.2 Firm productivity, profitability and firm survival	426
7.3 Job reallocation and productivity	427

8 Effects on workers trajectories	428
9 Conclusion	429
References	430
6. Job search, unemployment insurance, and active labor market policies	435
<i>Thomas Le Barbanchon, Johannes Schmieder, and Andrea Weber</i>	
1 Introduction and background on UI	436
1.1 The origin of unemployment insurance and active labor market policy	437
1.2 Unemployment insurance today	439
2 Micro-foundations of job search among the unemployed	441
2.1 A brief history of job search theory	441
2.2 The basic job search model	442
2.3 Evidence from job finding rates and re-employment wages	451
2.4 New empirical moments	459
2.5 Refining the search model	473
2.6 Discussion	496
3 Design of UI policy	496
3.1 The structure of unemployment insurance policies	497
3.2 The welfare effects of unemployment insurance	497
3.3 Quantification of behavioral costs	503
3.4 Quantification of the social value of UI changes	507
3.5 The marginal value of public funds	518
3.6 UI effects beyond unemployment duration	523
3.7 Other UI design questions	530
3.8 Micro and macro effects of UI programs	534
3.9 Discussion	539
4 Active labor market policies	541
4.1 Meta analysis studies	542
4.2 Lesson 1: New insights in the role of caseworkers	544
4.3 Lesson 2: More focus on programs for special groups	546
4.4 Lesson 3: Program design takes demand side into account	550
4.5 Lesson 4: Internationalization of ALMP use and evaluations	552
4.6 Lesson 5: Advances in labor market design on online search platforms	555
4.7 Lesson 6: Growing awareness of spillover or displacement effects	558
4.8 Lesson 7: Discussion of cost effectiveness	560
4.9 Lesson 8: Wide range of outcome variables	563
4.10 Lesson 9: What are the mechanisms explaining program effects	568
4.11 Lesson 10: Novel identification strategies	568

Acknowledgments	570
Appendix A Supporting information	570
References	570
7. Families, public policies, and the labor market	581
<i>Gordon Dahl and Katrine V. Loken</i>	
1 Introduction	581
2 Conceptual framework	582
3 Public policies in OECD countries	585
4 Fertility	586
4.1 Rates and trends	586
4.2 Public policies affecting fertility	587
5 Marriage, divorce, and cohabitation	591
5.1 Rates and trends	591
5.2 Public policies affecting marriage, divorce, and cohabitation	594
6 Family labor supply	596
6.1 Rates and trends	596
6.2 Public policies and family labor supply	598
7 Gender inequality	604
7.1 Rates and trends	604
7.2 Public policies and gender inequality	605
8 Child outcomes	606
8.1 Public policies and child outcomes	606
9 Norms and spillovers	609
9.1 Public policies and norms and spillovers	609
10 Lessons learned and avenues for future research	610
10.1 Short summary	610
10.2 Avenues for future research	611
References	611
8. The evolution of gender in the labor market	619
<i>Claudia Olivetti, Jessica Pan, and Barbara Petrongolo</i>	
1 Introduction	619
2 Real world and academic developments in gender inequality	624
3 Women's labor supply and the gender gap	630
3.1 The labor supply of the secondary earner	631
4 Evolving perspectives on gender inequality	635
4.1 Preferences, traits, and constraints	636
4.2 Career-family trade-offs	639
5 The anatomy of the career costs of motherhood	644
5.1 Gender biology and productivity	644
5.2 Differential job sorting and the organization of work	646
5.3 Monopsonistic labor markets	649

6 The role of identity and norms in understanding gender inequalities	652
6.1 Relevance for labor supply and household specialization	653
6.2 Stereotypes, beliefs, and discrimination	655
6.3 What drives gender norms and how malleable are they?	658
7 Micro-macro linkages	664
8 Conclusion	665
Appendix A: Gender in economic journals	667
References	669
 9. Crime and the labor market	 679
<i>Randi Hjalmarsson, Stephen Machin, and Paolo Pinotti</i>	
1 Introduction	680
2 Descriptive statistics and stylized facts: an international perspective	683
3 Labor market impacts on crime	687
3.1 Wages and income	687
3.2 Unemployment	689
3.3 Youth labor markets: summer youth employment programs	694
3.4 Returns to crime: earnings and prices	696
4 Criminal record impacts on the labor market	698
4.1 Effects of a record on labor market outcomes	700
4.2 Firm willingness to hire workers with criminal records	703
4.3 Policies to improve labor outcomes for workers with criminal records	703
5 Education and crime	705
5.1 Causal impacts of education on crime	740
5.2 Incapacitation	741
5.3 Schooling quantity and quality	742
5.4 Crime impacts on education	743
5.5 Crime and education policies	744
6 Future directions	744
6.1 Future direction 1: Victimization	744
6.2 Future direction 2: Gangs and organized crime	748
7 Conclusions	750
References	751
 10. Monopsony power in the labor market	 761
<i>José Azar and Ioana Marinescu</i>	
1 Introduction	761
2 New quantitative models of monopsony power	764
2.1 What is monopsony power?	764
2.2 Oligopsony	767

2.3 Differentiated jobs	769
2.4 Search frictions	781
2.5 Discussion	791
3 Empirically measuring monopsony power	792
3.1 Definition of the markdown in the empirical literature	792
3.2 Elasticity of labor supply	793
3.3 Labor market concentration	799
3.4 Reduced-form approach based on workers' outside options	804
3.5 Calibration and simulation	804
3.6 Structural estimation	805
3.7 Production function approach	806
3.8 Summary and discussion of markdown estimates	810
4 Policy and monopsony power	812
4.1 Merger control	812
4.2 Non-competition agreements	815
4.3 Minimum wage	818
5 Conclusion	820
References	822
Index	829

List of Contributor

- José Azar** Centre for Economic Policy Research, London, United Kingdom
- Gordon Dahl** University of California, San Diego
- Arindrajit Dube** Umass Amherst, NBER, MA, United States
- Randi Hjalmarsson** Department of Economics, University of Gothenburg, Sweden
- Patrick Kline** University of California, Berkeley, USA
- Attila Lindner** University College London, London, United Kingdom
- Katrine V. Loken** Norwegian School of Economics
- Stephen Machin** Department of Economics and Centre for Economic Performance,
London School of Economics, London, United Kingdom
- Ioana Marinescu** University of Pennsylvania, PA, United States
- Claudia Olivetti** Dartmouth College, NH; National Bureau of Economic Research,
MA, United States
- Jessica Pan** National University of Singapore, Singapore; University of Bonn, IZA,
Bonn, Germany
- Barbara Petrongolo** Oxford University, Oxford, United Kingdom; The Centre for
Economic Performance, London School of Economics and Political Science, London,
United States; Centre for Economic Policy Research, London, United Kingdom
- Paolo Pinotti** Department of Social and Political Sciences and CLEAN Research Unit
on the Economic Analysis of Crime, Bocconi University, Milan, Italy
- Christopher Walters** UC Berkeley, NBER, and Amazon

This page intentionally left blank

Preface

It has been 14 years since the last volume of the *Handbook of Labor Economics* was published. In that time, labor economics has witnessed remarkable research progress. New ideas have emerged across various areas of the field, and the availability of powerful administrative data sources has pushed the boundaries of what is possible, leading to incredible advancements.

This new edition of the *Handbook* aims to collect and present the most important developments in labor economics over the past two decades. It is split into two volumes, with chapters that are designed both for graduate-level teaching and as a concise summary of the latest developments for scholars working in the respective areas. This first volume of the new *Handbook* (*Handbook of Labor Economics*, vol. 5) includes three methodologically oriented chapters, four chapters focused on labor market policies, and three chapters covering other areas of central interest in labor economics.

The methods chapters cover modern econometrics methods that are becoming essential parts of the empirical toolkit of labor economists. Instrumental variable (IV) methods remain an essential tool for causal inference using observational data in labor economics. The most important advances in the last 15 years have revolved around the theme of heterogeneity in treatment effects. The chapter “Instrumental Variables with Unobserved Heterogeneity in Treatment Effects” by Magne Mogstad and Alexander Torgovitsky synthesizes and critically reviews these modern IV methods.

New empirical methods in labor economics have also been developed to fully exploit the potential of rich administrative data that are now widely used by labor economists. The chapter “Firm Wage Effects” by Patrick Kline reviews recent advances in the literature on firm wage differences and the fixed effects methods—pioneered by Abowd, Kramarz, and Margolis (1999)—typically used to measure these differences. The firm wage effects extracted from these large administrative data sets are unit-specific parameters that may represent causal effects under reasonable assumptions discussed in the chapter.

When working with a large number of unit-specific estimates, Empirical Bayes (EB) methods help refine these estimates, leading to improved estimators and decision rules. The chapter “Empirical Bayes Methods in Labor Economics” by Christopher Walters illustrates the potential of these methods in the context of value-added studies measuring causal effects of individual units like firms, managers, neighborhoods, teachers, schools, doctors, hospitals, police officers, and judges.

The four policy chapters paint a broad picture of how research over the last 15 years has led to major advances in our understanding of the impact of a vast

array of labor market policies used across the world. Few policies have captured the imagination of labor economists as much as minimum wage policies. The chapter “Minimum Wages in the 21st Century” by Arindrajit Dube and Attila Lindner surveys the evolving literature on the impact of minimum wages on low-wage labor markets and provide a comprehensive assessment of how minimum wage affects firms, workers, and labor markets.

The last 15 years were marked by two major economic downturns associated with the Global Financial Crisis of 2007-08 and the COVID-19 pandemic. These events spurred major research efforts on labor market policies aimed at addressing unemployment and persistent labor market challenges among disadvantaged segments of the labor market. The chapter “The Micro and Macro Economics of Short-Time Work” by Pierre Cahuc provides an overview of the economic literature on short-time work, focusing on its effectiveness as a job preservation mechanism, drawing on theoretical models and empirical studies.

In the chapter “Job Search, Unemployment Insurance, and Active Labor Market Policies,” Thomas Le Barbanchon, Johannes Schmieder, and Andrea Weber provide a comprehensive overview of labor economists’ perspectives on job search among the unemployed. It also explores how job search is influenced by unemployment insurance and active labor market policies.

Another key development in the past 15 years is the remarkable evolution of family dynamics and labor market interactions, encompassing changes in fertility, marriage, divorce, cohabitation, family labor supply, gender inequality, and childrearing. In the chapter “Families, Public Policies, and the Labor Market,” Gordon Dahl and Katrine Loken analyze how government policies have influenced these trends, showcasing the latest research from various OECD countries.

The three other chapters in this new Handbook discuss some of the most fundamental developments in the field over the last 10-20 years. In “The Evolution of Gender in the Labor Market,” Claudia Olivetti, Jessica Pan, and Barbara Petrongolo trace the evolution of the study of gender, focusing on how academic thinking on this topic has evolved and how past insights inform current perspectives on addressing the remaining gender disparities in the labor market. It uses theory and empirics to address how forces like culture, technology, and institutions affect the convergence or persistence of gender gaps.

The chapter on “The Economics of Crime” by Randi Hjalmarsson, Stephen Machin, and Paolo Pinotti explains how research in the field has evolved in recent decades, with economists increasingly exploring the causes and consequences of criminal behavior. This chapter surveys key contributions and developments from labor economists, who investigate the connection between crime and labor market factors such as education, wages, and unemployment.

Another major development is that labor economics has steadily moved away from the perfectly competitive markets paradigm to explore how imperfect competition can better account for key empirical patterns and shed a

different light on the effectiveness and appropriateness of different policy interventions. The chapter “Monopsony Power in the Labor Market” by José Azar and Ioana Marinescu discusses the implication of different models on the wage markdown that summarizes how employers with market power can pay wages below the marginal product of labor. The three classes of models considered are oligopsony models, job differentiation models, and search and matching models.

We are deeply grateful to all the authors for their (quasi)-timely delivery of high-quality chapters. This has been a significant undertaking, and we appreciate their commitment and dedication to this project.

We would also like to express our gratitude to the ROCKWOOL Foundation Berlin (RFBerlin) for generously funding the Handbook Conference, held in October 2023. At this event, the authors presented their chapters and received valuable feedback, which significantly contributed to the success of this endeavor.

It is our hope that this latest edition of the *Handbook of Labor Economics* will serve as a valuable resource for both PhD students and senior scholars alike.

Christian Dustmann
Thomas Lemieux

This page intentionally left blank

Chapter 1

Instrumental variables with unobserved heterogeneity in treatment effects[☆]

Magne Mogstad^{a,b,c,*} and Alexander Torgovitsky^{a,†}

^aKenneth C. Griffin Department of Economics, University of Chicago, ^bStatistics Norway, ^cNBER

*Corresponding author. e-mail address: magne.mogstad@gmail.com

Chapter Outline

1 Introduction	2	3.8 Covariates	34
2 Background	4	3.9 Summary of reverse engineering	39
2.1 IV in a nutshell	4		
2.2 Why is there unobserved heterogeneity in treatment effects?	4	4 Forward engineering: estimating target parameters	44
2.3 From the classical linear IV model to potential outcomes	5	4.1 Assuming away the problem	45
2.4 Selection models	7	4.2 Estimating LATEs and ACRs in the presence of covariates	46
2.5 Full exogeneity	10	4.3 Marginal treatment effects	53
2.6 Target parameters	11	4.4 Binary treatments when monotonicity is violated	70
2.7 Testability	14	4.5 Ordered treatments	72
3 Reverse engineering: interpreting linear estimators	14	4.6 Unordered treatments	78
3.1 Estimators, estimands, and weak causality	15	4.7 No selection model	81
3.2 Binary treatment, binary instrument, no covariates	16	4.8 Summary of forward engineering	86
3.3 Multivalued instruments	18		
3.4 Violations of monotonicity	21	5 Recommendations for practice	87
3.5 Multiple instruments	26	5.1 Step 1: Assess the likely role of UHTE	87
3.6 Ordered, cardinal treatments	27	5.2 Step 2: Reverse engineer with caution	88
3.7 Unordered or non-cardinal treatments	29	5.3 Step 3: Forward engineer estimates of interpretable target parameters	90
		6 Conclusion	91

☆ We thank Deniz Dutz, Koichiro Ito, Pat Kline, Matt Masten, Vitor Possebom, Evan Rose, Henrik Sigstad, Tymon Słoczyński, Winnie van Dijk, Chris Walters, and Thomas Wiemann for helpful discussions and comments. We thank Koichiro Ito and Evan Rose for providing data. Ian Xu provided excellent research assistance.

† Research supported by National Science Foundation grant SES-1846832.

Appendix A Potential outcomes or latent variables?		Appendix E Derivations for marginal treatment effects	96
It's just notation ...	92	E.1 Derivations of weighting expressions	96
Appendix B Definition of a weakly causal estimand	92	E.2 The normal selection model	98
Appendix C Deriving the average causal response and an alternative decomposition	94	E.3 Saturated MTR specifications reproduce the LATE	98
Appendix D Estimating the average causal response with covariates	96	References	99

1 Introduction

Instrumental variable (IV) methods are fundamental to causal inference in economics. They are now also widely used across the social and biological sciences. Their attraction lies in allowing for unobserved confounders, which arise generically in economic applications due to private information, preference heterogeneity, and simultaneity, among other reasons. This chapter synthesizes and critically reviews the literature on modern IV methods that allow for unobserved heterogeneity in treatment effects (UHTE).

In Section 2, we briefly review the basic ideas behind IV methods. We argue that UHTE is a generic feature of many economic applications, especially those in labor economics. The classical linear IV model found in textbooks does not allow for UHTE. We clarify the problems created by this misspecification and we outline the conceptual trade-offs associated with various ways of solving these problems.

The rest of the chapter is then organized into two parts, reflecting the two main approaches to incorporating UHTE into IV models.

The first approach, which was pioneered by [Imbens and Angrist \(1994\)](#), is to interpret linear IV estimators designed for the classical linear IV model through the lens of a nonparametric IV model that allows for UHTE. Based on their results, it has become increasingly common in the empirical literature to describe linear IV estimators, such as two-stage least squares (2SLS), as reflecting local average treatment effects (LATEs). This interpretation is derived from a baseline setup with a binary treatment, a binary instrument, and no covariates, a setup which does not characterize most empirical work in practice.

In Section 3, we provide a comprehensive survey of how the LATE interpretation is affected by moving away from the baseline setup. We find that it is remarkably specific to the baseline setup. Deviating from the baseline setup by having a multivalued treatment, multivalued instrument, or by linearly controlling for covariates complicates, qualifies, or breaks the widespread interpretation that “linear IV is LATE.”

The interpretation problems we point to are orthogonal to the debate over whether LATEs are interesting objects, a debate which has been had many times before. Instead, the problems stem from the now-widespread methodological practice of trying to provide a misspecification-robust interpretation for a commonly-used estimator in the context of a less restrictive model for which it was not designed. We call this practice reverse engineering because it starts with an estimator rather than starting with a model. Reverse engineering arguments are increasingly fashionable in microeconomics, having been applied to selection on observables (Angrist, 1998), difference-in-differences and two-way fixed effects (e.g. Goodman-Bacon, 2021; Sun and Abraham, 2021), settings with multivalued treatments (Goldsmith-Pinkham et al., 2024), and regression discontinuity and kink designs (Lee, 2008; Card et al., 2015; Cattaneo et al., 2016). Our discussion in Section 3 shows that reverse engineering arguments for linear IV estimators are brittle.

The second approach, which has a longer and more diffuse history, is to forward engineer estimators of specific target parameters in models that explicitly allow for UHTE. This includes methods of directly estimating LATEs with estimators other than linear IV. It also includes the classical selection model developed by Gronau (1974) and Heckman (1974, 1976, 1979), and its nonparametric reincarnation in terms of the marginal treatment effect (Heckman and Vytlacil, 1999, 2005). Section 4 is devoted to surveying these forward engineering approaches. We discuss methods for estimating unconditional LATEs that avoid some of the pitfalls of reverse engineering encountered with linear IV. We then discuss a practical linear regression framework for conducting marginal treatment effect (MTE) analysis with binary treatments. In doing so we emphasize the underappreciated point, formalized by Vytlacil (2002), that the separable threshold-crossing model used in both the Gronau-Heckman selection model and modern MTE analysis imposes exactly the same “monotonicity condition” about selection as the model used by Imbens and Angrist (1994), just with different notation. The benefit of the MTE analysis is that it provides a vehicle for both clearly stating the target parameter and for imposing additional assumptions to aid in estimating it. We show how the linear regression framework for MTE extends to ordered and unordered treatments, even as the equivalence with the monotonicity condition is lost. The key theme that emerges is the model of treatment selection and under what assumptions it is identified. We contrast these selection model methods to those that allow for UHTE but do not impose restrictions on how the instrument affects treatment choice.

In Section 5, we distill our discussion into a list of recommendations for researchers using IV methods. Section 6 provides some brief concluding remarks. Example Stata and R code for implementing some of the main methods we discuss in the chapter is available at <https://a-torgovitsky.github.io/ivhandbook/>.

2 Background

In this section we provide some brief background on IV methods with an emphasis on the motivation for incorporating unobserved heterogeneity in treatment effects (UHTE).

2.1 IV in a nutshell

IV methods are used to estimate the causal effect of one variable, the treatment, on another variable, the outcome. The motivating concern is that the treatment is endogenous in the sense that it covaries with other unmeasured factors that are associated with the outcome. The association between the treatment and outcome conflates the effect of the treatment with these unmeasured factors. An IV method instead focuses on the association between the outcome and a third variable, the instrumental variable, or instrument for short. If the instrument is associated with the treatment, but not with the unmeasured factors, and if it has no direct effect on the outcome itself, then the association between the instrument and the outcome should only reflect the causal effect of the treatment on the outcome.

This line of reasoning relies on three assumptions that all IV methods invoke to one extent or another: exclusion, exogeneity, and relevance. Exclusion means that the instrument itself has no direct effect on the outcome. Exogeneity means that it is not associated with any unmeasured factors that are associated with the outcome. Relevance means that the instrument is associated with the treatment.

The exclusion and exogeneity assumptions are often controversial in practice. The purpose of this chapter is not to litigate their merits either in general or in specific applications. The enormous body of published empirical work using IV methods suggests that at least some researchers find these assumptions reasonable in at least some applications. Our focus instead is on how these assumptions can be implemented while also allowing for the possibility of unobserved heterogeneity in treatment effects (UHTE), meaning systematic variation in the effect of the treatment on the outcome that persists even after controlling for other observable variables.

2.2 Why is there unobserved heterogeneity in treatment effects?

UHTE creates many complications in IV methods. These complications can be entirely avoided by assuming that treatment effects are either constant (homogeneous) or, slightly more generally, idiosyncratic in the sense of being unassociated with the treatment variable. So before diving in, we should take a moment to reflect on why such an assumption is often unpalatable in empirical economics.

A workhorse example from labor economics illustrates the issues clearly. Suppose that the treatment variable is college attendance and the outcome variable is a labor market outcome, such as subsequent earnings (e.g. Card, 1999). The classic endogeneity concern is that there are unmeasured factors,

often described loosely as “ability,” that are correlated with both educational attainment and labor market performance (e.g. Becker, 1964; Griliches, 1977). This description is not particularly helpful because it obscures the role of choice; attending college is a choice and individuals make choices purposefully.

A more compelling explanation of the endogeneity problem is that individuals are heterogeneous in their anticipated returns to college due to unobserved private information about their skills, aptitudes, or outside options, and they use this information when making their attendance choices.¹ For example, some individuals have an aptitude for abstract reasoning that translates into strong labor market performance only with a college education. Other individuals have an aptitude for trade skills (welder, plumber, carpenter) that are equally remunerative with or without a college education. Individuals choose whether to attend college at least in part because of its anticipated effect on their future earnings. This explanation results in UHTE that is systematically related to the treatment variable itself: those who choose to attend college tend to be those who would benefit from it.

The essential ingredients of this story are common for causal inference questions involving human actors. Interesting treatment variables are often choices. Interesting outcome variables often reflect substantive consequences for the human beings under consideration. Human beings don’t make choices randomly; they likely consider, at least in part, the effect that their choices may have on the outcome. These choices then become treatment variables that are associated with their effects on the outcome. Unless there’s a compelling domain-specific reason to believe that the effect of the treatment cannot vary for some physical or institutional reason, then there will be UHTE that is systematically associated with the observed treatment choices.

2.3 From the classical linear IV model to potential outcomes

Classical textbook treatments of IV (e.g. Theil, 1971; Wooldridge, 2010) start with an equation like

$$Y_i = \alpha_0 + \alpha_1 D_i + \epsilon_i, \quad (1)$$

where Y_i is the outcome variable, D_i is the treatment variable, and ϵ_i is an unobservable that collects all other unmeasured factors in Y_i . The instrument, Z_i , does not appear in this equation due to the exclusion assumption. The exogeneity assumption is that Z_i is uncorrelated with ϵ_i . The relevance assumption is that Z_i is correlated with D_i . All variables are indexed by a unit of observation, i , which we will think of as an individual for concreteness.

¹ Becker (1967), Willis and Rosen (1979), and Card (2001) develop models with this property. The following is adapted from Willis and Rosen (1979, pp. S10–S11).

[Equation \(1\)](#) says that a one unit increase in D_i —holding all else e_i fixed—causes a change of α_1 in Y_i for everyone (all i). This is restrictive in two ways: (i) it implies that the treatment effect of D_i on Y_i is linear, and (ii) it rules out heterogeneous treatment effects. In much of the literature and much of this chapter, it is assumed that D_i is binary (takes values 0 or 1), in which case (i) is not restrictive. Our focus is on relaxing (ii).

One way to relax (ii) is to allow for treatment effects to vary with some observable covariates, X_i . For example, [\(1\)](#) could be augmented to

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \alpha_3 X_i D_i + \epsilon_i, \quad (2)$$

so that the causal effect of D_i on Y_i is now $\alpha_1 + \alpha_3 X_i$, which can vary with i through $\alpha_3 X_i$. [Equation \(2\)](#) allows for observed heterogeneity in treatment effects, but not for unobserved heterogeneity (UHTE). So this relaxation of (ii) does not address self-selection created by factors such as private information and heterogeneity in skills or preferences. For many economic applications—especially those that appeal to an IV strategy—it is precisely the unobserved heterogeneity that is the concern.

Relaxing (ii) to allow for UHTE instead requires interacting the treatment variable with latent variables. One way to do this is to postulate a relationship like

$$Y_i = f(D_i, \epsilon_i), \quad (3)$$

for some function f . Such a relationship is called “nonseparable” because D_i and ϵ_i are not additively separable, as in [\(1\)](#). Unlike the classical model [\(1\)](#), a nonseparable relationship allows for unobserved heterogeneity in treatment effects because, $f(d', \epsilon_i) - f(d, \epsilon_i)$ still depends on ϵ_i .

Comparing $f(d, \epsilon_i)$ to $f(d', \epsilon_i)$ involves the mental exercise of considering the value that the outcome would have taken if the treatment variable had been fixed at a potentially counterfactual value d or d' while keeping all other factors e_i the same. This exercise can be represented succinctly in the potential outcomes notation introduced by Neyman (republished as [Splawa-Neyman et al., 1990](#)) for experiments and ported to observational settings by [Rubin \(1974\)](#).² In potential outcomes notation, the function f and unobservable ϵ_i are replaced by a collection of potential outcomes $Y_i(d)$, one for each value that the treatment D_i can conceivably take.³ Each potential outcome $Y_i(d)$ is a random variable that answers the same counterfactual question as $f(d, \epsilon_i)$: what would Y_i have been had D_i been fixed to d , keeping all other factors the same? The outcome actually observed, Y_i , corresponds to the potential outcome of the

² [Heckman and Vytlacil \(2007a\)](#) detail the many other authors across numerous disciplines who have independently invented similar notation. This is perhaps a testament to its intuitive appeal.

³ The exclusion restriction is already implicitly embedded in this form of potential outcomes notation. To be more explicit, one could start by postulating potential outcomes $Y_i(d, z)$, state the exclusion restriction as $Y_i(d, z) = Y_i(d, z')$ for all z and z' , and then define $Y_i(d) \equiv Y_i(d, z)$.

observed treatment state, $Y_i = Y_i(D_i)$, while all other potential outcomes are unobserved for individual i . When D_i is binary, this is often written as

$$Y_i = (1 - D_i) Y_i(0) + D_i Y_i(1), \quad (4)$$

which is the potential outcomes analog of (3).

Some researchers have strong opinions about the merits of working with notation involving latent variables and nonseparable models versus working with potential outcomes notation. Sometimes these opinions seem to border on suggesting that the notation itself has some special powers. As we show in [Appendix A](#), the difference between the two notations is indeed fully notational: every model written in form (3) implies one written in form (4), and conversely. Good notation is essential for clearly communicating arguments and assumptions. But at the end of the day, it is just notation.

In this chapter, we will use both types of notation. Our default is to use potential outcomes, which tends to be simpler for models that make fewer assumptions. As we will see, however, the challenges created by UHTE often demand more assumptions. Latent variable notation turns out to be useful for this purpose. A good example of the relationship between potential outcomes and latent variables arises when thinking about models of treatment selection or selection models.

2.4 Selection models

The reason that UHTE complicates IV methods is that it matters “who” takes treatment. Individuals with different treatment effects will also tend to make different treatment choices. And if the instrument indeed affects treatment choice, then the distribution of treatment effects conditional on the treatment will further vary conditional on the instrument. Modeling the selection process of how the instrument affects treatment provides a way to keep track of this relationship and restrict it through additional assumptions.

Modeling selection requires taking a stance on the dimensions of the treatment D_i and instrument Z_i . Consider the simplest setting in which both D_i and Z_i are binary (0 or 1).

A large body of research pioneered by [Gronau \(1974\)](#), [Lewis \(1974\)](#), [Heckman \(1974, 1976, 1979\)](#), [Willis and Rosen \(1979\)](#) and others, modeled selection with a threshold-crossing model like

$$D_i = \mathbb{1}[V_i \leq \gamma Z_i], \quad (5)$$

where γ is an unknown parameter, and V_i is a continuously distributed latent variable.⁴ Empirical implementations of this model often incorporate additional

⁴ In the early literature, the focus was typically on a one-sided selection problem where D_i indicated whether Y_i was observed for individual i ([Heckman, 1976](#), is an exception). This problem is a simplified counterpart to evaluating the causal effect of a binary treatment, which can be seen as a two-sided selection problem. Models like (5) are now frequently described as Roy models after [Roy \(1951\)](#); see [Heckman and Honoré \(1990\)](#) for a discussion of the rationale.

control variables X_i , necessitating some functional form assumptions, notably on the distribution of V_i , which is often taken to be normally distributed. These parameterizations should be understood as specific practical implementation choices rather than an assumption inherent to (5).

[Imbens and Angrist \(1994\)](#) took an ostensibly different approach to modeling selection by applying the potential outcomes notation to potential *treatments* that vary with the instrument. We denote these $D_i(z)$ for $z = 0$ and $z = 1$, so that

$$D_i = (1 - Z_i)D_i(0) + Z_iD_i(1), \quad (6)$$

in analogy with (4). With a binary treatment there are four configurations of the pair $(D_i(0), D_i(1))$, which can be thought of as individual i 's choice group. [Angrist et al. \(1996\)](#) later described these groups as never-takers, always-takers, compliers, and defiers, for $(D_i(0), D_i(1)) = (0, 0), (1, 1), (0, 1)$, and $(1, 0)$, respectively.

[Imbens and Angrist \(1994\)](#) assumed that either $D_i(1) \geq D_i(0)$ for all i or $D_i(0) \geq D_i(1)$ for all i , a condition they described as monotonicity. An alternative way to state the condition, which ends up being easier to extend and modify, is in terms of probability: $\mathbb{P}[D_i(1) \geq D_i(0)] = 1$ or $\mathbb{P}[D_i(0) \geq D_i(1)] = 1$. There is no substantive difference between these two formulations. The monotonicity condition implies that there are either no defiers or no compliers. It is usually reasonable to simplify this to the assumption of no defiers, since there are few situations in which monotonicity is a compelling assumption but the direction of monotonicity is not known.

How does the [Imbens and Angrist \(1994\)](#) potential choices model with monotonicity differ from the classical threshold-crossing model (5)? It doesn't. A causal interpretation of (5), in which V_i represents “all other factors,” implies that treatment choice would be given by $D_i(0) = \mathbb{1}[V_i \leq 0]$ if $Z_i = 0$, and by $D_i(1) = \mathbb{1}[V_i \leq \gamma]$ if $Z_i = 1$. Since the same latent variable V_i appears in both implied treatment choices, this implies that either $\mathbb{P}[D_i(1) \geq D_i(0)] = 1$, if $\gamma > 0$, or the reverse if $\gamma \leq 0$, which is exactly the monotonicity condition. Conversely, given potential choices, one can always construct a threshold-crossing model of form (5) that implies exactly the same potential choices by constructing V_i according to individual i 's group. Assuming that $\gamma \geq 0$, make every always-taker have $V_i \leq 0$, every complier have V_i in $(0, \gamma]$, and every never-taker have $V_i > \gamma$ (see Fig. 1). The threshold-crossing model (5) is therefore *equivalent* to the [Imbens and Angrist \(1994\)](#) potential choices model with the monotonicity condition.

The argument just outlined is a special case of a more general equivalence theorem due to [Vytlačil \(2002\)](#). The key to the argument is the additive separability of Z_i and V_i in the threshold-crossing model. The equivalence continues to hold if (5) is replaced by $D_i = \mathbb{1}[V_i \leq \nu(Z_i)]$ for some unknown function ν . It breaks down in a more general model in which Z_i and V_i interact,

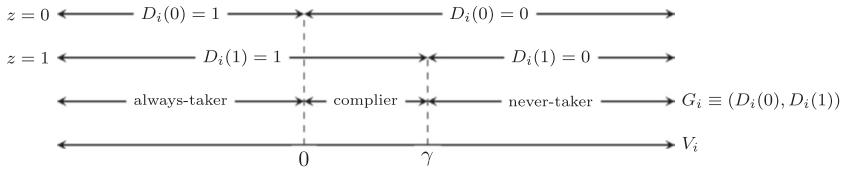


FIG. 1 The Vytlacil (2002) equivalence theorem. Notes: The figure illustrates the case in which monotonicity is in the direction $D_i(1) \geq D_i(0)$ for all i . Individuals with potential no-instrument choice $D_i(0) = 1$ get mapped to $V_i \leq 0$. Because there are no defiers, these individuals are always-takers. Individuals with potential with-instrument choice $D_i(1) = 0$ get mapped to $V_i > \gamma$. Again because there are no defiers, these individuals are never-takers. Compliers get mapped to the remaining region of $0 < V_i \leq \gamma$.

such as $D_i = \mathbb{1}[v(Z_i, V_i) \geq 0]$, which no longer necessarily produces or is produced by potential treatments that satisfy monotonicity.

The implication of Vytlacil's equivalence theorem is that rather than presenting a new model, Imbens and Angrist (1994) were in fact continuing with the same selection model developed in econometrics in the 1970s–1980s, but using a different notation. Their contribution was not so much in the model itself, but in establishing an important nonparametric identification result, the local average treatment effect, which we discuss ahead in Section 3.2. Perhaps inadvertently, they also contributed to clarifying that the additive separability between Z_i and V_i in the threshold-crossing model has a behavioral interpretation as the monotonicity condition.

Vytlacil's equivalence theorem is quite specific to the case of a binary treatment, although within that context it extends naturally to multivalued instruments and covariates. A wider variety of selection models are used for non-binary treatments, many of which we discuss throughout this chapter. These models are more complicated, but they are motivated by a recognition that capturing the relevant treatment variation is important for an IV argument. While it can be tempting to turn a multi-valued treatment into a binary one by collapsing its values together, this can create violations of the exclusion condition.

When modeling with potential treatments, it is often clarifying to think of individuals i as being partitioned into latent groups depending on their potential treatments. Suppose that there are $K + 1$ instrument values z_0, z_1, \dots, z_K with potential treatments $D_i(z_0), D_i(z_1), \dots, D_i(z_K)$. If the treatment can take say four values, then an individual can be in one of 4^{K+1} possible choice groups, which we write as $G_i = (D_i(z_0), D_i(z_1), \dots, D_i(z_K))$.⁵ Assumptions like the monotonicity condition can be viewed as requiring some choice groups to not exist

⁵ Choice groups are an example of what Frangakis and Rubin (2002) call a principal stratification (see also Robins and Greenland (1992)), and what Heckman and Pinto (2018) describe as response vectors or types. Manski (2007) used the same idea for discrete choice analysis.

(have zero probability). So, for example, with $K = 1$, $z_0 = 0$, $z_1 = 1$, and a binary treatment, there are $2^2 = 4$ choice groups—always-takers $G_i = (1, 1)$, never-takers $G_i = (0, 0)$, compliers $G_i = (0, 1)$, and defiers $G_i = (1, 0)$ —and the monotonicity condition is the assumption that $\mathbb{P}[G_i = (1, 0)] = 0$, so there are no defiers. We use this group notation extensively ahead.

2.5 Full exogeneity

Both [Imbens and Angrist \(1994\)](#) and prior work using the threshold-crossing model (5) assumed the instrument to be exogenous with respect to both the outcome and the treatment. In the potential choices notation with a binary treatment and binary instrument the assumption is that Z_i is independent of $(Y_i(0), Y_i(1), D_i(0), D_i(1))$, while the equivalent assumption in latent variable notation is that Z_i is independent of (ϵ_i, V_i) . In contrast, in the classical linear IV model, Z_i is only required to be exogeneous with respect to latent factors determining the outcome: $(Y_i(0), Y_i(1))$ or ϵ_i .⁶ We call these contrasting assumptions outcome and full exogeneity for emphasis. Most approaches to incorporating UHTE impose something like full exogeneity.

Full exogeneity exposes an important distinction between selection models and statistical first stages of the sort that show up in discussions of the two stage least squares (2SLS) estimator. A statistical first stage satisfies

$$D_i = \pi Z_i + \eta_i \quad \text{where } \mathbb{E}[Z_i \eta_i] = 0, \quad (7)$$

an equation often written in tandem with (1) for the classical IV model. [Equation \(7\)](#) can always be satisfied without any assumptions (beyond existence of moments) by taking π to be the population regression coefficient from regressing D_i onto Z_i , so that η_i are the population residuals. This is a statistical relationship, not a model of causality; the population residuals are simply the difference between D_i and its best linear predictor using Z_i . In contrast, the way in which selection models are typically used for IV models with UHTE presupposes a causal interpretation. This requires the unobservables to be viewed as “everything else,” whether stated using a latent variable V_i , or potential choices $D_i(z)$. Assuming full exogeneity implies that the first stage coefficient π represents a causal effect of Z_i on D_i .⁷

The difference between outcome and full exogeneity has important and underappreciated practical implications for IV analysis. Under full exogeneity, different instruments for the same treatment variable cannot be

⁶ Exogeneity in the classical linear IV model is usually stated in terms of orthogonality, correlation or mean independence. Full independence is often necessary for analyzing UHTE. The substantive economic interpretation of the exogeneity of an instrument rarely depends on whether the mathematical formulation is independence or something weaker, like orthogonality.

⁷ For example, if Z_i is binary, and if a constant term is included in (7), then $\pi = \mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]$ as a matter of regression algebra. If full exogeneity is additionally maintained, then $\pi = \mathbb{E}[D_i(1) - D_i(0)]$ is the average causal effect of Z_i on D_i .

considered in isolation. For example, [Card \(1995\)](#) used distance as an instrument for college, while [Kane and Rouse \(1995\)](#) used public college tuition. Suppose that the outcome is future earnings. Evaluating the outcome exogeneity of these instruments means considering whether they are correlated with any other omitted correlates of future earnings. This task is familiar from classical IV analysis with constant treatment effects. Evaluating the *full* exogeneity of these instruments means also considering whether they are correlated with any other omitted correlates of college attendance. In particular, if distance and tuition are correlated, then neither distance nor tuition will satisfy full exogeneity when used in isolation. Full exogeneity requires using both instruments together, or controlling for the omitted instrument as a covariate. For further discussion, see [Heckman \(2010, Section 3.6\)](#) and [Mogstad et al. \(2021, Section III.E\)](#).

Covariates are often used in IV analysis to try to weaken instrument exogeneity. An instrument that is not exogenous unconditionally might still be exogenous conditional on a vector of covariates X_i . This type of reasoning is routine in empirical work, see for example [Gelbach \(2002, pg. 309\)](#), [Dinkelman \(2011, pg. 3091\)](#) or [Maestas et al. \(2013, pp. 1811–1812\)](#), to name just a few. In the classical IV model, covariates are introduced by including them in the outcome [equation \(1\)](#) and then assuming that they are orthogonal to any remaining latent variation in Y_i :

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \epsilon_i \quad \text{where } \mathbb{E}[X_i \epsilon_i] = 0. \quad (8)$$

This outcome equation is the same as [\(2\)](#) with the interaction term removed, a common specification when covariates are used to support exogeneity rather than for estimating heterogeneity along observables.

Covariates are introduced more directly in modern IV analysis by making exogeneity conditional on covariates. With a binary treatment and binary instrument, conditional full exogeneity is the assumption that $(Y_i(0), Y_i(1), D_i(0), D_i(1))$ is independent of Z_i , conditional on X_i . Conditional outcome exogeneity is that $(Y_i(0), Y_i(1))$ is independent of Z_i , conditional on X_i . Conditional independence is a nonparametric concept. This makes it easier to reason about than the orthogonality condition in the classical IV model [\(8\)](#), which also requires considering functional form: is a linear function of X_i enough to ensure orthogonality with ϵ_i , or are quadratic or more exotic terms needed as well? Separating the economic question of exogeneity from the statistical question of functional form is useful conceptually, even if (perhaps especially when) parametric functional forms end up being used for estimation in practice.

2.6 Target parameters

The classical IV model has a single homogeneous treatment effect, the coefficient α_1 on D_i . Allowing for treatment effect heterogeneity replaces this single effect with a distribution of effects across individuals. How do we want

to summarize this distribution? The answer to this question inherently depends on the researcher's motivation for causal inference.

We see two broad and not necessarily exclusive motivations: policy and "science." Policy means inference with the intent to evaluate a change in the way the treatment variable is assigned. For example, [Ito et al. \(2023\)](#) evaluate the welfare impacts of different incentive policies designed to encourage users to adopt electricity plans with dynamic pricing. The "science" motivation is a bit of a residual category, but could perhaps be thought of as knowledge for the sake of knowledge, without necessarily being used to guide a concrete decision. Understanding the effect of education on labor market outcomes is important for understanding fundamental issues about human capital, something which has value independent of any policy implications.

Both motivations produce empirical questions. *What would the gain in welfare be if an \$x incentive for dynamic pricing were provided? What is the average effect of a college degree on future earnings?* Answering either type of empirical question requires estimating quantities that summarize the distribution of treatment effects. We call these quantities target parameters.

The choice of target parameter tends to be clearer for policy questions. One way to evaluate a policy change is to estimate the average outcome that would occur under the new policy and compare that to the status quo average observed in the data. [Heckman and Vytlacil \(2001a, 2005\)](#) call this the policy-relevant treatment effect (PRTE). If the conjectured policy change is not observed in the data, then estimating a PRTE requires extrapolation. Policy changes that involve mandating or forbidding a treatment do not require modeling selection because treatment choice is fully determined in the counterfactual. Policy changes that involve changing incentives to take treatment do require a model of how those changes affect selection into treatment.

Choosing a target parameter for the vaguer "scientific" motivation is more open-ended, but can be guided by two reasonable principles. Fix a population of interest. For example, the entire population in a representative survey, the population of individuals at risk of disability or unemployment, or the subset of females in the population covered by a given study. Target parameters that reflect larger subpopulations of the population of interest are more interesting than those that reflect smaller and more specific subpopulations. Target parameters that can be clearly interpreted in terms of basic statistical quantities, such as means and quantiles, should also be preferred. These two principles are consistent with the way randomized controlled trials are evaluated: with simple quantities such as means and quantiles, and with an understanding that the results are specific to the population being studied.

If different target parameters answer different questions, then it stands to reason that some target parameters will be harder to estimate than others. "Harder" here means, loosely, that less can be learned under the same assumptions, and that stronger assumptions are needed to learn the same amount; it could also involve various measures of statistical difficulty.

Acknowledging the existence of this trade-off does not mean that there shouldn't be guiding principles to the choice of a target parameter. A reasonable approach to resolving this trade-off is to estimate a variety of target parameters that answer questions of different ambition under assumptions of different strengths.

An example of a target parameter that is often difficult to estimate is the average treatment effect (ATE), which is the overall average effect of a (binary) treatment on the population under study. The ATE may or may not answer an interesting policy question. In the [Ito et al. \(2023\)](#) study of dynamic electricity pricing, the ATE compares a policy in which all consumers have static pricing to one in which all consumers are mandated to have dynamic pricing. In the context of active labor market programs, the ATE compares a policy that mandates training to one that prohibits it, a mental exercise that probably has little policy relevance (e.g. [Heckman et al., 1999](#)). However, on the two guiding principles for a scientific motivation, the ATE scores at the top. Averages are perhaps the easiest summary of a distribution to understand, and the population reflected in the ATE is the overall population under study, same as in a randomized controlled trial. The ATE is, however, usually difficult to estimate with an IV while allowing for UHTE; it is generally not identified without additional assumptions beyond full exogeneity and the sharp bounds on the ATE are often too wide to be of practical interest.

A target parameter that is easier to estimate is the average treatment effect for the compliers to a binary instrument, the so-called local average treatment effect (LATE), the mechanics of which we discuss extensively ahead. Whether the LATE answers an interesting policy question depends on what the instrument is. [Ito et al. \(2023\)](#) randomly assigned an incentive of \$60 for adopting dynamic pricing relative to a baseline of no incentive. The LATE derived from this contrast provides a comparison of exactly this policy, which might be one potential policy of interest. The average effect for compliers is as easy to interpret as the ATE, but it only concerns the compliers, which are a smaller subset of the overall population. All things equal, LATEs that represent larger shares of compliers should be more interesting on a scientific basis.

Much ink has been spilt on the question of whether the LATE is an interesting target parameter compared to say, the ATE, or something else, such as the average treatment effect on the treated.⁸ In our view, extreme positions on this question are indefensible. Whether a given target parameter is interesting or relevant depends on the context and the empirical question, which is itself necessarily driven by the researcher's motivation for pursuing causal inference. How interesting a target parameter is also cannot be divorced from

⁸ An exhausting but not exhaustive list is [Angrist et al. \(1996\)](#), [Robins and Greenland \(1996\)](#), [Heckman \(1997\)](#), [Imbens \(2010\)](#), [Deaton \(2010\)](#), [Heckman and Urzua \(2010\)](#), [Pearl \(2011\)](#), and [Swanson and Hernán \(2014\)](#).

the difficulty involved in estimating it; there are trade-offs involved and reasonable people can disagree on how these trade-offs are resolved. Instead, the important and hopefully less controversial point is that the target parameter should be clearly stated and correctly interpreted. Not doing so obscures the empirical question that the analysis is intended to answer.

2.7 Testability

The traditional route for testing the classical linear IV model is an over-identification test with multiple instruments (Sargan, 1958). The logic of an overidentification test can be viewed as comparing the equality of multiple possible IV estimates of the same constant treatment effect; see Windmeijer (2019) for a precise statement. Such a test might reject because of UHTE rather than because the instrument fails to be excluded or exogenous.

There's a well-developed literature that provides alternative tests for IV models that allow for UHTE. These tests do not require multiple instruments and instead are based on whether statistical quantities that should reflect well-defined treatment effects or potential outcome distributions actually have the properties of such objects. For example, if the outcome Y_i is known to lie in $[-1, 1]$, then does an estimator that should reflect an average causal effect for a subpopulation actually lie in $[-1, 1]$? If not, then model can be rejected. We do not discuss testability in this chapter out of length considerations, but see Balke and Pearl (1997), Imbens and Rubin (1997), and Heckman and Vytlacil (2005) for discussions of the testable implications, Bhattacharya et al. (2012), Huber and Mellace (2014), Kitagawa (2015), Mourifié and Wan (2016), and Kédagni and Mourifié (2020) for various ways of turning these implications into formal statistical tests, and Carr and Kitagawa (2023), Frandsen et al. (2023), and Sun (2023) for more recent developments and applications.

3 Reverse engineering: interpreting linear estimators

If there is UHTE then the classical linear IV model is misspecified, but a linear IV estimator can still be computed. Perhaps it has an interpretation that is robust to omitted UHTE? This line of reasoning has been popular in the recent microeconomics literature. We describe it as reverse engineering because it starts with a tool—the estimator—and attempts to reverse engineer an interpretation for it. This section contains a comprehensive survey and synthesis of reverse engineering results for linear IV estimators.

In the next subsection, we begin by first introducing some concepts used in reverse engineering exercises. Then we use these concepts to review the well-known LATE interpretation that applies in the baseline case of a binary treatment, binary instrument, and no covariates. The remainder of the section then considers in turn what happens as one deviates from the baseline case by having either a non-binary instrument, a non-binary treatment, or by including covariates.

3.1 Estimators, estimands, and weak causality

Reverse engineering arguments start with an estimator and consider its associated estimand, meaning the population quantity to which the estimator can be expected to converge under a law of large numbers. For example, the estimand for the ordinary least squares estimator of the coefficient on D_i in a regression of Y_i on D_i and a constant is $\mathbb{C}[Y_i, D_i]/\mathbb{V}[D_i]$. The estimand is then decomposed into terms involving the underlying causal model, typically using potential outcome notation. The focus is on how assumptions about the underlying causal model affect the properties of the decomposition. In this way the estimator is taken as the starting point and its interpretation is then reverse engineered from an underlying causal model.

A minimal criterion for a successful interpretation is typically taken to be whether the estimand can be written in the form of a weighted average of mutually exclusive subgroup-specific average treatment effects with weights that are all non-negative. In an IV framework, an individual's group is defined by their unobserved potential choice behavior—the variable G_i introduced in Section 2.4—together with their observed covariates, X_i . Being able to write an estimand β as a non-negatively weighted average of group-specific treatment effects means that there exist weights $\omega(g, x) \geq 0$ such that

$$\beta = \sum_{g,x} \underbrace{\omega(g, x)}_{\text{weights}} \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|G_i = g, X_i = x]}_{\text{subgroup-specific treatment effects}}. \quad (9)$$

The non-negativity of the weights is seen as important because it guarantees that the estimand cannot systematically have the “wrong” sign. Suppose that all underlying covariate- and group-specific treatment effects are non-negative. If β can be written as (9) with weights that are non-negative, then β must also be non-negative. [Blandhol et al. \(2022\)](#) generalize this reasoning to include estimands that do not have decompositions like (9) either because the treatment takes multiple values or because the weights applied to the treatment arms are asymmetric. They call an estimand “weakly causal” if β is non-negative whenever the causal effect of all covariate- and group-specific treatment contrasts are non-negative. [Appendix B](#) provides the formal definition and a generalization to unordered treatments.

Weak causality is, as the name suggests, an extremely minimal requirement for an estimand to be viewed as “causal.” Any target parameter appropriate for the scientific motivation of estimating an average treatment effect for a subpopulation will be weakly causal. But a weakly causal estimand with non-constant weights need not reflect the average treatment effect for any single subpopulation.⁹ If all that is known about an estimand is that it is weakly

⁹ [Poirier and Słoczyński \(2024\)](#) show that a weakly causal estimand can, however, reflect the average treatment effect for a new subpopulation formed by combining subsets of multiple covariate and/or choice groups.

causal, then the scientific question it answers is an easy one based on a strong premise. *Assuming that everyone has either a positive or negative treatment effect, is the common sign positive or negative?*

Weak causality is not necessarily a useful property for policy purposes. A target parameter that answers a policy question might not be weakly causal if the policy shifts some individuals into treatment and others out of treatment. If such a target parameter had form (9) then it would give positive weight to groups induced by the policy change to increase treatment, but negative weights to those induced to decrease treatment. The opposite possibility can also arise: an estimand that is not weakly causal but *is* useful for policy. [Kline and Walters \(2016, Section V.C\)](#) provide an example of this in the context of unordered treatments, where the estimand conflates two different treatment contrasts but still reflects an important target parameter for a policy counterfactual.

Most weakly causal estimands also have weights that are convex in the sense of being both non-negative and summing to one across all subgroups: $\sum_{g,x} \omega(g, x) = 1$. The additional sum-to-one property ensures that if treatment effects are actually homogeneous, then β is equal to that common single effect. More generally, it ensures that β lies somewhere between the smallest and largest subgroup treatment effects, which seems like an intuitively attractive property.

The same estimand can have multiple different weakly causal interpretations. As a simple example, suppose that D_i is binary and randomly assigned, and let $\beta = \mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$ be the population difference in means. Then

$$\underbrace{\beta}_{\text{difference in means (estimand)}} = \underbrace{\frac{\text{overall ATE (interpretation #1)}}{\mathbb{E}[Y_i(1) - Y_i(0)]}}_{(10)} = \underbrace{\sum_x \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]\mathbb{P}[X_i = x]}_{\text{weighted average of conditional ATEs (interpretation #2)}}.$$

The first equality is the usual interpretation of β as the overall ATE, which is a convex weighted average of one term. The second equality shows that β can alternatively be interpreted as a convex weighted average of the conditional ATEs across all covariate groups.

3.2 Binary treatment, binary instrument, no covariates

The leading example in which the linear IV estimator has a clear interpretation that is robust to misspecification occurs in the simplest possible setting. Suppose that $D_i \in \{0, 1\}$ and $Z_i \in \{0, 1\}$ are both binary, and there are no additional covariates X_i . Assume that the monotonicity condition (or threshold-crossing model) in [Section 2.4](#) holds together with the full exogeneity condition discussed in [Section 2.5](#). [Imbens and Angrist \(1994\)](#) showed that under

these assumptions the average treatment effect for the compliers—what they called the local average treatment effect, or LATE—is identified:

$$\frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \underbrace{\mathbb{E}\left[Y_i(1) - Y_i(0) \mid \begin{array}{l} \text{subpopulation of compliers } (G_i=(0,1)) \\ D_i(0) = 0, D_i(1) = 1 \end{array}\right]}_{\text{average treatment effect for compliers}} \\ \beta_{\text{WALD}} \equiv \text{Wald estimand} \\ \equiv \text{LATE.} \quad (11)$$

Angrist and Imbens (1995) later called the left-hand side the Wald (1940) estimand. The same result appeared in the biostatistics literature in lesser-known papers by Permutt and Hebel (1989) and Baker and Lindeman (1994).¹⁰

Equation (11) is a natural and intuitive nonparametric identification result. In the absence of additional assumptions, the only subpopulation whose treatment effects could possibly be identified are the individuals whose decisions are causally affected by the instrument. Treatment effects for always-takers cannot be identified without some sort of extrapolation because they are never observed in the untreated state; the instrument has no causal effect on their treatment choice behavior. The same is true for never-takers, who are never observed in the treated state.

In general, both compliers and defiers are affected by the instrument, and the numerator of the Wald estimand reflects the aggregate change in outcomes that results from shifting compliers into treatment and defiers out of treatment. The monotonicity condition eliminates the defiers, leaving only the impact on compliers. The numerator of the Wald estimand—the reduced form—then reflects the aggregate change in outcomes caused by the instrument, which reflects both the size of the complier group and the impact that treatment has on their outcomes. The denominator—the first stage—adjusts for the size of the complier group:

$$\frac{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}{\text{denominator of the Wald estimand}} = \overbrace{\mathbb{E}[D_i(1) - D_i(0)]}^{\text{by full exogeneity}} = \overbrace{\mathbb{P}\left[\begin{array}{l} D_i(0) = 0, D_i(1) = 1 \\ \text{the complier choice group } G_i = (0,1) \end{array}\right]}^{\text{by monotonicity}}. \quad (12)$$

The ratio of the reduced form to the first stage—the Wald estimand—then reflects the average per-unit treatment effect among the compliers, which is the LATE defined in (11).

¹⁰The analysis of Permutt and Hebel (1989) is informal, but remarkably clear, elegant, and precise.

See in particular the bottom of page 621, where the authors recognize the four choice groups created by a binary treatment and binary instrument, followed by the monotonicity condition, the treatment effect for the compliers, the implied identification of the shares of each choice group, and the attenuation result for multivalued treatments formalized by Angrist and Imbens (1995, Section 3.1). The analysis of Baker and Lindeman (1994) is more formal, but a bit obscured by its embedding inside a “paired availability design;” however see the beginning of Section 3, Section 5, and Appendix I.

The misspecification-robust interpretation of the linear IV estimand as a LATE comes from the relationship

$$\underbrace{\text{LATE} = \beta_{\text{WALD}}}_{\text{from (11)}} \equiv \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \underbrace{\frac{\mathbb{C}[Y_i, Z_i]}{\mathbb{C}[D_i, Z_i]}}_{\text{simple IV estimand}} \equiv \beta_{\text{IV}}, \quad (13)$$

where the simple IV estimand β_{IV} refers to population coefficient on D_i for the linear IV estimator that instruments $[1, D_i]'$ with $[1, Z_i]'$. This equality between the Wald and IV estimands is specific to the case in which Z_i is binary and there are no covariates. It does not hold more generally. It is the source of the misspecification-robust interpretation that “linear IV is LATE.”

3.3 Multivalued instruments

Suppose that we generalize the setting slightly to allow Z_i to take on multiple values z_0, z_1, \dots, z_K , but that D_i is still binary.¹¹ Each instrument value now has an associated potential treatment $D_i(z_0), D_i(z_1), \dots, D_i(z_K)$.

[Imbens and Angrist \(1994\)](#) generalized the monotonicity condition by assuming that the values of the instrument can be placed in order of their impact on treatment choice, with an ordering that does not vary across individuals i . As in the binary case, this ordering does not need to be known a priori. Suppose that the instrument is indexed in increasing order, so that the monotonicity condition becomes $\mathbb{P}[D_i(z_0) \leq D_i(z_1) \leq \dots \leq D_i(z_K)] = 1$, meaning that larger instrument values make everyone more likely to take treatment. [Vytlačil \(2002\)](#) showed that the [Imbens and Angrist \(1994\)](#) assumptions are the same as modeling selection with the threshold-crossing model,

$$D_i = \mathbb{1}[V_i \leq \nu(Z_i)], \quad (14)$$

where ν is an unknown function. Full exogeneity is still required, and now means that Z_i is independent of $(Y_i(0), Y_i(1), D_i(z_0), D_i(z_1), \dots, D_i(z_K))$ or, equivalently, that Z_i is independent of $(Y_i(0), Y_i(1), V_i)$.

With $K + 1$ instrument values, there are 2^{K+1} treatment choice groups in general. The monotonicity condition implies that only $K + 2$ of these can exist: the always-takers, the never-takers, and K different complier groups, one for each subsequent pair of values for the instrument. [Table 1](#) illustrates. The same argument that produces (11) shows that the average treatment effect for each

¹¹ We focus on a discrete number of instrument values for simplicity. The continuous instrument case is conceptually similar and intuitively requires replacing sums with integrals. See [Alvarez and Toneto \(2024\)](#) for details.

TABLE 1 Group definitions when D_i is binary and Z_i takes four values ($K = 3$).

$D_i(z_0)$	$D_i(z_1)$	$D_i(z_2)$	$D_i(z_3)$	G_i	Group description
1	1	1	1	AT	Always-takers
0	1	1	1	CP ₁	z_1 -compliers ($G_i = CP_1$)
0	0	1	1	CP ₂	z_2 -compliers ($G_i = CP_2$)
0	0	0	1	CP ₃	z_3 -compliers ($G_i = CP_3$)
0	0	0	0	NT	Never-takers
0	1	0	1	DF	Defier (one of $2^4 - (3 + 2) = 11$ types)

Notes: When $K = 3$ the monotonicity condition allows for the six possible configurations of $G_i \equiv (D_i(z_0), D_i(z_1), D_i(z_2))$ shown here.

complier group is identified from the Wald estimand using these subsequent pairs:

$$\frac{\mathbb{E}[Y_i|Z_i = z_k] - \mathbb{E}[Y_i|Z_i = z_{k-1}]}{\mathbb{E}[D_i|Z_i = z_k] - \mathbb{E}[D_i|Z_i = z_{k-1}]} = \underbrace{\mathbb{E}\left[Y_i(1) - Y_i(0) \mid \overline{D_i(z_{k-1}) = 0, D_i(z_k) = 1}\right]}_{\text{the average treatment effect for } k-\text{compliers (LATE}_k\text{)}} \quad \text{one of several possible Wald estimands} \quad (15)$$

for each $k = 1, \dots, K$. We use the short-hand LATE _{k} for the right-hand side of (15), a notation which clarifies that there are now multiple LATEs. Even in this simple extension from the previous section, the statement that “linear IV is LATE” already doesn’t make sense: *which* LATE?

On top of that, *which* linear IV? With a single binary instrument, a binary treatment, and no covariates, the IV estimand given in (13) was the only one to consider. When the instrument takes multiple values, there are now many possible IV estimators, each producing a different IV estimand. A general formulation is to instrument for $[1, D_i]'$ with $[1, \zeta(Z_i)]'$, where ζ is a scalar function of Z_i . This nests using Z_i directly as an instrument for D_i , in which case the IV estimand is the same as in (13). It also nests any 2SLS estimand, in which case $\zeta(Z_i)$ are the population fitted values from the first stage.

Imbens and Angrist (1994) showed that this IV estimand can be decomposed as

$$\underbrace{\frac{\mathbb{C}[Y_i, \zeta(Z_i)]}{\mathbb{C}[D_i, \zeta(Z_i)]}}_{\text{linear IV estimand}} = \sum_{k=1}^K \underbrace{\frac{\mathbb{P}[G_i = CP_k] \mathbb{C}[\zeta(Z_i), \mathbb{I}[Z_i \in \{z_\ell\}_{\ell \geq k}]]}{\mathbb{C}[D_i, \zeta(Z_i)]}}_{\text{weights}} \text{LATE}_k. \quad (16)$$

The weights sum to one and can be shown to be non-negative if $\zeta(z)$ is non-decreasing in z . This is satisfied if $\zeta(z) = z$, so that the estimand is again the simple IV estimand on the right-hand side of (13). It is also satisfied for the 2SLS specification whose first stage includes an indicator for each value of the instrument, in which case $\zeta(z) = \mathbb{P}[D_i = 1|Z_i = z]$ becomes the propensity score. In either of these cases the IV/2SLS estimand on the left-hand side of (16) is weakly causal.

The weights in (16) are larger for larger complier groups, a feature that seems intuitive. However, the linear IV weights also vary with a second term that reflects how $\zeta(Z_i)$ and Z_i covary. One implication is that different choices of estimator—different choices of ζ —estimate different objects. A second implication is that the weights—and so also the estimand—depend on the marginal distribution of the instrument.

As an example, suppose that a researcher conducted a randomized experiment in which individuals were encouraged to take a treatment. Some individuals were given no additional incentive, while others were randomly assigned an incentive of \$10 or \$50 for taking treatment. Only 5 % of subjects in the unincentivized arm took treatment, while 20 % and 35 % took treatment in the \$10 and \$50 arms.¹² There are two complier groups in this setting: those who wouldn't take treatment if unincentivized, but would if offered \$10, and those who would only take treatment if offered \$50. Suppose that the average treatment effect for the former group is 1, and for the latter group is -1 , a difference which might reflect the second group's greater reluctance to participate. What will the IV estimand be? Fig. 2 shows that the answer depends delicately both on how the incentives were randomly assigned and on which IV estimand is being considered. The two lines indicate the values of two different IV estimands, one in which Z_i is used directly as an instrument as in (13), and one in which indicators for the different incentive arms are used in the first stage and combined through 2SLS. The left panel depicts a scenario in which the researcher assigns 75 % of individuals to the no incentive arm and assigns some proportion of the remaining 25 % to either the \$10 or \$50 incentive. When all of the incentivized individuals are in the low incentive arm, both estimands are the same and equal to the average treatment effect of 1 for the low incentive compliers. As the proportion assigned to the high incentive arm increases, the estimands begin to differ, and the value of the IV estimand starts to decrease as more high-incentive compliers are reflected in the estimand. The center and right panels of Fig. 2 show the same comparisons when a larger share of individuals are assigned to receive any incentive. The difference between the estimands grows and in many cases they even have the opposite sign.

This scenario is one in which the premise of weak causality is not met because the low incentive compliers have positive treatment effects, while the high incentive compliers have negative treatment effects. There are three LATEs: one

¹² For example, Dutz et al. (2022, 2023a,b) implemented incentivized surveys with this design, Ito et al. (2023) randomly assigned incentives for switching to dynamic electricity pricing, and Lee et al. (2019) randomly assigned incentives to connect to the electrical grid.

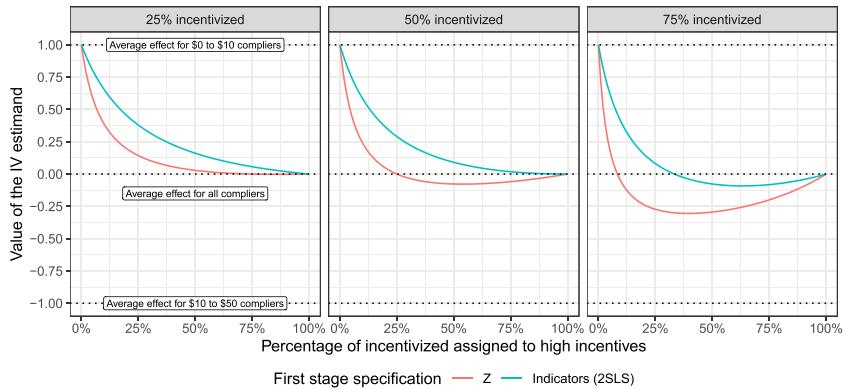


FIG. 2 The marginal distribution of the instrument affects the IV estimand. Notes: Values of the linear IV estimand in a hypothetical experiment in which individuals were incentivized to take a binary treatment. There are three incentive arms: no incentive, low (\$10) and high (\$50). The panels show the proportion of individuals assigned to the low or high arm. The x-axis shows the proportion of individuals in the high arm relative to the low arm. Two different estimands are shown: one that uses the level of the incentive ($Z_i = 0, 10, 50$) and the 2SLS estimand that uses indicators for each incentive level.

for each complier group separately, and a third when both complier groups are combined into a single group. The latter can generally be written as

$$\begin{aligned} \text{LATE}_{0 \rightarrow K} &\equiv \mathbb{E}[Y_i(1) - Y_i(0) | \underbrace{D_i(z_0) = 0, D_i(z_K) = 1}_{\text{complies with any instrument}}] \\ &= \sum_{k=1}^K \mathbb{P}[G_i = \text{CP}_k] \text{LATE}_k, \end{aligned} \quad (17)$$

which is like (16), but with weights that only depend on the complier proportions. As Fig. 2 shows, neither of the two IV estimands is generally equal to any of these three LATEs when both incentives are assigned. If only low incentives are assigned, then it's as if we are back in the binary instrument case, and both IV estimands are equal to the low incentive LATE of one. If only high incentives are assigned, then we are again in a binary instrument case, and both IV estimands are equal to the combined LATE, which is zero in this example.¹³ Outside of these two polar cases—that is, when the instrument actually takes multiple values—the IV estimand is not equal to any individual LATE.

3.4 Violations of monotonicity

The monotonicity condition plays a central role in the LATE identification result (11). There are many settings in which it is usually uncontroversial, such

¹³ The low and high compliers are $.20 - .05 = .15$ and $.35 - .20 = .15$ of the population, so the combined LATE is $\text{LATE}_{0 \rightarrow 2} = 1 \times .15 + (-1) \times .15 = 0$.

as the type of incentivized experiment just discussed, where the instrument is a monetary incentive for treatment. In other contexts there can often be more scope for contention.

As one example, consider the [Angrist and Evans \(1998\)](#) study of the effect of fertility on labor supply. The authors used the sex composition of a family's existing children as an instrument for further childbearing. They found that among families that have at least two children, those in which the first two children had the same biological sex were more likely to go on to have a third child than those that had both a male and female child. They interpreted this as reflecting a preference for sex mix among children. For the monotonicity condition to hold requires *all* families to have this same preference for sex mix. Monotonicity would be violated if there are families whose fertility stopping rule is to have two male children.

[Angrist et al. \(1996\)](#), Section 5.2) show how to conduct a sensitivity analysis in the binary instrument case. They show that when the monotonicity condition does not hold,

$$\begin{aligned}
 & \frac{\text{difference between Wald and LATE}}{\frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} - \text{LATE}} \\
 &= \underbrace{\left(\frac{\mathbb{P}[G_i = \text{DF}]}{\mathbb{P}[G_i = \text{CP}] - \mathbb{P}[G_i = \text{DF}]} \right)}_{\text{relative size of defier group}} \underbrace{(\text{LATE} - \mathbb{E}[Y_i(1) - Y_i(0)|G_i = \text{DF}])}_{\text{difference in treatment effects}}. \tag{18}
 \end{aligned}$$

The bias of the Wald estimand for the LATE has a product structure that is increasing in the size of the defier group and scaled by the difference between complier and defier average treatment effects. Both terms in the product need to be large for the Wald estimand to differ substantially from the LATE.

[Fig. 3](#) illustrates this point using estimates from [Angrist and Evans \(1998\)](#). The authors report a Wald estimate of approximately $-.13$, the denominator of which is estimated to be $.06$. If the monotonicity condition holds, then this implies that 6 % of the population are compliers and that the Wald estimate is the LATE. If the monotonicity condition doesn't hold because actually 2 % of the population are defiers and 8 % are compliers, then the Wald estimate would differ from the LATE, but not necessarily by much. For example, even if the 2 % defiers have treatment effects that are $-.24$ —nearly twice as negative as the Wald estimate of $-.13$ —this would still imply a LATE of approximately $-.16$.¹⁴

¹⁴ See [Noack \(2021\)](#) for a formal analysis of this example in terms of the more general concepts of the breakdown frontier and falsification region used in the literature on sensitivity (e.g. [Kline and Santos, 2013](#); [Masten and Poirier, 2020,2021](#)).

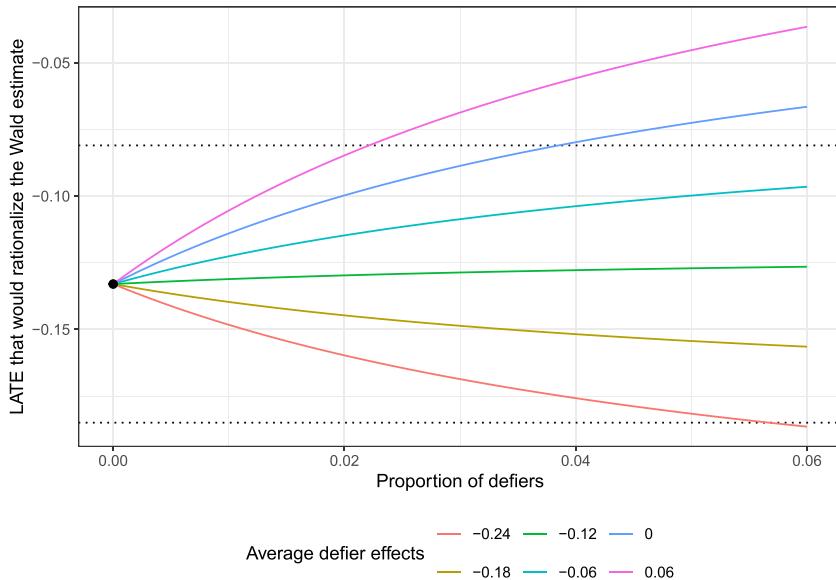


FIG. 3 Sensitivity of [Angrist and Evans \(1998\)](#) to violations of monotonicity. Notes: The Wald point estimate is taken to be $-.133$ in the graph, matching the estimate in Table 5 of [Angrist and Evans \(1998\)](#) for the “worked for pay” binary outcome. [Angrist and Evans \(1998\)](#) report a standard error of $-.026$, and the associated 95 % confidence intervals are indicated with dotted lines. We keep the denominator of the Wald estimate (the first stage) fixed at .060, consistent with the point estimate in Table 5 of [Angrist and Evans \(1998\)](#).

Judge designs are a common example of an IV strategy in which the monotonicity condition can be suspect. These designs are based on institutionally-pre-scribed random assignment of a judge or other examiner to cases in which the judge chooses treatment. If certain judges are systematically more likely to assign treatment, then the judge identities serve as an instrument for treatment (e.g. [Kling, 2006; Doyle, 2007; Dahl et al., 2014; Bhuller et al., 2020](#)).

The monotonicity condition places strong restrictions on the behavior of judges. Suppose that judge A is stricter than judge B in the sense that judge A assigns treatment in a higher proportion of cases. Then the monotonicity condition requires judge A to *always* assign treatment to any case in which judge B would assign treatment. This effectively prevents judges from systematically disagreeing. These types of settings were actually discussed in the original work by [Imbens and Angrist \(1994, Example 2\)](#) as an example where monotonicity might be an unattractive assumption.

[Frandsen et al. \(2023\)](#) propose a weaker alternative to monotonicity that they describe as “average monotonicity.” Their motivation is a judge design, although the assumption could be considered in other contexts as well. Suppose that the instrument Z_i denotes judge identity for $K + 1$ judges labeled

z_0, z_1, \dots, z_K . [Frandsen et al. \(2023\)](#) observe that the 2SLS estimand produced by using an indicator for each judge as an excluded variable can be written as

$$\beta_{2SLS} = \frac{\mathbb{E}[w_{AM}(G_i)\mathbb{E}[Y_i(1) - Y_i(0)|G_i]]}{\mathbb{E}[w_{AM}(G_i)]}, \quad (19)$$

where w_{AM} are weights defined as

$$\begin{aligned} w_{AM}\left(\underbrace{D_i(z_0), D_i(z_1), \dots, D_i(z_K)}_{\equiv G_i}\right) &\equiv \sum_{k=0}^K \mathbb{P}[Z_i = z_k](\mathbb{P}[D_i = 1|Z_i = z_k] - \mathbb{P}[D_i = 1]) \\ &\times \left(D_i(z_k) - \sum_{\ell=0}^K \mathbb{P}[Z_i = z_\ell]D_i(z_\ell)\right). \end{aligned} \quad (20)$$

In particular, (19) is true whether or not the monotonicity condition holds.

[Frandsen et al. \(2023\)](#) define average monotonicity as the assumption that the weights w_{AM} are all non-negative, meaning that $\mathbb{P}[w_{AM}(G_i) \geq 0] = 1$, or equivalently that $\mathbb{P}[G_i = g] = 0$ for any choice group g such that $w_{AM}(g) < 0$. As can be seen from (19), average monotonicity is equivalent to the assumption that β_{2SLS} is weakly causal. In this sense it assumes away the problems raised by failures of the usual monotonicity condition.

[Fig. 4](#) illustrates the content of the average monotonicity assumption using the data from [Stevenson \(2018\)](#) (provided by [Cunningham, 2021](#)), which has eight judges. Without any assumptions on potential treatment choice behavior there are $2^{K+1} = 2^8 = 256$ treatment choice groups, one for each configuration of potential judge decisions. The weight that each individual i contributes to β_{2SLS} is given by $w_{AM}(G_i)$, which depends only on their choice group, G_i . Although the share of each choice group is not identified, the weights $w_{AM}(g)$ are identified for each value of g . [Fig. 4](#) plots estimates of $w_{AM}(g)$ as a histogram taken over all 256 treatment choice groups.

The weights are symmetric around zero. If all groups occur, then exactly half will have non-negative weights while the other half will have non-positive weights.¹⁵ To see why, consider a pair of individuals i and i' who satisfy $D_i(z_k) = 1 - D_{i'}(z_k)$ for all k . For these individuals,

$$D_{i'}(z_k) - \sum_{\ell=0}^K \mathbb{P}[Z_i = z_\ell]D_{i'}(z_\ell) = -1 \times \left(D_i(z_k) - \sum_{\ell=0}^K \mathbb{P}[Z_i = z_\ell]D_i(z_\ell)\right), \quad (21)$$

which implies that $w_{AM}(G_{i'}) = -w_{AM}(G_i)$. An example of this symmetry is highlighted in [Fig. 4](#). One choice group is only assigned treatment for the third and sixth judge; their weights are positive. The other choice group is assigned treatment for all judges except the third and sixth; their weights are negative.¹⁶

¹⁵ Two groups always have zero weight: the always-takers and the never-takers.

¹⁶ The third and sixth judges are not particularly remarkable: the third one has the sixth highest propensity score and has the fifth most cases, while the sixth one has the seventh highest propensity score and has the sixth most cases.

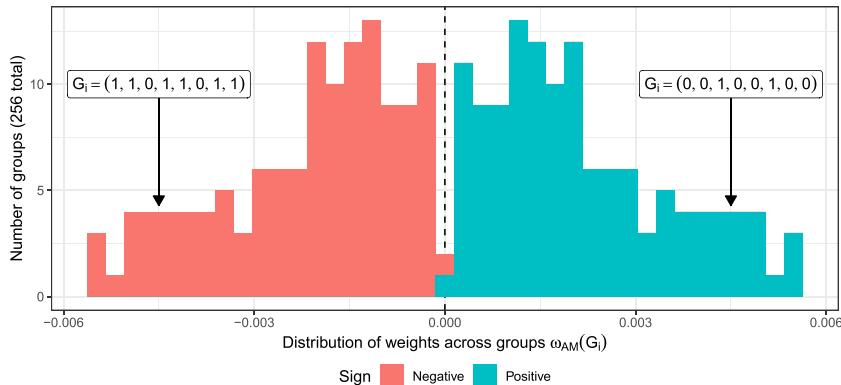


FIG. 4 Weights for the 2SLS estimand using the Stevenson (2018) data. Notes: The eight judges in the Stevenson (2018) data imply $2^8 = 256$ treatment choice groups (values of G_i). The weight $\omega_{AM}(G_i)$ that each individual i contributes to β_{2SLS} is shown on the x-axis for each value that G_i can assume. Without any monotonicity assumption, half of the groups contribute to the 2SLS estimand with negative weight. Average monotonicity is satisfied if and only if the group shares for this half are zero.

Justifying average monotonicity requires explaining why all the negatively-weighted groups in Fig. 4 do not exist. This seems challenging to do here, even with only eight judges. A judge design with one hundred judges has 2^{100} choice groups to consider, so that justifying average monotonicity requires arguing that half (2^{99}) of these groups do not occur, presumably an even more difficult feat. Compounding this difficulty, the identity of the groups that must not occur depends on the observable data through the propensity scores (leniencies) of all judges and the frequency with which each judge is observed. This makes any attempt at justifying average monotonicity inherently application-specific.¹⁷ While average monotonicity is mathematically weaker than the usual monotonicity condition, it is not clear that it is any easier to justify on substantive grounds.¹⁸

Reverse engineering arguments grind to a halt without some sort of monotonicity condition. de Chaisemartin (2017) shows that the IV estimand can still be interpreted as representing average treatment effects for some

¹⁷ Chyn et al. (2024, Appendix A) consider the content of average monotonicity in a hypothetical context with four judges. They show that average monotonicity can be satisfied under an assumption that reduces individuals to four latent types: minority/majority and with/without a criminal background. Their justification involves stylized assumptions on the frequency of these types in the population and the way in which the four judges respond to these types.

¹⁸ Sigstad (2024b) estimates treatment group shares using judge panels under the assumption that judges rule in panels the same way that they would individually. He finds evidence against the usual monotonicity condition and somewhat weaker evidence against the average monotonicity condition.

subset of the compliers, even if the monotonicity condition is not satisfied, as long as the distribution of treatment effects between the compliers and defiers is not too different. The result effectively provides conditions under which a single weighted average with some negative weights—the IV estimand without a monotonicity condition—is equal to a weighted average of some subset of its non-negatively weighted components. Whether an average treatment effect for the group represented by such a subset is interesting (or indeed, even uniquely defined) is another question, but the result provides a thought-provoking examination of the logical limits of reverse engineering. If all we want to know is whether a given estimate represents a treatment effect for *someone*, then we can do that under general conditions. Is that robustness or superficiality?

3.5 Multiple instruments

The credibility of the monotonicity condition is not about the number of values that the instrument takes. Compare monetary incentives to judges. If it's reasonable to assume that everyone prefers treatment with \$10 to treatment without compensation, then it's also reasonable to assume that everyone prefers treatment with \$50 to treatment without compensation. Adding another \$100 incentive arm would not jeopardize this argument. On the other hand, if we're worried that any two of eight judges may disagree on some cases, then we would probably also have that concern in a setting with only two judges. The issue instead is whether the instrument has a natural ordering: ordering monetary incentives from small to large is natural, but ordering judges, say in terms of their leniency, attempts to reduce something complicated (a judge) down to one dimension.

This ordering issue arises when Z_i is a vector containing multiple distinct components, a case we describe as multiple instruments. Multiple instruments are always multivalued. For example, if Z_i is a vector of two binary instruments, then Z_i takes four values. If the monotonicity condition holds for these four values, then all of the discussion in the previous section for multivalued instruments continues to apply with multiple instruments.¹⁹ But the fact that the four values represent a combination of two distinct components usually makes the monotonicity condition an extremely unattractive assumption.

To see why, suppose that incentives for treatment vary along two dimensions: a monetary incentive, and the distance to the location where treatment is administered. An experimental design like this was used by Thornton (2008) to study the demand for learning about HIV status in Malawi. For simplicity, suppose that both instruments are binary, with $Z_{i1} \in \{0, 1\}$ denoting a monetary incentive (no or yes), and $Z_{i2} \in \{0, 1\}$ denoting distance (far or near), so that there are four potential treatments $D_i(z_1, z_2)$. The monotonicity condition requires these potential treatments to be ordered for all individuals; in particular, it requires either

¹⁹ Imbens and Angrist (1994, pg. 470) explicitly mention the case in which Z_i is a vector in deriving the decomposition (16).

$\mathbb{P}[D_i(0, 1) \geq D_i(1, 0)] = 1$ or that $\mathbb{P}[D_i(1, 0) \geq D_i(0, 1)] = 1$. The first condition says that there is no one who would take treatment if they were given a monetary incentive but not take treatment if they were assigned to a close location. The second condition says the opposite: there is no one who would take treatment if they were assigned to a close location but not if they were given a monetary incentive. Either condition assumes that there is no meaningful heterogeneity in the opportunity cost of time (responsiveness to distance).²⁰

If the monotonicity condition is dropped entirely, then a linear IV estimand will include negatively-weighted average treatment effects for some groups, and so fail to be weakly causal. The problem here is the model of treatment choice: some model is needed for a weakly causal interpretation, but the monotonicity (threshold-crossing) model is too strong to be credible. Mogstad et al. (2021) consider an intermediate model of treatment choice called partial monotonicity. Partial monotonicity requires the usual monotonicity condition to be satisfied for each instrument separately, while holding all other instruments fixed. For example, it is satisfied if all individuals are more likely to take treatment when given a monetary incentive and when closer to a treatment center, but it does not require one or the other to be a uniformly more effective inducement to treatment. Mogstad et al. (2021) show that partial monotonicity can be sufficient for a weakly causal interpretation: with two binary instruments, a 2SLS estimand with a saturated first stage will be weakly causal as long as the instruments are not negatively correlated.

3.6 Ordered, cardinal treatments

The monotonicity condition is a model of treatment choice, so it must be reconsidered when the treatment has more than two values. Suppose that the treatment takes $J + 1$ values labeled in increasing order as d_0, d_1, \dots, d_J .²¹ If the treatment is ordered, then a natural generalization of the monotonicity condition is $\mathbb{P}[D_i(1) \geq D_i(0)] = 1$. Both the notation and content are the same as in the binary case: receiving the instrument causes treatment to weakly increase for everyone.

The natural generalization of the binary threshold-crossing model (5) to an ordinal treatment would be to an ordered response model (e.g. Greene and Hensher, 2009). Depending on how the ordered response model is specified, it may or may not entail the same restrictions on treatment choice behavior as the monotonicity condition (Vytlačil, 2006). We return to this point in Section 4.5,

²⁰This point was first made by Heckman and Vytlačil (2005, Section 6) and Heckman et al. (2006, Section III.D). As those authors observed, the “monotonicity” condition is not really about monotonicity, but about uniformity in treatment choice behavior. The two descriptions often coincide when the instrument is scalar, but they become meaningfully different with multiple instruments.

²¹All of the conceptual issues in the following discussion extend to the case of a continuous treatment, essentially by replacing sums with integrals and finite differences with derivatives; see (Angrist et al., 2000).

where it becomes particularly salient. In this section, we consider reverse engineering under the [Angrist and Imbens \(1995\)](#) monotonicity condition.

Let $Y_i(d_0), Y_i(d_1), \dots, Y_i(d_J)$ be potential outcomes for each treatment state. Even for a given individual i there is no longer a single treatment effect because $Y_i(d_2) - Y_i(d_1)$ could be different than $Y_i(d_1) - Y_i(d_0)$ if the size of the treatment increment differs and/or treatment effects are nonlinear. [Angrist and Imbens \(1995\)](#) showed that if Z_i is binary then the Wald estimand, which is still equal to the simple IV estimand, has the following decomposition:

$$\beta_{IV} \equiv \frac{\mathbb{C}[Y_i, Z_i]}{\mathbb{C}[D_i, Z_i]} = \sum_{j=1}^J \omega_{ACR}(j) \mathbb{E}[Y_i(d_j) - Y_i(d_{j-1}) | D_i(1) \geq d_j > D_i(0)],$$

where $\omega_{ACR}(j) \equiv \frac{\mathbb{P}[D_i(1) \geq d_j > D_i(0)]}{\sum_{\ell=1}^J \mathbb{P}[D_i(1) \geq d_\ell > D_i(0)](d_\ell - d_{\ell-1})}$. (22)

The decomposition allows for treatment effects that are both nonlinear and heterogeneous.²² The treatment variable D_i should have a cardinal interpretation to consider β_{IV} . If it is ordered but not cardinal (e.g. low, medium, high), then β_{IV} will be sensitive to the arbitrary coding of the values d_j , making estimators that use different treatment indicators more appropriate. This case is discussed in the next section.

[Angrist and Imbens \(1995\)](#) described the right-hand side of (22) as the average causal response (ACR). The unit “causal response” is the effect of increasing treatment from d_{j-1} to d_j , and the “average” is a weighted one taken across all treatment indices j . The weights in the average are proportional to the probability of the event that $D_i(1) \geq d_j > D_i(0)$, and the causal effects are conditioned on this event. This event includes all individuals i that would have treatment value larger than d_j with the instrument, but strictly smaller than d without it. Individuals unaffected by the instrument ($D_i(1) = D_i(0)$) do not contribute to the ACR. The effect of increasing treatment from d_{j-1} to d_j for individuals who would never have either treatment level (i.e. $D_i(0) \geq d_j$ or $D_i(1) \leq d_{j-1}$) does not enter into the ACR. These properties are sensible analogs of the binary treatment LATE identification result. The weights are non-negative, so the ACR is weakly causal. They do not sum to one unless $d_j - d_{j-1} = 1$ for all j , which is the case discussed in [Angrist and Imbens \(1995\)](#) and [Angrist and Pischke \(2009\)](#).²³

²² Expression (22) is slightly more general than the one in [Angrist and Imbens \(1995\)](#) because they assume that the treatment is coded as $d_j = j$, so that each treatment value is one increment apart. A derivation of (22) is in [Appendix C](#).

²³ The fact that the weights do not sum to one is not necessarily a concern. Compare one IV estimand with $J = 2$ in which d_0, d_1 , and d_2 are coded as 0, 1 and 2, to another in which they are coded as 0, 2, 4. The latter estimand will be half the size of the former, and its weights will also sum to one half. This is because the unit causal response does not depend on the coding of the treatment, implying that the weights must depend on it.

The ACR has been criticized on the grounds that the conditioning events it represents are not mutually exclusive. That is, an individual with $D_i(1) = d_2$ and $D_i(0) = d_0$ gets “double counted” in both of the events $D_i(1) \geq d_1 > D_i(0)$ and $D_i(1) \geq d_2 > D_i(0)$. [Angrist and Imbens \(1995\)](#), pg. 435–436) and [Heckman et al. \(2006\)](#) both discuss this criticism, although the former authors downplay it on the grounds that they do not expect the instrument in their example to have more than a one unit effect. [Heckman et al. \(2006, Section VI\)](#) provide comparable reverse engineering results under an ordered response model and observe that the same criticism does not arise; see [Section 4.5](#) for more discussion.

One way to address this criticism while sticking with the [Angrist and Imbens \(1995\)](#) monotonicity condition is to write the ACR as a different weighted average in which individuals only appear once.²⁴ Using the group notation $G_i \equiv (D_i(0), D_i(1))$, let $\mathcal{G} \equiv \{(g(0), g(1)) : g(1) \geq g(0)\}$ denote the subset of the $(J + 1)^2$ possible groups that can have non-zero probability under the monotonicity condition. In [Appendix C](#), we show that

$$\beta_{IV} = \sum_{g \in \mathcal{G}} \underbrace{\frac{\mathbb{P}[G_i = g](g(1) - g(0))}{\sum_{g' \in \mathcal{G}} \mathbb{P}[G_i = g'](g'(1) - g'(0))}}_{\text{weights reflect group size and instrument effect on treatment}} \mathbb{E} \left[\underbrace{\frac{Y_i(g(1)) - Y_i(g(0))}{g(1) - g(0)}}_{\text{average per-unit treatment effect for group } g} \right]. \quad (23)$$

The conditioning events in this decomposition are mutually exclusive because each individual belongs to exactly one choice group. The treatment effects being weighted are expressed in per-unit averages across the range of values that the instrument shifts a group’s treatment choice. Researchers who find the ACR hard to appreciate due to overlapping conditioning events may find (23) more attractive.²⁵

Even so, multivalued treatments are undoubtedly more complicated than binary treatments. Researchers are often tempted to binarize a treatment in order to avoid this complication. [Angrist and Imbens \(1995\)](#) show that this practice leads to a Wald estimand that is larger in magnitude than the ACR. [Marshall \(2016\)](#), [Andresen and Huber \(2021\)](#), and [Rose and Shem-Tov \(2023\)](#) consider additional assumptions under which the binarized Wald estimand has a more attractive interpretation.

3.7 Unordered or non-cardinal treatments

The previous reverse engineering results all consider linear IV specifications with the treatment as the sole endogenous variable. These specifications only

²⁴ This point was made by [Frölich \(2007\)](#), pg. 50). [Equation \(23\)](#) is an alternate phrasing of his [equation \(19\)](#).

²⁵ For the ACR decomposition (22), one could also multiply the weights by $d_j - d_{j-1}$ and divide the unit causal response by $d_j - d_{j-1}$ to get a third expression in which the weights sum to one.

make sense if the treatment has a natural cardinal ordering. If it doesn't, either because it's ordered but not cardinal, or because it's unordered, then the natural specification to consider is one with multiple endogenous variables that are indicators for different treatment states or sets of states.

The simplest setting is when D_i takes one of three treatment states, d_0 , d_1 , or d_2 , which are coded up using two binary endogenous variables, $D_{i1} \equiv \mathbb{I}[D_i = d_1]$ and $D_{i2} \equiv \mathbb{I}[D_i = d_2]$. The order condition requires at least two excluded variables for a linear IV estimand to be defined. Suppose that we have access to an instrument Z_i that takes three values, 0, 1, and 2, which have been similarly coded into two binary variables, Z_{i1} and Z_{i2} . Potential outcomes $Y_i(d_j)$ and potential treatments $D_i(z)$ are defined as before, with $D_{ij}(z) \equiv \mathbb{I}[D_i(z) = d_j]$ giving the implied potential binary treatment indicators. Full exogeneity is assumed, as usual.

With three treatment states and three instrument values there are $3^3 = 27$ a priori possible choice groups G_i reflecting different combinations of $(D_i(0), D_i(1), D_i(2))$. Suppose that we generalize the monotonicity condition to the assumption that instrument values one and two are targeted towards the corresponding first and second treatment states, and that receipt of these instrument values weakly pushes all individuals towards those states. The formal assumption is that $D_{i1}(1) \geq D_{i1}(0)$ and $D_{i2}(2) \geq D_{i2}(0)$, so that receiving $Z_i = j$ makes choosing $D_i = d_j$ more likely than when $Z_i = 0$. This eliminates 17 of the 27 choice groups, leaving the first ten shown in [Table 2](#). The final row of [Table 2](#) gives an example of a group that doesn't satisfy this monotonicity condition: $G_i = (d_1, d_0, d_2)$ would choose $D_i = d_0$ when $Z_i = 1$, but would choose $D_i = d_1$ when $Z_i = 0$, violating the assumption that $Z_i = 1$ encourages takeup of state d_1 for everyone.

The linear IV estimand for this case is the one with outcome equation linear in a constant, D_{i1} , and D_{i2} , and first stage variables mirrored with a constant, Z_{i1} , and Z_{i2} . The coefficients on D_{i1} and D_{i2} can be written as a vector by partialling out the constant (demeaning) from the excluded variables:

$$\beta_{IV} \equiv \begin{bmatrix} \beta_{IV,1} \\ \beta_{IV,2} \end{bmatrix} = \mathbb{E} \left[\begin{bmatrix} \tilde{Z}_{i1} \\ \tilde{Z}_{i2} \end{bmatrix} \begin{bmatrix} D_{i1} \\ D_{i2} \end{bmatrix}' \right]^{-1} \mathbb{E} \left[\begin{bmatrix} \tilde{Z}_{i1} \\ \tilde{Z}_{i2} \end{bmatrix} Y_i \right] \quad \text{where } \tilde{Z}_{ij} \equiv Z_{ij} - \mathbb{E}[Z_{ij}]. \quad (24)$$

For this unordered case, $\beta_{IV,j}$ is weakly causal if it is non-negative whenever $\mathbb{E}[Y_i(d_j) - Y_i(d_0)|G_i = g]$ is non-negative for all g ([Appendix B](#)).

[Kirkeboen et al. \(2016\)](#) and [Heinesen et al. \(2022\)](#) show that neither of the components of β_{IV} are even close to weakly causal under the given monotonicity condition. Instead, $\beta_{IV,1}$ captures a complicated weighted average of $Y_i(d_1) - Y_i(d_0)$ and $Y_i(d_2) - Y_i(d_0)$ involving all of the seven non-always-taker groups not ruled out by monotonicity.²⁶ Many of the weights will be negative.

²⁶The full expression can be found in Proposition 3 of [Heinesen et al. \(2022\)](#).

TABLE 2 Choice groups for an unordered treatment with three states.

G_i	Mon.	EM	IR	NB	KLM
$(d_0, d_0, d_0) \equiv g_1$	✓	✓	✓	✓	✓
$(d_0, d_1, d_0) \equiv g_2$	✓	✓	✓	✓	✓
$(d_0, d_0, d_2) \equiv g_3$	✓	✓	✓	✓	✓
$(d_0, d_1, d_2) \equiv g_4$	✓	✓	✓	✓	✓
$(d_1, d_1, d_1) \equiv g_5$	✓	✓	✓		
$(d_2, d_2, d_2) \equiv g_6$	✓	✓	✓		
$(d_1, d_1, d_2) \equiv g_7$	✓		✓		
$(d_2, d_1, d_2) \equiv g_8$	✓		✓		
$(d_0, d_1, d_1) \equiv g_9$	✓			✓	
$(d_0, d_2, d_2) \equiv g_{10}$	✓			✓	
:					
(d_1, d_0, d_2)					
:					

Notes: The groups shown are for a treatment that takes three states d_0, d_1, d_2 and an instrument that takes three values 0, 1, 2 so that choice groups are determined by a combination of $G_i \equiv (D_i(0), D_i(1), D_i(2))$. There are 27 groups possible a priori. Only those satisfying the natural extension of the monotonicity condition (Mon.) are shown, together with one example of a group that violates that condition. Acronyms: EM (extended monotonicity), IR (irrelevance), NB (next-best), and KLM (Kirkeboen et al., 2016).

So $\beta_{IV,1}$ fails to reflect the sign of $E[Y_i(d_1) - Y_i(d_0)|G_i = g]$ both because it negatively weights some groups and because it also reflects potential outcomes for treatment state d_2 . Symmetric conclusions apply to $\beta_{IV,2}$.

The reason this happens is that the monotonicity condition does not sufficiently restrict treatment choice behavior. Groups with $D_i(1) = d_1$ who choose the first state when its instrument is switched on ($Z_i = 1$) could switch to choosing either d_0 (groups g_2, g_5, g_9) or d_2 (group g_8) when switched off. A contrast between $Z_{i1} = 1$ and $Z_{i1} = 0$ would not isolate a single treatment contrast even if it were able to keep Z_{i2} fixed, which the linear IV estimand (24) does not. More assumptions on treatment choice behavior are needed.

An early proposal by Behaghel et al. (2013) is to impose an “extended monotonicity” (EM) condition that eliminates groups g_7 through g_{10} . Their motivation was a multi-armed encouragement design in which individuals are given a specific encouragement to take the first or second treatment. The EM restriction rules out groups like g_7 who would enter into treatment d_2 when encouraged to, but would switch to treatment d_1 when encouraged to or when not encouraged at all. Under EM, the authors show that $\beta_{IV,1}$ is equal to $E[Y_i(d_1) - Y_i(d_0)|D_{i1}(1) = 1, D_{i1}(0) = 0]$, which is the average treatment effect of state d_1 relative to d_0 for the combined

complier group $G_i \in \{g_2, g_5\}$ that responds to $Z_i = 1$. Similarly, $\beta_{IV,2}$ is equal to $\mathbb{E}[Y_i(d_2) - Y_i(d_0)|D_{i2}(2) = 1, D_{i2}(0) = 0] = \mathbb{E}[Y_i(d_2) - Y_i(d_0)|G_i = g_4 \text{ or } g_5]$. These quantities have the hoped-for interpretation of average treatment effects for their respective treatment states relative to the omitted state d_0 for their respective sub-population of compliers.

[Kirkeboen et al. \(2016\)](#) point out that EM is a strong restriction on preferences. Under EM, an individual who chooses state d_2 when its cost is low cannot switch to state d_1 when the cost of d_2 becomes high. Even in an experimental setting, this may not be an attractive assumption unless there is one-sided noncompliance. [Behaghel et al. \(2013\)](#) note that EM has the testable implication that $\mathbb{P}[D_i = d_2|Z_i = 1] = \mathbb{P}[D_i = d_2|Z_i = 0]$, meaning that the share of those choosing d_2 should be the same among those encouraged to choose d_1 and those not encouraged to choose either d_1 or d_2 , with a similar implication for the probability that $D_i = d_1$. They reject EM for part of their experiment, as do [Kirkeboen et al. \(2016\)](#) in a different context.

[Kirkeboen et al. \(2016\)](#) consider two alternative assumptions. The first, which they call “irrelevance” (IR), is that if the instrument value that encourages d_2 doesn’t actually induce d_2 , then it doesn’t instead induce d_1 . This rules out group g_9 , who has $D_i(2) = d_1$ but $D_i(0) = d_0$. It also rules out group g_{10} symmetrically. The second assumption, which they call “next best” (NB), is that they can observe in the data who would choose d_0 if assigned $Z_i = 0$. In their application to estimating the returns to field of study, they justify the NB assumption by restricting their sample to individuals who list the same second-preferred field of study in a centralized college admissions process. Conditioning on $D_i(0) = d_0$ excludes members of groups g_7 and g_8 .²⁷ Together, IR and NB leave the same choice groups as under EM, and so also imply that $\beta_{IV,1}$ and $\beta_{IV,2}$ reflect average treatment effects among their respective complier groups (g_2, g_5 and g_4, g_5), although now conditioned on the sample selection rule of having a particular next best alternative.

[Bhuller and Sigstad \(2024\)](#) show that the conditions used by [Behaghel et al. \(2013\)](#) and [Kirkeboen et al. \(2016\)](#) are necessary for β_{IV} to be weakly causal interpretation.²⁸ In particular, the necessary (and sufficient) condition is that the each instrument essentially only affects the indirect utility of one treatment state, and that the excluded treatment state d_0 is always either the preferred or next best alternative.²⁹ The only groups in [Table 2](#) that satisfy this description

²⁷ It also eliminates the always-taker groups g_3 and g_6 , however these groups get differenced out regardless.

²⁸ The [Bhuller and Sigstad \(2024\)](#) results apply to any number of J discrete treatment states with β_{IV} defined as the coefficients on $J - 1$ indicators that are instrumented for by $J - 1$ binary instruments.

²⁹ As usual in discrete response, the minor caveat suggested by “essentially” is the possibility that an instrument shifts the indirect utility of choosing a different state as long as the shift is inframarginal and does not result in a discrete change in choice behavior.

are g_1 through g_6 . The implication of the [Bhuller and Sigstad \(2024\)](#) result is that the linear IV estimand will only have a sensible interpretation if heterogeneity in choice behavior is heavily restricted, as in [Behaghel et al. \(2013\)](#), or if one has data on next-best alternatives, as in [Kirkeboen et al. \(2016\)](#). [Heinesen et al. \(2022\)](#) show how to conduct a sensitivity analysis similar to the one in [Section 3.4](#), which suggests that given the monotonicity condition, the interpretation of the 2SLS estimand given in [Kirkeboen et al. \(2016\)](#) will be relatively insensitive to modest violations of either IR or NB (but not both) if the amount of treatment effect heterogeneity is also modest.

In some cases, the values d_0 , d_1 , and d_2 might have a natural ordering even if not a cardinal interpretation. An example studied by [Arteaga \(2023\)](#), [Humphries et al. \(2023b\)](#), and [Kamat et al. \(2024\)](#) is criminal case outcomes, where d_2 is incarcerated, d_1 is convict with no incarceration, and d_0 is do not convict. [Bhuller and Sigstad \(2024\)](#) characterize sufficient and necessary conditions for weak causality of the linear IV estimand in (24) coded up as $D_{i1} = \mathbf{1}[D_i \geq d_1]$ and $D_{i2} = \mathbf{1}[D_i = d_2]$. With constant effects, $\beta_{IV,1}$ would be interpreted as the causal effect of d_1 relative to d_0 and the coefficient on $\beta_{IV,2}$ would be interpreted as the incremental causal effect of d_2 relative to d_1 . [Bhuller and Sigstad \(2024\)](#) show that if treatment follows an ordered response model, then an additional parametric assumption of linearity between the first stage fitted values for the two treatments is necessary for a weakly causal interpretation. [Humphries et al. \(2023b\)](#) and [Kamat et al. \(2024\)](#) both conclude that the implied restrictions on choice behavior are unattractive in their setting.

Given these difficulties, one might consider using a linear IV estimand with only a single treatment indicator, such as D_{i2} , and instrumenting for it with Z_{i2} alone. The concern here is substitution bias (e.g. [Heckman et al., 2000](#)). As several authors have noted, the resulting IV estimand will be equal to a weighted average of treatment effects relative to both d_0 and d_1 depending on which treatment state would have been chosen if $Z_{i2} = 0$ (e.g. [Kirkeboen et al., 2016](#); [Kline and Walters, 2016](#); [Mountjoy, 2022](#)). The conflation of two different treatment contrasts makes such a quantity not weakly causal and generally makes it difficult to interpret even under restrictive assumptions on choice behavior; see [Mountjoy \(2022\)](#) for one clear example. [Kline and Walters \(2016\)](#) show that a quantity that conflates two different treatment contrasts can, however, still be useful for evaluating policy questions that do not require accounting for substitution bias. [Humphries et al. \(2023b\)](#) show how an interpretation in terms of a single treatment contrast can be restored under restrictions on choice behavior if the instruments are probabilities of assignment to the treatment states, as in a judge design, and one of the instruments is conditioned on as a covariate.³⁰

³⁰In particular, the authors maintain the unordered partial monotonicity condition used by [Mountjoy \(2022\)](#). They point out that this restriction has some unattractive implications for behavior in a judge design, which they find some evidence against in their data.

What about multiple different treatment variables collected into a vector? This would be an unordered treatment even if the individual components are all ordered and cardinal. We are not aware of any work for this case on reverse engineering linear IV with UHTE. The difficulties encountered with the scalar unordered case suggest that the conclusion is unlikely to be inspiring.

3.8 Covariates

Covariates X_i have so far been absent from our discussion of reverse engineering. Yet they often play an important role in bolstering the credibility of the exogeneity assumption. Their inclusion complicates reverse engineering interpretations considerably.

3.8.1 Controlling for covariates nonparametrically

Conditioning on covariates nonparametrically doesn't have any substantive implications for reverse engineering, since it only changes the subpopulation to which the arguments and assumptions must apply. For example, dropping observations that don't fall into the conditioning set leads to conditional versions of any of the previous interpretations. These can be aggregated or compared across conditioning sets as desired. Estimates from a crude binning approach like this may be too noisy to be useful. Nonparametric machine learning estimators are an appealing alternative, but they are not linear IV (see [Section 4.2](#)). The current dominant practice is instead to condition on covariates by controlling for them linearly.

3.8.2 Controlling for covariates linearly

Linearly controlling for a vector of covariates X_i changes the IV estimand to

$$\beta_{IV} \equiv \frac{\mathbb{E}[Y_i \tilde{Z}_i]}{\mathbb{E}[D_i \tilde{Z}_i]} \quad \text{where } \tilde{Z}_i \equiv Z_i - \underbrace{X' \frac{\mathbb{E}[X_i X']^{-1} \mathbb{E}[X_i Z_i]}{\mathbb{E}[X_i X']^{-1} \mathbb{E}[X_i Z_i]}}_{\text{population fitted values from linear regression of } Z_i \text{ onto } X_i}$$

$$\equiv Z_i - X' \delta. \tag{25}$$

[Equation \(25\)](#) has the same reduced-form-to-first-stage structure as the simple IV estimand without covariates (the right-hand side of [equation \(13\)](#)), except that the instrument is residualized against the covariates first, leaving an effective instrument, \tilde{Z}_i . This effective instrument contains variation due to both the instrument *and* the covariates. The two sources of variation can be separated in the reduced form (the numerator):

$$\frac{\text{numerator of IV estimand}}{\mathbb{E}[Y_i \tilde{Z}_i]} = \mathbb{E}[\mathbb{E}[Y_i \tilde{Z}_i | X_i]] = \frac{\text{variation in } Y_i \text{ caused by } Z_i}{\mathbb{E}[\mathbb{C}[Y_i, Z_i | X_i]]} + \frac{\text{covariation between } Y_i \text{ and } X_i}{\mathbb{E}[Y_i \mathbb{E}[\tilde{Z}_i | X_i]]}. \tag{26}$$

The first term contains the type of variation we hope to extract with an IV estimator: the relationship between Y_i and Z_i , *conditional* on X_i . In contrast, the second term reflects variation between Y_i and a function of X_i , something that isn't part of the nonparametric rationale of an IV strategy.

The variation in the second term *is* part of the rationale of controlling for covariates in the classical model (8). This is because the random variable $\mathbb{E}[\tilde{Z}_i|X_i]$ contained in the second term is also orthogonal to X_i .³¹ As a consequence, the second term—like the first term—still reflects the constant treatment effect, β_1 :

$$\text{substitute (8), } Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 X_i + \varepsilon_i \\ \overbrace{\mathbb{E}[Y_i \mathbb{E}[\tilde{Z}_i|X_i]]}^{} = \underbrace{\beta_1 \mathbb{E}[D_i \mathbb{E}[\tilde{Z}_i|X_i]]}_{\text{treatment effect (scaled)}} + \beta_2 \underbrace{\mathbb{E}[X_i \mathbb{E}[\tilde{Z}_i|X_i]]}_{\mathbb{E}[X_i \tilde{Z}_i] = 0}. \quad (27)$$

This happy simplification only occurs because of the structure of (8): a constant coefficient on D_i , and a linear adjustment for X_i . The very point of reverse engineering an interpretation for the IV estimand is to avoid these assumptions.

3.8.3 Level-dependence caused by covariates

Heterogeneous treatment effects in particular imply that the second term of (26) will generally reflect problematic variation. Consider the binary treatment, binary instrument case. The intuition of the baseline LATE argument is not that the always-takers and never-takers disappear, but rather that their contribution is differenced out in the instrument contrast represented by the numerator of the Wald estimand. This differencing occurs in the first term of (26), but not in the second. As a consequence, the second term reflects outcomes for the always- and never-takers, outcomes that inherently do not involve the causal effect of the treatment, because treatment does not vary for these groups. The implication is that the IV estimand is picking up not just the effect of the treatment on outcomes, but also the level of the outcome itself. [Blandhol et al. \(2022\)](#) call this phenomenon level-dependence and show that an estimand that is level-dependent cannot be weakly causal.³²

The size of the second term of (26) is mediated by the instrument residual, \tilde{Z}_i , and in particular by its conditional mean

$$\mathbb{E}[\tilde{Z}_i|X_i] = \mathbb{E}[Z_i|X_i] - X'\delta. \quad (28)$$

This quantity reflects the difference between the conditional mean of Z_i given X_i and the linear regression of Z_i onto X_i . The latter provides the best linear approximation to the conditional mean in terms of mean squared error, so it is only zero if the covariate specification is sufficiently flexible to exactly

³¹ Because \tilde{Z}_i is the residual from a linear regression of Z_i onto X_i , $0 = \mathbb{E}[X_i \tilde{Z}_i] = \mathbb{E}[X_i \mathbb{E}[\tilde{Z}_i|X_i]]$.

³² The intuition is that an estimand that depends on potential outcome levels could have any value—and any sign—even if all of the underlying treatment effects are positive. By contrast, for an estimand that only depends on treatment effects, weak causality is only a matter of whether the weights on these treatment effects are non-negative.

reproduce the conditional mean of the instrument. [Blandhol et al. \(2022\)](#) say that IV specifications with this property have “rich covariates.” They show that IV estimands for specifications that do not have rich covariates will necessarily be level-dependent and will therefore not be weakly causal.

There are two cases in which rich covariates is not a controversial assumption.

The first is when the instrument is independent of the covariates, so that $\mathbb{E}[Z_i|X_i] = \mathbb{E}[Z_i]$ is fit exactly as long as X_i contains a constant. This can happen when the instrument is experimentally assigned, but also in some natural experiments. Fuzzy regression discontinuity designs implemented with local IV estimators (e.g. [Hahn et al., 2001](#); [Calonico et al., 2014](#)) are another example where this occurs, albeit in a limiting sense. While controlling for covariates is not necessary for exogeneity in these settings, researchers still often do so to reduce residual variation in the outcome and/or treatment variables, which can improve statistical precision.

The second case is when the covariate specification is so flexible that it cannot fail to be rich. When discussing linear IV estimands with covariates, [Angrist and Pischke \(2009\)](#) analyze a specification they call “saturate and weight,” which controls for covariates nonparametrically by including an indicator for each covariate value.³³ Controlling for covariates in this way ensures that $\mathbb{E}[\tilde{Z}_i|X_i] = 0$, so that the second term of (26) disappears, and the IV estimand is not level-dependent. However, saturate and weight is a ravenously data-hungry specification, and it tends to produce noisy and poorly-behaved IV estimators. [Blandhol et al. \(2022\)](#) report a survey of IV papers which indicates that saturating in covariates is uncommon in practice.

Outside of these two cases, assuming that a specification has rich covariates is a parametric assumption. The [Blandhol et al. \(2022\)](#) analysis shows that having rich covariates is necessary for an IV estimand to have a weakly causal interpretation, and therefore also necessary for the IV estimand to have some sort of interpretation as a convex average of LATEs when the treatment is binary. Interpreting IV estimands with covariates as reflecting LATEs therefore implicitly requires the parametric assumption that $\mathbb{E}[Z_i|X_i]$ is linear in X_i , a conclusion that is uncomfortably at odds with the motivation of reverse engineering as providing an interpretation that is robust to misspecification. [Blandhol et al. \(2022\)](#) point out that rich covariates can be tested, for example with [Ramsey's \(1969\)](#) RESET test.

3.8.4 Weighting expression for linear IV under rich covariates

If rich covariates holds, then (25) can be written as

$$\beta_{IV} = \mathbb{E} \left[\frac{\mathbb{C}[D_i, Z_i|X_i]}{\mathbb{E}[\mathbb{C}[D_i, Z_i|X_i]]} \frac{\mathbb{C}[Y_i, Z_i|X_i]}{\mathbb{C}[D_i, Z_i|X_i]} \right] = \mathbb{E} \left[\frac{\mathbb{C}[D_i, Z_i|X_i]}{\mathbb{E}[\mathbb{C}[D_i, Z_i|X_i]]} \beta_{IV}(X_i) \right], \quad (29)$$

³³This requires assuming that the covariates are discrete or have been adequately discretized. The saturate and weight specification was originally Theorem 3 in [Angrist and Imbens \(1995\)](#).

where $\beta_{IV}(x)$ is the linear IV estimand with no covariates (other than a constant), but now conditional on the subpopulation with $X_i = x$.³⁴ The interpretation of $\beta_{IV}(x)$ can be considered for each $X_i = x$ subpopulation without concern about linear extrapolation across these subpopulations. If $\beta_{IV}(x)$ is weakly causal for each x , then (29) shows that β_{IV} will be weakly causal if and only if $\mathbb{C}[D_i, Z_i|X_i = x] \geq 0$ for all x . This latter condition says that the sign of the first stage relationship is the same for all $X_i = x$. It has a close connection to the monotonicity condition.

3.8.5 Monotonicity-correct first stage specifications

The monotonicity condition we have been considering so far requires the instrument to operate in the same direction for all individuals. One can weaken this to allow the direction of the effect to depend on each individual's covariates, so that

$$\mathbb{P}[D_i(1) \geq D_i(0)|X_i = x] = 1 \quad \text{or} \quad \mathbb{P}[D_i(0) \geq D_i(1)|X_i = x] = 1 \quad \text{for all } x. \quad (30)$$

[Sloczyński \(2020\)](#) describes (30) as weak monotonicity. Strong monotonicity by contrast is the assumption that we have previously been working with, where the directional effect of the instrument does not change when conditioning on covariates. [Vytlačil's \(2002\)](#) equivalence theorem continues to hold under weak monotonicity if the instrument and covariates are interacted in the threshold-crossing model.

Is weak monotonicity appreciably weaker than strong monotonicity? Many objections to the monotonicity condition are about ruling out heterogeneity in treatment choice due to unobservables, such as preferences or costs. Weak monotonicity doesn't address those concerns. In fact, in judge designs researchers commonly test monotonicity by examining the sign of the first stage relationship across cases with different covariates (see e.g. [Dobbie and Song, 2015](#); [Dobbie et al., 2018](#); [Bhuller et al., 2020](#); [Norris et al., 2021](#)). This exercise can be interpreted as a suggestive test against strong monotonicity, but it wouldn't be appropriate as a test of weak monotonicity, which would allow for these sign changes. This leads us to

³⁴ Multiplying and dividing by $\mathbb{C}[D_i, Z_i|X_i]$ in (29) raises the potential concern that this term may be zero with positive probability. This turns out to not be a concern in the case considered here because $\mathbb{C}[Y_i, Z_i|X_i = x]$ will be zero whenever $\mathbb{C}[D_i, Z_i|X_i = x]$ is zero. To see why, suppose that $\mathbb{C}[D_i, Z_i|X_i = x] = \mathbb{E}[D_i(1) - D_i(0)|X_i = x]\mathbb{V}[Z_i|X_i = x] = 0$. If this is because the first term is zero, then the monotonicity condition implies that $\mathbb{P}[D_i(1) - D_i(0) = 0|X_i = x] = 1$. In this case, full exogeneity implies that $\mathbb{C}[Y_i, Z_i|X_i = x] = \mathbb{E}[(D_i(1) - D_i(0))(Y_i(1) - Y_i(0))|X_i = x]\mathbb{V}[Z_i|X_i = x] = 0$ as well. Alternatively, if $\mathbb{V}[Z_i|X_i = x] = 0$, then $\mathbb{C}[Y_i, Z_i|X_i = x] = \mathbb{C}[D_i, Z_i|X_i = x] = 0$. Instead of indicating for the event that $\mathbb{C}[D_i, Z_i|X_i] \neq 0$ in (29), we just define $\mathbb{C}[D_i, Z_i|X_i = x]\beta_{IV}(x) = 0$ whenever $\mathbb{C}[D_i, Z_i|X_i = x] = 0$ to keep the expression cleaner.

believe that researchers are concerned with failures of strong monotonicity, not weak monotonicity, at least in these contexts. An exception is given by [Mueller-Smith \(2015\)](#), who is explicit in preferring weak monotonicity to strong monotonicity.

[Sloczyński \(2020\)](#) considers the implications of weak and strong monotonicity for interpreting linear IV estimands in a setting with a binary treatment and binary instrument, taking rich covariates as given. He shows that if weak monotonicity holds but strong monotonicity does not, then the linear IV estimand (29) that uses only Z_i as an excluded instrument will reflect negatively-weighted complier treatment effects, and so will fail to be weakly causal. The reason is that when D_i and Z_i are both binary, $\beta_{IV}(x)$ is a LATE for the subgroup with $X_i = x$, while $C[D_i, Z_i | X_i = x]$ is positive if the direction of monotonicity is the first case of (30) and negative if it is the second case. If weak monotonicity holds but strong monotonicity does not, then the sign will be different for some values of x , so compliers for some subpopulations will necessarily be negatively-weighted.

These sign reversals arise because the first stage for β_{IV} has a single excluded instrument Z_i and so is not flexible enough to capture the changes in the direction of monotonicity allowed under weak monotonicity. A 2SLS specification that includes interactions between X_i and Z_i as excluded variables is more flexible in this regard and can restore non-negativity in the weights ([Sloczyński, 2020](#)). [Blandhol et al. \(2022\)](#) extend this concept to 2SLS specifications with non-binary endogenous variables and instruments, describing a specification as “monotonicity-correct” if it is sufficiently flexible to capture covariate-mediated changes in monotonicity. They show that given rich covariates, monotonicity-correctness is the additional sufficient and necessary condition for a 2SLS estimand to be weakly causal.

3.8.6 Specification considerations with covariates

There are two specification considerations when using covariates with linear IV. First, are the covariates rich? If not, then the linear IV estimand reflects potential outcome levels, not just treatment effects, and so is not weakly causal. Second, assuming that covariates are rich, does strong or weak monotonicity hold and, if only weak monotonicity holds, does enough instrument-covariate interactions been included in the first stage to capture changes in monotonicity? If not, then the linear IV estimand reflects some negatively-weighted treatment effects, and so is also not weakly causal.

The saturate and weight specification in [Angrist and Pischke \(2009\)](#) addresses both considerations by including a full set of instrument-covariate interactions in the first stage. Doing so is a dangerous invitation to many instruments bias, the phenomenon whereby an overfit first stage leads to a IV

estimate that is similar to an OLS estimate (e.g. [Bekker, 1994](#)). [Blandhol et al. \(2022\)](#) provide both simulation and empirical evidence that the saturate and weight specification is irredeemably contaminated by many instruments bias in realistic use cases.³⁵

On the other hand, (29) shows that if one is willing to maintain strong monotonicity, then including instrument-covariate interactions in the first stage is usually not necessary for securing a weakly causal interpretation. The sole requirement for the IV estimand to be weakly causal in this case is that it has rich covariates, which is a matter of the flexibility in which covariates are controlled for, not their interactions with the instrument in the first stage. In particular, a specification that is saturated in covariates, but still only uses a single excluded variable Z_i , will generally produce a weakly causal linear IV estimand under strong monotonicity.³⁶ This is the natural counterpart to the [Angrist and Pischke \(2009\)](#) saturate and weight specification, but does not suffer from the pitfalls of many instruments.

Even so, saturating in X_i will often still be too statistically demanding. For these cases, rich covariates must be maintained as a substantive parametric assumption to ensure that linear IV is a weakly causal estimand. A natural alternative is to use flexible machine learning methods to help select the functional form in which the covariates enter, but this takes one outside of the realm of linear IV; see [Section 4.2](#) for more detail.

3.9 Summary of reverse engineering

[Table 3](#) summarizes reverse engineering interpretations for linear IV estimands across the different cases we have considered: the baseline case with a binary treatment and binary instrument, multivalued (or multiple) instruments, multivalued ordered treatments, multivalued unordered treatments, and specifications that control for covariates. The only case in which the linear IV estimand has an unqualified interpretation as a LATE is the first one, the baseline case. The baseline case is commonly the exclusive focus of discussions of the LATE idea, see e.g. textbook discussions by [Wooldridge \(2010, Section 21.4.3\)](#) and [Hansen \(2022b, Section 12.34\)](#), the Nobel lectures by [Angrist \(2022\)](#) and [Imbens \(2022\)](#), or the scientific background provided by the Nobel committee

³⁵ [Blandhol et al. \(2022\)](#) provide simulation evidence that jackknife IV estimators ([Phillips and Hale, 1977](#); [Angrist et al., 1999](#); [Kolesár, 2013](#)) can struggle with the herculean task of correcting the massive many instruments bias introduced by saturate and weight. The limited information maximum likelihood (LIML) estimator, which is also sometimes suggested as a solution for many instruments (e.g. [Bekker, 1994](#); [Hansen et al., 2008](#)), was shown by [Kolesár \(2013\)](#) to generally not be weakly causal.

³⁶ There are some minor caveats here if the instrument is multivalued; see [Blandhol et al. \(2022\)](#) for a precise statement.

TABLE 3 Reverse engineering linear IV estimands.

Treatment (D_j)	Instruments (Z_j)	Covariates (X_j)	Summary
Bin.	Bin.	No	The Wald and simple linear IV estimands are equal to each other and equal to the LATE under monotonicity and full exogeneity.
Bin.	Mul.	No	Each pair of instrument values defines a different complier group with an associated LATE. Different linear IV estimands produce different weighted averages of LATEs, 2SLS with a saturated instrument specification leads to non-negative weights. The weights can be negative with non-saturated specifications, but will be non-negative if the specification reproduces the monotonicity order of the instruments. The monotonicity condition can be especially unattractive if the multivalued instrument is not ordered, for example in judge designs, or when there are multiple instruments.
Ord.	Any	No	If the instrument is binary, and the treatment is a single scalar cardinal variable, then the linear IV estimand can be interpreted as the average causal response (ACR). The ACR can in turn be interpreted either as an average treatment effect among overlapping groups whose treatment choice is shifted by the instrument, or an average per-unit treatment effect across all (disjoint) complier groups. The second interpretation is a natural generalization of the LATE from the binary treatment case. If the instrument is multivalued, then these generalized LATEs get averaged according to different instrument contrasts, the same way as in the binary treatment case, and with the same caveats. Ordered treatments that are not cardinal are better analyzed through the unordered treatment case.

Uno.	Any	No	The linear IV estimand in this case has indicators for each treatment state, except for the excluded state, which is captured by a constant. If there are instruments that affect each treatment state, then the two linear IV estimands will be weakly causal if and only if each instrument affects choices only in its targeted treatment state and the excluded state is always the preferred or next best choice. Achieving this requires strong behavioral restrictions or data on next best choices. With ordered treatments that are not cardinal there are possibilities for restoring a weakly causal interpretation, but they are complicated; see main text.
Any	Any	Yes	Two assumptions are required for a linear IV estimand to be interpretable as a convex weighted average of LATEs: rich covariates and a monotonicity-correct first stage. Rich covariates is often satisfied in randomized experiments, but may not be satisfied when the instrument is not independent of covariates. The first stage will usually be monotonicity-correct under strong monotonicity, but under weak monotonicity it will only be monotonicity-correct if it includes covariates in a way that is flexible enough to account for changes in the direction of monotonicity across covariates.

Notes: This table summarizes the discussion in Section 3.

itself ([Nobel Committee, 2021](#)). But it is rarely the setting in which empirical work using linear IV is actually conducted.³⁷

[Blandhol et al. \(2022\)](#) find that empirical researchers routinely describe their linear IV estimates using LATE language as if they were in the baseline case. The source of this confusion may be due to the way LATE interpretations have been discussed in some influential texts. An early example is the *Handbook of Labor Economics* chapter by [Angrist and Krueger \(1999, pg. 1326\)](#), who wrote.

Finally, we note that the discussion of IV in heterogeneous and non-linear models so far has ignored covariates ... IV estimates in models with covariates can be thought of as producing a weighted average of covariate-specific Wald estimates [conditional LATEs] as long as the model for covariates [satisfies “saturate and weight”]. In other cases it seems reasonable to assume that some sort of approximate weighted average is being generated, but we are unaware of a precise causal interpretation that fits all cases.

The precise (sufficient and necessary) causal interpretation that Angrist and Krueger conjecture about was only recently provided by [Blandhol et al. \(2022\)](#), whose results show that their reasonable assumption is not true (see [Section 3.8](#)). In their hugely influential monograph, [Angrist and Pischke \(2009, pg. 173\)](#) make a similar but less circumspect assertion:

The econometric tool remains 2SLS and the interpretation remains fundamentally similar to the basic LATE result, with a few bells and whistles ... These results provide a simple causal [sic] interpretation for 2SLS in most empirically relevant settings.

³⁷ [Blandhol et al. \(2022\)](#) show that the vast majority of empirical studies using IV control for covariates in a way that suggests they are not just being used to improve precision, while [Mogstad et al. \(2021\)](#) find that over 40 % use multiple instruments. We have not included a discussion of longitudinal settings out of space considerations. The few reverse engineering results on this topic are quite negative, even in simple settings. [Blundell and Dias \(2009, pp. 589–591\)](#) considered difference-in-differences IV strategies in which differential changes over time in the treatment between two groups serves as an instrument. They showed that the corresponding Wald estimand can be interpreted as a LATE for the exposed group if the treatment effect is constant over time and treatment rates do not change in the unexposed group. If they do change in the unexposed group, then the estimand is a weighted average of LATEs between the two groups, with weights that can be negative. This result later appeared in [de Chaisemartin and D'Haultfoeuille \(2018, Theorem 1\)](#), who emphasized the importance and restrictiveness of the time-constant treatment effect assumption. [Miyaji \(2024a,b\)](#) has recently considered the implementation of IV-DID with staggered events through two-way fixed effects, complications that certainly don't make reverse engineering any more attractive ([Bailey and Goodman-Bacon, 2015; Sun and Abraham, 2021](#)). [de Chaisemartin and Lei \(2023\)](#) and [Hahn et al. \(2024\)](#) have shown that it is difficult to ensure a weakly causal interpreted for linear IV estimands used with Bartik instruments.

Sweeping descriptions like these have been repeated more recently, for example in Cunningham's (2021, pg. 351) popular monograph:

The intuition of LATE generalizes to most cases where we have continuous endogenous variables and instruments, and additional control variables, as well.

Statements like these are, at best, only true subject to many unstated caveats.

One reaction to the many caveats of reverse engineered LATE interpretations is to downplay the theory. Does any of this actually matter “in practice?” It’s an interesting question because by their nature reverse engineering arguments are creatures of theory: they do not change practice, they only change interpretation. This probably explains their seductive appeal to busy empirical researchers, and the understandable desire to stretch the theory to fit cases that it does not.

In our view, focusing such intense attention on the reverse-engineered interpretation of a single number like the linear IV estimand makes it more—not less—important for the theory to accurately reflect practice. With forward engineering, the consequences of changing the estimation procedure can be seen directly in the results; the estimates change, but the estimand stays fixed. Reverse engineering, by contrast, cannot be “seen” in the results, but is instead a matter of theoretical justification, the subtle assumptions of which are easy to sweep under the rug in practice.

Reverse engineering can also create problems in other, unexpected places. A clear if mundane example of this was emphasized by Lee (2018), who points out that the usual standard errors for overidentified 2SLS estimators, such as those commonly used with multiple or multivalued instruments, are not correct if there is treatment effect heterogeneity and heteroskedasticity. The classical derivation of these standard error formulas makes use of the assumption that the linear IV model is correctly specified as having constant treatment effects. The derivation omits a term that shows up if the model is misspecified due to UHTE. Imbens and Angrist (1994, Theorem 3) recognized this point, but it seems to have been forgotten in the subsequent empirically-oriented literature, including Angrist and Pischke (2009). Whether this point is substantively important is debatable, and Lee’s empirical illustrations suggest that it may not be. Nevertheless, it provides a clear example of how reverse engineering makes it easy for practitioners to ignore the theory that justifies their practice.

There’s a separate issue of whether these reverse-engineered interpretations—when properly invoked—actually answer useful questions. The baseline LATE result in Section 3.2 is clearly useful in some settings, but this is arguably also not really an example of reverse engineering, as it is based on a nonparametric estimand (the Wald estimand) that simply happens to be equal to a linear IV estimand in the baseline setting. The interpretations with a multivalued ordered treatment and binary instrument in Section 3.6 also seem

useful: with a single binary instrument it is difficult to explore nonlinearity, so a sensibly-averaged summary of the nonlinearity seems like the best one can hope to identify nonparametrically.

For the other cases, it is less clear that these reverse engineering interpretations provide much useful information about causality. An ideal interpretation with convex weights allows one to conclude that the estimand lies somewhere between the smallest and largest average subgroup-specific treatment effects. This conclusion is more informative the less treatment effect heterogeneity there is. With substantial treatment effect heterogeneity—perhaps even effects of different signs—knowing that the weights are convex is not particularly conclusive. The implication is that reverse engineering interpretations work best at providing an interpretation robust to omitted UHTE exactly when this form of misspecification is a lesser concern.

The underlying problem is the form of the weights. For many reverse engineering interpretations of linear IV, the weights reflect statistical features rather than substantive concerns. The clearest example of this is with multivalued instruments, where linear IV estimands were seen to be interpretable as a weighted average of different LATEs, but the weights depended on the marginal distribution of the instrument, as well as on the choice of linear IV estimator. Different experimenters operating under different budgets or making different but sensible choices of evaluation could reach different conclusions in the same economic environment. This is clearly unappealing.³⁸

Changing the weights would lead to more interpretable estimands without these drawbacks. In the multivalued instrument case, weights that were equal or given by the respective complier shares would produce a quantity with a clear interpretation. It would be invariant to the distribution of the instrument and could be defined without reference to a choice of estimator. But for the same reason, such a quantity is unlikely to arise from a reverse engineering mindset as the estimand to some commonly-used estimator: reverse engineering *starts* with the estimator. Estimating target parameters with purposefully-chosen weights requires forward engineering.

4 Forward engineering: estimating target parameters

A forward engineering approach starts with a model and then constructs estimators under the assumption that the model is correctly specified. This is arguably traditional approach taken by economists, with the earliest examples

³⁸ Statistically-driven weights that appear in reverse engineering expressions have also been argued to be objectionable on other grounds. [Słoczyński \(2022\)](#) argues that the weights that appear in reverse-engineered interpretations of the OLS estimand under selection on observables have counterintuitive properties. [Słoczyński \(2020\)](#) argues that the weights that appear in linear IV estimands can be difficult to interpret in some cases. [Balla-Elliott \(2023\)](#) argues that the weights that appear in linear IV estimands are likely to systematically underestimate the causal effects of beliefs in information provision experiments.

being the Gronau-Heckman selection model (Gronau, 1974; Heckman, 1974, 1976). In this section we survey approaches to forward engineering with a eye towards more recent practice.

4.1 Assuming away the problem

The simplest “solution” to the difficulties raised by UHTE is to assume that, in fact, there is no such unobserved heterogeneity. Angrist and Fernández-Val (2013) reintroduce this assumption under the description of “conditional effect ignorability” (CEI). Stated in our notation for the binary treatment, binary instrument case with covariates, their Assumption 3 is that

$$\underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|G_i = g, X_i = x]}_{\text{LATE}(x) \text{ when } g=(0,1)} = \overbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x]}^{\text{ATE}(x) \text{ — the average treatment effect given } X_i=x} \quad (31)$$

for all groups g .

The CEI assumption allows for treatment effect heterogeneity across observable covariates X_i but assumes that the unobservably-different choice groups of always-takers, never-takers, and compliers have the same average treatment effects.

Under the CEI assumption, the ATE is equal to the average of covariate-specific LATEs. This can be seen by applying the law of iterated expectations with (31):

$$\mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}\left[\underbrace{\mathbb{E}[Y_i(1) - Y_i(0)|X_i]}_{\text{ATE}(X_i)}\right] = \overbrace{\mathbb{E}[\text{LATE}(X_i)]}^{\text{by force of CEI assumption (31)}}. \quad (32)$$

Angrist and Fernández-Val (2013) propose estimating covariate-specific LATEs and then averaging them into the ATE, or into the average treatment on the treated/untreated (ATT/ATU), to which similar arguments apply. They call this argument “LATE-Reweighting,” but given the strength of the CEI assumption, one could equally well describe it as “ATE-Reweighting.” Such a reweighting scheme can be implemented in the classical linear model by including treatment-covariate interaction terms and then summarizing the resulting observable treatment effect heterogeneity however one sees fit. Angrist and Fernández-Val (2013) maintain both monotonicity and full exogeneity, but neither are required given the strong CEI assumption.

Assuming away unobserved heterogeneity in treatment effects—whether phrased in the classical model or with CEI—is a strong assumption. Assumptions are necessary for causal inference and strong assumptions may be justified in difficult empirical problems. But it’s important to remember the economic considerations drove the concern about UHTE to begin with (Section 2.2).

In describing why fertility and female labor supply are endogenous—but before introducing the CEI—[Angrist and Fernández-Val \(2013, pg. 406\)](#) write:

Mothers with weak labor force attachment or low earnings potential may be more likely to have children than mothers with strong labor force attachment or high earnings potential.

This description nearly precludes the CEI: mothers with weak labor force attachment may be more likely to have children because the labor supply impacts are different than for mothers with strong labor force attachment. The only way in which the description of endogeneity can coexist with the CEI is if less fertile mothers would still work more than more fertile mothers even in the counterfactual world in which both types of mothers have the same number of children. A story like this rules out UHTE by disconnecting the labor supply and fertility decisions, thereby weakening the very motivation for using an IV to begin with.

4.2 Estimating LATEs and ACRs in the presence of covariates

The difficulties encountered when reverse engineering linear IV estimands with covariates ([Section 3.8](#)) were more mechanical than conceptual. Conceptually, nothing about the LATE identification argument changed, at least for the binary treatment and binary instrument case; conditional-on-covariate LATEs were identified and could be estimated cell-by-cell and then aggregated in whatever way desired. The mechanical problem was with the linear IV estimand, which was not guaranteed to implement a convex weighting without additional implicit (and typically unstated) assumptions. Even if these assumptions were met, the weighting implemented by the linear IV estimand was statistical, making it difficult to interpret the resulting quantity, and difficult to transfer it across settings.

A well-developed econometric literature solves these problems by forward engineering direct estimators of the unconditional LATE. The estimators are primarily designed for the case of a binary treatment and a binary instrument, although they can also be applied with a multivalued ordered treatment, in which case the target parameter is the unconditional ACR. In this section, we discuss two types of estimators and then illustrate their implementation in an empirical example.

4.2.1 Propensity score weighting

A binary instrument allows for a fruitful connection with the large literature on estimation under selection on observables (e.g. [Imbens, 2015](#)). Let

$$W_i(z) \equiv Y_i(D_i(z)) = Y_i(0) + D_i(z)(Y_i(1) - Y_i(0))$$

denote the potential outcome for Y_i associated with a manipulation of the instrument, Z_i , with $Y_i = (1 - Z_i)W_i(0) + Z_iW_i(1)$. Full exogeneity implies that

$W_i(z)$ is independent of Z_i , conditional on X_i . The average treatment effect of Z_i on Y_i —often called the intent to treat (ITT)—is then identified by averaging covariate-conditioned contrasts, assuming an appropriate overlap condition:

$$\underbrace{\mathbb{E}[\mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i]]}_{\text{assuming } 0 < \mathbb{P}[Z_i = 1|X_i] < 1 \text{ (instrument overlap)}} = \overbrace{\mathbb{E}[W_i(1) - W_i(0)]}^{\text{the ATE of } Z_i \text{ on } Y_i \text{ (the ITT)}}. \quad (33)$$

If D_i is binary and exclusion and strong monotonicity are satisfied, then the ITT only reflects treatment effects for the compliers:

$$\begin{aligned} \mathbb{E}[W_i(1) - W_i(0)] &= \mathbb{E}[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \\ &= \text{LATE} \times \mathbb{P}[G_i = \text{CP}], \end{aligned} \quad (34)$$

where $\text{LATE} \equiv \mathbb{E}[Y_i(1) - Y_i(0)|G_i = \text{CP}]$ is the unconditional LATE. The proportion of compliers is itself identified as the average treatment effect of Z_i on D_i , so can be written in an analogous form:

$$\mathbb{E}[\mathbb{E}[D_i|Z_i = 1, X_i] - \mathbb{E}[D_i|Z_i = 0, X_i]] = \mathbb{E}[D_i(1) - D_i(0)] = \mathbb{P}[G_i = \text{CP}]. \quad (35)$$

Putting together (33)–(35) gives

$$\text{LATE} = \frac{\mathbb{E}[W_i(1) - W_i(0)]}{\mathbb{E}[D_i(1) - D_i(0)]} = \frac{\mathbb{E}[\mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i]]}{\mathbb{E}[\mathbb{E}[D_i|Z_i = 1, X_i] - \mathbb{E}[D_i|Z_i = 0, X_i]]}, \quad (36)$$

which is an X_i -averaged reduced form divided by an X_i -averaged first stage.³⁹ A similar argument can be used when D_i is multivalued and ordered, with the change being that the right-hand side of (36) identifies the ACR in (22), rather than the LATE (see Appendix D).

Equation (36) was derived by Tan (2006) and Frölich (2007), who also explicitly made the connection to the problem of estimating the ATE under selection on observables. This connection is helpful because it suggests applying one of the roughly three types of approaches used for that problem: imputation, propensity score weighting, and doubly-robust estimators that combine imputation and weighting. The propensity score referred to here is for the instrument, denoted as $q(x) \equiv \mathbb{P}[Z_i = 1|X_i = x] = \mathbb{E}[Z_i|X_i = x]$ with $Q_i \equiv q(X_i)$. Notice that q was the same object that needed to be correctly specified for a linear IV estimand to satisfy the rich covariates condition necessary for a weakly causal interpretation (Section 3.8.3).

An imputation approach constructs estimators of each of the conditional means in (36) and then averages across the distribution of X_i . Frölich (2007) derived the asymptotic properties of nonparametric series and local polynomial imputation estimators, however the usual curse of dimensionality suggests

³⁹ Similar arguments can be used to construct an expression the average treatment effect for treated or untreated compliers (Hong and Nekipelov, 2010). A treated complier has $Z_i = 1$ while an untreated complier has $Z_i = 0$. They are probabilistically identical groups without covariates, but with covariates they can differ because the distribution of X_i varies conditional on Z_i .

these will tend to perform poorly with more than a couple of covariates. Hirano et al. (2000), Yau and Little (2001), and Tan (2006) considered imputation with parametric estimators. Matching on the instrument propensity score Q_i is another possible imputation approach, which was suggested by Frölich (2007, Section 4), but does not seem to have been pursued further in the literature.

Propensity score weighting has been more widely analyzed. In the context of (36), the appropriate weighting expressions are

$$\mathbb{E}[W_i(1)] = \mathbb{E}\left[\frac{Y_i Z_i}{Q_i}\right] \quad \text{and} \quad \mathbb{E}[W_i(0)] = \mathbb{E}\left[\frac{Y_i(1 - Z_i)}{1 - Q_i}\right], \quad (37)$$

with analogous expressions for $D_i(1)$ and $D_i(0)$. The attraction of propensity score weighting is that only a single function $q(x)$ needs to be modeled and estimated by $\hat{q}(x)$, after which simple sample analogs of the weighting expressions can be formed from $\hat{Q}_i \equiv \hat{q}(X_i)$ and combined to estimate LATE via (36):

$$\frac{\frac{1}{n} \sum_{i=1}^n Y_i Z_i / \hat{Q}_i - \frac{1}{n} \sum_{i=1}^n Y_i(1 - Z_i) / (1 - \hat{Q}_i)}{\frac{1}{n} \sum_{i=1}^n D_i Z_i / \hat{Q}_i - \frac{1}{n} \sum_{i=1}^n D_i(1 - Z_i) / (1 - \hat{Q}_i)}. \quad (38)$$

Weighting estimators like this were proposed by Frölich (2007), Tan (2006), Uysal (2011), MacCurdy et al. (2011), and Donald et al. (2014), and have been revisited more recently by Heiler (2022), Sun and Tan (2022), Singh and Sun (2024), and Słoczyński et al. (2024). The latter authors emphasize the importance of normalizing the weights so that the first term in the numerator of (38) is replaced by

$$\left[\sum_{i=1}^n Y_i Z_i / \hat{Q}_i \right] / \left[\sum_{i=1}^n Z_i / \hat{Q}_i \right], \quad (39)$$

and similarly for the other three terms.

The weighting expressions in (37) follow as special cases from a more general argument developed by Abadie (2003). Abadie (2003) showed that for any function ψ of Y_i , D_i , and X_i ,

$$\begin{aligned} \mathbb{E}[\psi(Y_i, D_i, X_i) | G_i = \text{CP}] &= \frac{1}{\mathbb{P}[G_i = \text{CP}]} \mathbb{E}[\kappa_i \psi(Y_i, D_i, X_i)] \\ \text{where } \kappa_i &\equiv 1 - \frac{D_i(1 - Z_i)}{1 - Q_i} - \frac{(1 - D_i)Z_i}{Q_i}, \end{aligned} \quad (40)$$

a result that has been called “Abadie’s κ ” by Angrist and Pischke (2009).⁴⁰ The most common use of Abadie’s result is to estimate the distribution of X_i among

⁴⁰ To obtain (37) from the more general (40), note that $D_i = Z_i$ for compliers and that

$$\mathbb{E}[W_i(1) - W_i(0)] = \text{LATE} \times \mathbb{P}[G_i = \text{CP}] = \mathbb{E}\left[\frac{Y_i D_i}{Q_i} - \frac{Y_i(1 - D_i)}{1 - Q_i} \mid G_i = \text{CP}\right] \mathbb{P}[G_i = \text{CP}].$$

Applying (40) and simplifying some algebra then produces the difference of the two terms in (37).

compliers by taking $\psi(X_i) = X_i$ (e.g. Marx and Turner, 2019; Leung and O'Leary, 2020; Goodman et al., 2020).

Another application of Abadie's result is to estimate parametric specifications of $\mathbb{E}[Y_i(1) - Y_i(0)|G_i = \text{CP}, X_i = x]$, which (40) shows can be achieved by taking ψ to be an appropriate criterion function. For example, if ψ is taken to be a least squares criterion for a linear regression of Y_i onto D_i and X_i , then minimizing the left-hand side of (40) corresponds to running this regression only among compliers, while minimizing the right-hand side corresponds to running a κ_i -weighted regression among the entire population. The former is of interest because $D_i = Z_i$ is exogenous for the subpopulation of compliers (conditional on X_i), but infeasible because compliers are not observed. The latter is feasible because κ_i can be estimated by substituting \hat{Q}_i for Q_i . Using Abadie's result in this way requires correctly specifying both q and the functional form of the linear controls in the weighted regression. The latter condition ends up being the same as requiring rich covariates (Blandhol et al., 2024). The regression-based approach can also potentially be used to examine heterogeneity in complier treatment effects along observables by including interactions between D_i and X_i ; for an example, see Angrist et al. (2013).

4.2.2 Double robustness and machine learning

Doubly robust approaches combine imputation and propensity score weighting to produce an estimator that is consistent under correct specification of either the propensity score or the conditional means, but not necessarily both. See Kang and Schafer (2007) and Sloczyński and Wooldridge (2018) for overviews in the context of selection on observables. For estimating LATEs, Tan (2006) showed that a doubly-robust estimator of the first term in the numerator of (36) is

$$\frac{1}{n} \sum_{i=1}^n \frac{Z_i}{\hat{q}(X_i)} Y_i - \frac{1}{n} \sum_{i=1}^n \left(\frac{Z_i}{\hat{q}(X_i)} - 1 \right) \hat{\mu}_1(X_i), \quad (41)$$

where \hat{q} is (as before) an estimator of the instrument propensity score q and $\hat{\mu}(x)$ is an estimator of $\mu(x) \equiv \mathbb{E}[Y_i|Z_i = 1, X_i = x]$. Analogous estimators would replace the other terms in (36). Uysal (2011), Ogburn et al. (2015), Sun and Tan (2022), and Sloczyński et al. (2022) analyze various doubly-robust estimators based on (41).

To see what double robustness means, suppose that \hat{q} and $\hat{\mu}$ consistently estimate functions \tilde{q} and $\tilde{\mu}$, so that (41) is consistent for

$$\mathbb{E} \left[\frac{Z_i}{\tilde{q}(X_i)} Y_i \right] - \mathbb{E} \left[\left(\frac{Z_i}{\tilde{q}(X_i)} - 1 \right) \tilde{\mu}_1(X_i) \right]. \quad (42)$$

If $\tilde{q} = q$ is correctly specified, then the first term of (42) is equal to $\mathbb{E}[W_i(1)]$, as shown in (37), while iterating expectations shows that the second term is

zero, regardless of whether $\tilde{\mu}_1 = \mu_1$. On the other hand, if $\tilde{\mu}_1 = \mu_1$, then the second term of (42) satisfies

$$\mathbb{E} \left[\left(\frac{Z_i}{\tilde{q}(X_i)} - 1 \right) \mu_1(X_i) \right] = \mathbb{E} \left[\frac{Z_i}{\tilde{q}(X_i)} Y_i \right] - \mathbb{E} [\mu_1(X_i)], \quad (43)$$

and so again (42) reduces to $\mathbb{E}[\mu_1(X_i)] = \mathbb{E}[W_i(1)]$, this time regardless of whether $\tilde{q} = q$. This is the double robustness property: the estimator (41) converges to (42), which is equal to $\mathbb{E}[W_i(1)]$ if either \hat{q} or $\hat{\mu}$ is consistent for q or μ , but not necessarily both.

The double robustness property gives the researcher two chances at correct specification. Whether this translates into better finite sample performance when both specifications are wrong is debated, see e.g. Kang and Schafer (2007) and the commenting articles, or Sloczyński and Wooldridge (2018) for a review and unifying analysis. For practitioners, doubly robust estimators may be less attractive than propensity score weighting because they require making more modeling choices.

Machine learning (ML) methods can lessen this concern by allowing for data-driven model selection. Belloni et al. (2017) and Chernozhukov et al. (2018) propose a method based on (42) and the other three analogous pieces, combined into a single moment condition. They estimate this moment condition using what they term double/debiased machine learning (DDML), which uses cross-fitting to fit flexible ML estimators to the functions $q(x)$, $\mathbb{E}[Y_i|Z_i = z, X_i = x]$, and $\mathbb{E}[D_i|Z_i = z, X_i = x]$ for $z = 0, 1$. With sufficiently flexible ML estimators of these five functions this procedure can be viewed as providing a nonparametric estimator of the unconditional LATE/ACR in (36). The doubly-robust formulation turns out to be important here as it makes the resulting moment condition Neyman orthogonal, an essential property for ensuring that nonparametric ML methods can be used despite their slower-than-parametric rates of convergence (see, e.g. Newey, 1994; Chernozhukov et al., 2018). Ahrens et al. (2024b) show how multiple ML estimators can be averaged together in DDML using “stacking” to reduce dependency on the approximation properties of any specific estimator.

4.2.3 Empirical illustration

We illustrate some of these methods with the well-known extract of the National Longitudinal Survey of Young Men used by Card (1993) in his analysis of the returns to schooling. The sample size of 3010. The outcome Y_i is log wage in 1976. The treatment D_i is years of education. The instrument Z_i is a binary indicator for living near a four-year college in 1966. Our aim is to estimate the unconditional ACR while accounting for the need to control for covariates X_i . Table 4 compares five estimators across five different sets of covariates. The first two rows report the OLS and linear IV estimators that linearly control for sets of geographic and demographic covariates. Column (4)

TABLE 4 Methods of controlling for covariates in Card's (1993) data.

	(1)	(2)	(3)	(4)	(5)
OLS	0.052 (0.003)	0.040 (0.003)	0.039 (0.003)	0.075 (0.004)	0.073 (0.004)
Linear IV	0.188 (0.026)	0.091 (0.056)	0.092 (0.056)	0.132 (0.054)	0.133 (0.055)
PLIV (DDML)	—	—	0.097 (0.050)	—	0.124 (0.051)
ACR (weighting)	—	0.051 (0.053)	0.041 (0.053)	0.073 (0.047)	0.066 (0.050)
ACR (DDML)	—	—	0.032 (0.033)	—	0.063 (0.046)
Geographic controls		✓	✓	✓	✓
Geographic interactions			✓		✓
Demographic controls				✓	✓
Demographic interactions					✓

Notes: Point estimates and heteroskedasticity-robust standard errors in parentheses. Geographic controls are indicators for region of residence in 1966, residence in an SMSA in 1966 and 1976, and residence in the South in 1966 and 1976. Demographic controls are an indicator for Black, experience, and experience squared. The DDML estimates use an ensemble of ten differently-tuned random forest, gradient boosting, and neural network algorithms, with weights chosen by non-negative least squares through the short-stacking procedure of Ahrens et al. (2024b). DDML estimates are given uninteracted lists of covariates but reported under the with-interactions columns because the methods potentially incorporate interactions on their own. Five folds are used for cross-fitting. The reported point estimate is the median across one hundred replications (different sample splits), with standard errors for the median computed according to Chernozhukov et al. (2018). The propensity score ACR estimates used normalized weights computed with a logit model.

corresponds to the set of covariates used in Card's Table 5, column (3). Columns (2)–(3) only control for geographic covariates, while column (5) augments Card's specification with interactions between geographic and demographic covariates. The linear IV estimator is larger than the OLS estimator throughout all specifications, a finding that is common in the empirical literature, but conflicts with the classic constant effects reasoning about “ability bias” (Card, 2001).

Reverse engineering an interpretation for the linear IV estimator that allows for UHTE is challenging. The rich covariates condition is required for the

corresponding estimand to have an interpretation as a non-negatively weighted average of causal effects (Section 3.8). A [Ramsey \(1969\)](#) RESET test rejects the null of rich covariates in all specifications, with p-values smaller than 10^{-4} . This provides strong statistical evidence that the linear IV estimator cannot be interpreted as weakly causal in this application.

A modern reaction is to use data-driven machine learning techniques to select the functional form of covariates. The third row of [Table 4](#) reports the DDML estimator for the partially linear IV (PLIV) specification discussed by [Chernozhukov et al. \(2018\)](#), which can be implemented using the `ddml` package for Stata or R([Ahrens et al., 2023, 2024a](#)). Three functions are fit in this approach: $\mathbb{E}[Y_i|X_i = x]$, $\mathbb{E}[D_i|X_i = x]$, and $\mathbb{E}[Z_i|X_i = x]$. If the learners used to fit these functions are sufficiently expressive, then the PLIV DDML estimator will converge to the statistically-weighted average of covariate-specific ACRs given in (29), with $\beta_{IV}(x)$ being interpreted as the covariate-specific ACR via conditional versions of (22) or (23). While weakly causal, this object has a convoluted counterfactual interpretation because the weights depend on the joint distribution of X_i and Z_i .

The linear IV and PLIV estimates are close both when using only geographic controls and in Card's specification that includes demographic controls. Comparing the linear IV and PLIV estimates underscores an important point about reverse engineering. Even if the linear IV and PLIV estimates were identical, this would not justify interpreting the linear IV estimate as weakly causal. There are, of course, an infinite number of ways to write the same single number as a weighted average, whether the weights are non-negative or not. The obvious consequence is that two estimates can be similar even if one estimates a weakly causal estimand and the other does not.

The fourth row of [Table 4](#) reports propensity score weighting estimators (38) with the weights normalized as in (39). The only choice needed for this estimator is a model of $q(x) \equiv \mathbb{E}[Z_i|X_i = x]$. We use a logit model with the same covariate specifications as in the corresponding linear estimators. [Słoczyński et al. \(2024\)](#) show how to construct analytical standard errors for the estimator and provide a Stata package for implementation.⁴¹ The fifth row of [Table 4](#) reports DDML estimates of the unconditional ACR, again implemented using the `ddml` package ([Ahrens et al., 2024a](#)). The DDML estimates are computationally intensive to implement and depend on many choices and tuning parameters, which we made only modest attempts to explore.

The weighting and DDML estimates of the unconditional ACR are similar to one another, but substantially smaller than the linear and partially linear estimates, even while the standard errors for all estimates are comparable. The implication is that the difference between an unconditional ACR and a

⁴¹ The package is called `kappalate`. We used our own R code together with bootstrapped standard errors. The analytical standard errors reported by `kappalate` are 10–20 % smaller.

statistically-weighted average of ACRs is considerable in Card's application. Both sets of unconditional ACR estimates are comparable to the OLS estimates, and in some cases even smaller. This provides one answer to Card's (2001) puzzle of why linear IV estimates often exceed their OLS counterparts: the linear IV estimator is estimating an odd statistically-weighted object. Estimates of a more interpretable parameter like the unconditional ACR are not in fact larger than their OLS counterparts.

4.3 Marginal treatment effects

This section contains a development and selected review of marginal treatment effect (MTE) methods for binary treatments. Our focus is on an empirically tractable formulation of the MTE idea that leads to an implementation via linear regression. Surveys on MTE with different emphases are provided by Heckman and Vytlacil (2007b), Cornelissen et al. (2016), and Mogstad and Torgovitsky (2018).

4.3.1 Definitions

MTE methods for binary treatments start with the same underlying assumptions used for identification of LATEs: full exogeneity and monotonicity. Instead of representing these selection assumptions with potential treatments, most authors prefer to use the latent variable notation (14), which we reproduce here, now augmented explicitly with covariates X_i :

$$D_i = \mathbb{1}[V_i \leq \nu(Z_i, X_i)]. \quad (44)$$

The idea is to view V_i as a random variable that captures the unobserved tendency to take treatment and then model the relationship between V_i and $(Y_i(0), Y_i(1))$.

The "marginal" descriptor comes from viewing an individual with $X_i = x$ and $V_i = \nu(z, x)$ as being on the margin between choosing $D_i = 0$ and $D_i = 1$ when faced with an instrument value $Z_i = z$. The average treatment effect for these marginal individuals is $\mathbb{E}[Y_i(1) - Y_i(0)|V_i = \nu(z, x), X_i = x]$. Björklund and Moffitt (1987) appear to have been the first to make use of this interpretation in a Gronau-Heckman normal selection model, but it did not attract much attention until being reintroduced in a nonparametric form by Heckman and Vytlacil (1999, 2005).

Working with (44) is cumbersome because both the function ν and the distribution of V_i are unknown. Yet full exogeneity implies that some features of these unknowns are identified by the propensity score:

$$p(z, x) \equiv \mathbb{P}[D_i = 1|Z_i = z, X_i = x] = \overbrace{\mathbb{P}[V_i \leq \nu(z, x)|X_i = x]}^{\text{by (44) and full exogeneity}} \equiv F_{V|X}(\nu(z, x)|x),$$

the (treatment) propensity score

where $F_{V|X}$ is the distribution of V_i conditional on X_i . The model can be simplified while incorporating this identified relationship by reparameterizing (or "normalizing") the distribution of V_i . The simplest way to do this is to assume

that V_i is continuously distributed and then apply $F_{V|X}$ to both sides of (44), defining a new random variable $U_i \equiv F_{V|X}(V_i|X_i)$.⁴²

$$D_i = \underbrace{\mathbb{1}[F_{V|X}(V_i|X_i) \leq F_{V|X}(\nu(Z_i, X_i)|X_i)]}_{\equiv U_i} \equiv \underbrace{\mathbb{1}[U_i \leq p(Z_i, X_i)]}_{p(Z_i, X_i)}. \quad (45)$$

The distribution of U_i conditional on $X_i = x$ is always uniform on $[0,1]$ for any x , a textbook result known as the probability integral transform (e.g. Hansen, 2022a, pg. 35). Note that this implies that U_i is independent of X_i . However, it's important to remember that U_i is defined as a rank *conditional* on X_i ; comparing U_i across different values of X_i can be misleading.

The normalized selection equation (45) is easier to work with because the distribution of U_i is known and the propensity score is identified. The normalization gives U_i an interpretation of the quantile of resistance to treatment. An individual with $U_i = .05$ is more prone to take treatment than 95 % of the population or, equivalently, less resistant to taking treatment than only 5 % of the population. The Vytlacil equivalence theorem reminds us that these statements should be interpreted as relative to hypothetical variation in the instrument Z_i ; the selection model is a model of how treatment choice varies with Z_i , not X_i , and U_i is defined relative to Z_i . If Z_i is a cost shifter, then those with lower U_i require less cost reduction to take treatment than those with higher U_i . A different Z_i would mean a different model of selection and so a different U_i .⁴³ Even if two different binary instruments have the same propensity scores, their compliers need not reflect the same individuals, and so their U_i 's also need not be comparable.

The MTE is defined as the ATE among subpopulations with the same propensity to take treatment:

$$\text{MTE}(u, x) \equiv \mathbb{E}[Y_i(1) - Y_i(0)|U_i = u, X_i = x].$$

The MTE is a useful definition because it uses the selection model to partition the population based on all unobservable and observable determinants of their treatment choice except for the instrument, which is the source of exogeneous variation.⁴⁴ UHTE is captured through variation in the u component of the MTE, while observed treatment effect heterogeneity is captured through variation in the x component. For modeling purposes it can be advantageous to

⁴² Assuming that V_i is continuously distributed does not change the Vytlacil equivalence theorem; recall Fig. 1.

⁴³ This subtlety is perhaps one benefit of the potential treatment notation, which makes it harder to forget the interpretation of the latent variables being modeled. The latent variable notation makes it tempting to include several instruments without acknowledging the strong implications for choice behavior discussed in Section 3.5 (e.g. Carneiro et al., 2011), or to attempt to port U_i across different environments (e.g. Kowalski, 2023c).

⁴⁴ While usually considered in the context of instruments, the definition of the MTE only depends on (44), which allows for $\nu(Z_i, X_i) = \nu(X_i)$. Briggs et al. (2024) use this observation to consider an MTE analysis based on subjective expectations data rather than instruments.

work with the conditional mean of each treatment arm separately. [Mogstad et al. \(2018\)](#) call these conditional means the marginal treatment response (MTR):

$$\text{MTR}(d|u, x) \equiv \mathbb{E}[Y_i(d)|U_i = u, X_i = x]. \quad (46)$$

Any target parameter that reflects a mean or a mean contrast of potential outcomes can be written as a weighted average of the MTR function.⁴⁵ For example, the average treatment on the treated (ATT) can be written as

$$\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1] = \mathbb{E} \left[\int_0^1 (\text{MTR}(1|u, X_i) - \text{MTR}(0|u, X_i)) \underbrace{\frac{\mathbb{I}_{\{u \leq p(Z_i, X_i)\}}}{\mathbb{P}[D_i = 1]}}_{\omega(1|u, Z_i, X_i)} du \right], \quad (47)$$

where the weights $\omega(d|u, z, x)$ are as indicated. For the ATT, the weights are symmetric in the sense that $\omega(0|u, z, x) = -\omega(1|u, z, x)$, but asymmetric weights can arise for target parameters that reflect only one treatment arm, or both arms but weighted differently. [Table 5](#) reports weights for some of the more commonly considered target parameters. [Appendix E.1](#) briefly discusses how to derive weighting expressions like these. Key to these weighting expressions is that the weights themselves are identified. The MTR function is the sole unknown.

4.3.2 Motivation

So far, these are just definitions. No additional assumptions beyond monotonicity and full exogeneity have been imposed. The purpose of the definitions is to provide a framework under which additional assumptions can be imposed and their identifying content exploited. The additional assumptions are used to construct estimates of a specific target parameter of interest, such as one of the ones listed in [Table 5](#).

The [Ito et al. \(2023\)](#) study of dynamic electricity pricing provides a concrete example of why a researcher may want to do this. The treatment D_i in their setting indicates whether a household adopts dynamic pricing, meaning that instead of paying a fixed rate throughout the day, they pay considerably more during afternoon peak hours and somewhat less during off-peak hours. The instrument Z_i is a binary indicator of whether a household was randomly assigned a \$60 incentive to adopt dynamic pricing. The outcome Y_i is electricity usage.

This example falls into the baseline LATE setting of [Section 3.2](#): a binary treatment and a binary instrument that is unconditionally randomly assigned

⁴⁵ The MTE idea can be extended beyond means as well. [Carneiro and Lee \(2009\)](#) and [Martinez-Iriarte and Sun \(2024\)](#) consider quantiles, while [Acerenza et al. \(2024\)](#) consider duration outcomes with censoring.

TABLE 5 Marginal treatment response weights for common target parameters.

Target parameter	Expression	$\omega(1 u, z, x)$	$\omega(0 u, z, x)$	MTR weights
Average treated outcome	$\mathbb{E}[Y_i(1)]$	1	0	
Average untreated outcome	$\mathbb{E}[Y_i(0)]$	0	1	
Average treatment effect (ATE)	$\mathbb{E}[Y_i(1) - Y_i(0)]$	1	-1	
Conditional ATE	$\mathbb{E}[Y_i(1) - Y_i(0) X_i \in \mathcal{X}]$	$\frac{\mathbb{I}_{\{X_i \in \mathcal{X}\}}}{\mathbb{P}[X_i \in \mathcal{X}]}$	$-\omega(1 u, z, x)$	
Average treatment on the treated (ATT)	$\mathbb{E}[Y_i(1) - Y_i(0) D_i = 1]$	$\frac{\mathbb{I}_{\{u \leq p(z, x)\}}}{\mathbb{P}[D_i = 1]}$	$-\omega(1 u, z, x)$	
Average treatment on the untreated (ATU)	$\mathbb{E}[Y_i(1) - Y_i(0) D_i = 0]$	$\frac{\mathbb{I}_{\{u > p(z, x)\}}}{\mathbb{P}[D_i = 0]}$	$-\omega(1 u, z, x)$	
Generalization of the LATE to $U_i \in [\underline{u}, \bar{u}]$	$\mathbb{E}[Y_i(1) - Y_i(0) U_i \in [\underline{u}, \bar{u}]]$	$\frac{\mathbb{I}_{\{u < z \leq \bar{u}\}}}{\bar{u} - \underline{u}}$	$-\omega(1 u, z, x)$	
Average selection on treatment effects	$\mathbb{E}[Y_i(1) - Y_i(0) D_i = 1] - \mathbb{E}[Y_i(1) - Y_i(0) D_i = 0]$	$\frac{\mathbb{I}_{\{u \leq p(z, x)\}}}{\mathbb{P}[D_i = 1]} - \frac{\mathbb{I}_{\{u > p(z, x)\}}}{\mathbb{P}[D_i = 0]}$	$-\omega(1 u, z, x)$	
Average selection bias	$\mathbb{E}[Y_i(0) D_i = 1] - \mathbb{E}[Y_i(0) D_i = 0]$	$\frac{\mathbb{I}_{\{u \leq p(z, x)\}}}{\mathbb{P}[D_i = 1]} - \frac{\mathbb{I}_{\{u > p(z, x)\}}}{\mathbb{P}[D_i = 0]}$	0	
Policy relevant treatment effect (PRTE)	$\frac{\mathbb{E}[Y_i^o] - \mathbb{E}[Y_i]}{\mathbb{E}[D_i^o] - \mathbb{E}[D_i]}$	$\frac{\mathbb{P}[p^\circ(X_i, Z_i^o) \geq u] - \mathbb{P}[p(X_i, Z_i) \geq u]}{\mathbb{E}[p^\circ(X_i, Z_i^o)] - \mathbb{E}[p(X_i, Z_i)]}$	$-\omega(1 u, x)$	

The weights show how to produce the specified target parameter through the formula target parameter = $\mathbb{E} \left[\int_0^1 \text{MTR}(1|u, X_i) \omega(1|u, Z_i, X_i) du - \int_0^1 \text{MTR}(0|u, X_i) \omega(0|u, Z_i, X_i) du \right]$.

and undoubtedly satisfies the monotonicity condition. The authors estimate the LATE, which provides an evaluation of the effect of an incentive policy matching their experimental policy of a \$60 incentive. But how does it compare to other potential policies?

Answering this question requires understanding which households would be drawn into dynamic pricing under different incentive policies and how dynamic pricing would change their usage. As Ito et al. (2023) discuss, willingness to participate and impact are likely linked: households that can more easily adjust their electricity usage may be both more willing to adopt dynamic pricing and more affected by it. This creates UHTE because “ease of adjustment” is unobserved. The MTE function captures the UHTE by how it changes with u and captures observed heterogeneity by how it changes with x .

The authors’ MTE estimates are reproduced in Fig. 5 which shows the point estimate of $MTR(u, x)$ with x evaluated at the sample mean, along with 95 % confidence intervals. The estimated MTE indicates dramatic UHTE with dynamic pricing having much larger impacts on electricity usage for lower values of u (more willing households) than larger values of u (less willing households). This implies that different incentive policies would also have substantially different impacts. Households with the highest impacts are drawn in by relatively small incentives. Larger incentives draw more household into dynamic pricing, but with less impact on usage. Ito et al. (2023) develop a welfare framework that incorporates these considerations, while also accounting for potential costs of adoption (Eisenhauer et al., 2015). They use the framework to estimate optimal incentive policies.

The Ito et al. (2023) MTE estimates in Fig. 5 rely on additional assumptions about the MTR/MTE function beyond full exogeneity and monotonicity,

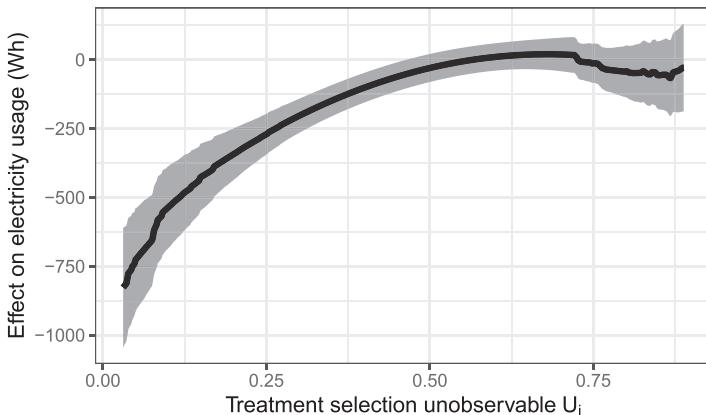


FIG. 5 Marginal treatment effect estimates from Ito et al. (2023). Notes: Authors’ reproduction of Figure 10, Panel A of Ito et al. (2023). We thank Koichiro Ito for providing the necessary data. The point estimate is the estimated MTE evaluated at the sample average of the covariates. The shaded region indicates 95 % bootstrapped confidence intervals.

assumptions which we discuss in detail ahead. Their estimates of the impacts of alternative incentive policies are necessarily less credible than their LATE estimate, because the former rely on strictly stronger assumptions. Yet the motivation of their analysis was precisely to estimate the impact of alternative policies and to characterize a potentially optimal one. Estimating the LATE alone does not speak to this motivation, but an MTE analysis can. The fact that their MTE analysis relies on stronger assumptions is part of the bargain.

4.3.3 A linear regression formulation

In this section, we describe a general but simple linear regression formulation of the MTE idea. Suppose that the MTR functions are linear in parameters, meaning that

$$\text{MTR}(d|u, x) = \sum_{k=1}^{d_\theta} \theta_k b_k(d|u, x), \quad (48)$$

where b_k are known “basis” functions specified by the researcher and θ_k are unknown parameters, collected into a d_θ dimensional vector θ . If the MTR satisfies (48), then the conditional means of the observed outcomes also turn out to be linear in θ . The relationship is

$$\begin{aligned} \mathbb{E}[Y_i|D_i, P_i, X_i] &= \sum_{k=1}^{d_\theta} \theta_k B_{ik} \\ \text{where } B_{ik} &\equiv \left(\frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 b_k(0|u, X_i) du + \frac{D_i}{P_i} \int_0^{P_i} b_k(1|u, X_i) du, \end{aligned} \quad (49)$$

and where $P_i \equiv p(X_i, Z_i)$ is the propensity score evaluated at X_i and Z_i . The derivation of (49) is given in [Appendix E.1](#), but it follows the same type of logic as the weight derivations in [Table 5](#). The regressors B_{ik} are known functions of D_i , P_i , and X_i . The integrals in B_{ik} can be computed analytically for common examples of b_k like polynomials, while numerical integration can be used in other cases.

[Equation \(49\)](#) brings us into standard linear model territory.⁴⁶ Collect the regressors into a vector $B_i \equiv [B_{i1}, \dots, B_{id_\theta}]'$. If the Gram matrix $\mathbb{E}[B_i B']$ is invertible, then θ is identified:

$$\theta = \mathbb{E}[B_i B']^{-1} \mathbb{E}[B_i Y_i]. \quad (50)$$

Because the B_i are functions of the propensity score, P_i , the invertibility of the Gram matrix is a statement about instrument relevance. Whether it holds will depend on the variation in the propensity score and the flexibility of the MTR

⁴⁶ The approach can be viewed as an example of a two-stage “control function” argument. An early example of it can be found in [Heckman and Robb \(1985, Section 3.4\)](#), although not stated in terms of the MTE. [Wooldridge \(2015\)](#) provides a history and exposition of the general idea of a control function. See also [Vella \(1998\)](#) for a discussion on the various control function approaches for one-sided sample selection models, many of which can be seen as progenitors of the approaches discussed ahead.

specification. If θ is identified, then so too is any target parameter that can be written as an identified function of the MTR, such as the quantities in Table 5.

For example, suppose that there are no covariates and that the MTR is specified as linear in u with different parameters for each treatment arm:

$$\text{MTR}(d|u) = \underbrace{\theta_1(1-d) + \theta_2(1-d)u}_{\text{linear in } u \text{ (untreated)}} + \underbrace{\theta_3d + \theta_4du}_{\text{linear in } u \text{ (treated)}}, \quad (51)$$

so that, $b_1(d|u) = (1-d)$, $b_2(d|u) = (1-d)u$, $b_3(d|u) = d$, and $b_4(d|u) = du$. Then (49) becomes

$$\mathbb{E}[Y_i|D_i, P_i] = \theta_1(1-D_i) + \theta_2(1-D_i)\frac{(1+P_i)}{2} + \theta_3D_i + \theta_4D_i\frac{P_i}{2}, \quad (52)$$

which specifies the observed conditional mean as a different linear function of the propensity score for each treatment arm. An alternative way to write (52) is as two separate regressions stratified by treatment arm:

$$\begin{aligned} \mathbb{E}[Y_i|D_i = 0, P_i] &= \theta_1 + \theta_2\frac{(1+P_i)}{2}, \\ \text{and } \mathbb{E}[Y_i|D_i = 1, P_i] &= \theta_3 + \theta_4\frac{P_i}{2}. \end{aligned} \quad (53)$$

From (53), we see clearly what is required for identification of θ : the propensity score must have at least two points of support in each treatment arm. This is satisfied if Z_i is binary and $p(0) \neq p(1)$ with $p(0), p(1) \in (0, 1)$.⁴⁷

The general linear-in-parameters formulation (48) can flexibly accommodate more complex MTR specifications. Covariates can be included for each treatment arm, as can interactions between x and u . Nonlinear functions of u such as higher-order polynomials or splines can also be incorporated, just as in a standard linear model framework. The complexity in the u component is limited by the identification requirement that the Gram matrix be invertible, which is in turn determined by the amount of variation in the propensity score net of covariates. We discuss this further in the next section.

If θ is identified via (50) then it can then be consistently estimated as the coefficients in a linear regression of Y_i on B_i . Some components of B_i will generally depend on the propensity score, $P_i \equiv p(Z_i, X_i)$, as in (52), so to make this regression feasible B_i needs to be replaced by an estimate \hat{B}_i based on an estimated propensity score $\hat{P}_i \equiv \hat{p}(Z_i, X_i)$. The same might also be true of the weights in the target parameter. We discuss this further in Section 4.3.6.

⁴⁷ If $p(z) = 0$ or $p(z) = 1$, then there is one-sided non-compliance. For example, suppose that Z_i is treatment assignment and treatment cannot be obtained without being assigned to it, so that $p(0) = 0$. Then $\mathbb{P}[Z_i = 0|D_i = 1] = 0$, so P_i only has one point of support ($P_i = p(1)$) in the treated arm, and consequently θ_3 and θ_4 are not separately identified. In this case only the compliers are treated, so there is no scope for interpolation or extrapolation among the treated.

4.3.4 Identification

The linear MTR specification (48) was first studied by Brinch et al. (2012, 2017).⁴⁸ As those authors observed, the support of the instrument determines how much linearity can be relaxed without losing point identification. If the instrument has three points of support, and if these three points yield three distinct propensity score points per treatment arm, then an MTR that is quadratic in u is point identified. A cubic is identified with four points of support, and so on.

The traditional Gronau-Heckman normal selection model can also be interpreted as specifying (48). For simplicity, suppose there are no covariates. The normal selection model assumes that $(V_i(d), V_i)$ are bivariate normal with mean zero for $d = 0$ and $d = 1$, where $V_i(d) \equiv Y_i(d) - \mathbb{E}[Y_i(d)]$ and V_i has variance one, where V_i is the pre-normalization selection unobservable. Bivariate normals have linear conditional means, so

$$\begin{aligned} \text{MTR}(d|u) &= \mathbb{E}[Y_i(d)] + \mathbb{E}\left[V_i(d) \mid \frac{U_i = u \Leftrightarrow V_i = F_V^{-1}(u) = \Phi^{-1}(u)}{\underbrace{F_V(V_i)}_{U_i}}\right] \\ &= \mathbb{E}[Y_i(d)] + \mathbb{C}[V_i(d), V_i]\Phi^{-1}(u), \end{aligned} \quad (54)$$

where Φ^{-1} is the inverse of the standard normal cumulative distribution function. This can be written in the linear in parameters form (48) as

$$\begin{aligned} \text{MTR}(d|u) &= \frac{\theta_1}{\mathbb{E}[Y_i(0)](1-d)} \underbrace{\frac{b_1(d|u)}{\mathbb{C}[V_i(0), V_i](1-d)\Phi^{-1}(u)}}_{\theta_2} + \frac{\theta_2}{\mathbb{E}[Y_i(1)]d} \\ &\quad + \underbrace{\frac{\theta_3 b_3(d|u)}{\mathbb{C}[V_i(1) - V_i(0), V_i]d\Phi^{-1}(u)}}_{\theta_4}. \end{aligned} \quad (55)$$

Appendix E.2 shows that computing the integrals in (49) with (55) produces

$$\mathbb{E}[Y_i|D_i, P_i] = \theta_1(1 - D_i) + \theta_2(1 - D_i)\lambda(-\Phi^{-1}(P_i)) + \theta_3D_i - \theta_4D_i\lambda(\Phi^{-1}(P_i)), \quad (56)$$

where $\lambda(v) \equiv \phi(v)/\Phi(v)$ is the inverse Mills' ratio. Equation (56) is recognizable as the Heckman selection correction applied to both treatment arms (e.g. Hansen, 2022b, pg. 883, equation 27.7).⁴⁹ The coefficients θ_3 and θ_1 are average treated and untreated outcomes for the entire population.

If the instrument is binary and there are no covariates, then both the linear and normal specifications lead to saturated regressions: D_i and P_i have four

⁴⁸ See Kowalski (2016, 2023c) for an alternative exposition of the same idea with an application to the impacts of health insurance. Closely-related linear control function assumptions have been used without the context of the MTE by Garen (1984) and Card (1999, 2001); see Wooldridge (2015).

⁴⁹ Most discussions of the Gronau-Heckman normal selection model would set the index function $v(z) = v_1 + v_2z$ to be linear. This would then lead $\Phi^{-1}(P_i) = v_1 + v_2Z_i$, which is the more familiar argument in the inverse Mills' ratio.

points of support and θ has four components. The population fitted values from these regressions will therefore exactly reproduce the conditional means $\mathbb{E}[Y_i|D_i = d, P_i = p]$ for $(d, p) \in \{0, 1\} \times \{p(0), p(1)\}$. As Brinch et al. (2017) observe, this means that for either specification (or any other saturated specification) the LATE implied by the MTR coefficients using the weighting in Table 5 must exactly match the usual LATE, even if the MTR specification is incorrect. See Appendix E.3 for a formal justification and Kline and Walters (2019) for an elaboration.⁵⁰ The equivalence is particular to saturated specifications and generally breaks in unsaturated specifications, which are more typical in practice, for example when including covariates.

While saturated MTR specifications produce the same LATE, they do not generally produce the same value for other parameters. The reason is that other parameters are not nonparametrically identified; they are identified given the MTR specification because the parametric assumptions allow for extrapolation (or interpolation). For example, the ATE weights all values of u equally, so the value of the ATE implied by a given MTR specification depends on how that specification extrapolates and interpolates from the observed propensity score support to other values of u . The extrapolation produced by the linear MTR specification is linear in the quantiles of latent resistance, u , whereas the extrapolation produced by the Gronau-Heckman normal selection model is highly nonlinear and diverges at its extremes because $\Phi^{-1}(u)$ diverges as u tends to zero or one. The relative transparency in how the linear MTR specification extrapolates is a good reason to prefer it over the traditional normal selection model.

Another advantage of the linear MTR specification is the transparency with which it can be made more flexible. The practice of including a quadratic or cubic term is familiar to practitioners from standard linear models. The linear-in-parameters specification also allows for the use of more local specifications, such as splines, which are even more transparent in how they extrapolate. A continuous instrument allows for a nonparametric specification, which for the linear-in-parameters formulation can be interpreted as linear sieve (Chen, 2007). This allows for nonparametric interpolation throughout the support of the propensity score, but does not solve the issue of extrapolation beyond the support.

An important distinction about propensity score variation arises when there are covariates. Suppose that $X_i \in \{0, 1\}$ is binary, for simplicity. A non-separable linear MTR specification interacts the covariate with the linear terms:

$$\text{MTR}(d|u, x) = \underbrace{(1 - d)(\theta_1 + \theta_2 u + \theta_3 x + \theta_4 ux)}_{\text{untreated arm}} + \underbrace{d(\theta_5 + \theta_6 u + \theta_7 x + \theta_8 ux)}_{\text{treated arm}}. \quad (57)$$

⁵⁰The result for the normal selection model is also implicit in Baker and Lindeman (1994, Appendix I).

Compared to (51), this specification now allows the slope of the linear-in- u MTR to be different for $x = 0$ and $x = 1$. It implies that the conditional mean of Y_i given $D_i = d$ and $X_i = x$ is linear in P_i for each of the four combinations of d and x . Identification requires $p(1, x) \neq p(0, x)$ for each x , so that the instrument is relevant conditional on x .

A separable linear MTR specification sets θ_4 and θ_8 to be zero in (57). This requires the slope of the MTR to be the same for $x = 0$ and $x = 1$ while still allowing the level to be different. In this case, identification only requires $p(1, x) \neq p(0, x)$ for either $x = 0$ or $x = 1$. To see this, consider the implied conditional mean of the observed outcome for the treated arm:

$$\mathbb{E}[Y_i|D_i = 1, P_i, X_i] = \theta_5 + \theta_6 \frac{P_i}{2} + \theta_7 X_i. \quad (58)$$

The coefficients are identified as long as P_i and X_i are not perfectly correlated (given $D_i = 1$). If we write the propensity score out in saturated form, so that

$$P_i = \pi_1 + \pi_2 Z_i + \pi_3 X_i + \pi_4 Z_i X_i \quad (59)$$

we can see that P_i and X_i will not be perfectly correlated if Z_i and X_i aren't and if either $\pi_2 \neq 0$ or $\pi_4 \neq 0$, which is the same as $p(0, x) \neq p(0, x)$ for either $x = 0$ or $x = 1$.

This discussion is related to the well-known critique about the Gronau-Heckman model being identified even without a relevant instrument (e.g. Goldberger, 1983; Puhani, 2000). The critique is really about the model of the propensity score: if both $\pi_2 = 0$ and $\pi_4 = 0$ in (59), so that Z_i is irrelevant, then P_i is perfectly correlated with X_i , so that θ_6 and θ_7 are not separately identified in (58). On the other hand, suppose that we start with an unsaturated probit model for the propensity score, so that

$$p(x, z) = \Phi(\pi_1 + \pi_2 Z_i + \pi_3 X_i), \quad (60)$$

where Φ is the standard normal cumulative distribution function. Now even if $\pi_2 = 0$, $P_i = \Phi(\pi_1 + \pi_3 X_i)$ is not perfectly correlated with X_i , because Φ is a nonlinear function. This shows that the critique about not needing an instrument is not related to selection modeling per se, but rather to the fact that statistical models commonly used for the propensity score of a binary treatment are nonlinear. Viewed from this perspective, this classic critique of selection modeling seems much less damning.

Separable specifications require less instrument variation. The flip side of this statement is that separable specifications can be more flexible in u than the variation in the instrument alone would suggest. Brinch et al. (2017) show that a separable MTR that is quadratic in u is generally identified with a binary instrument and binary covariate:

$$\text{MTR}(d|u, x) = \underbrace{(1 - d)(\theta_1 + \theta_2 u + \theta_3 x + \theta_4 u^2)}_{\text{untreated arm}} + \underbrace{d(\theta_5 + \theta_6 u + \theta_7 x + \theta_8 u^2)}_{\text{treated arm}}. \quad (61)$$

The intuition can be seen in the linear separable conditional mean (58) and the linear (saturated) propensity score specification (59). There are two excluded variables in (59)— Z_i and $Z_i X_i$ —but only one “endogenous” variable $P_i/2$ in (58).⁵¹ So there’s room to include an additional endogenous variable in (58) by adding the u^2 term in (61). Nonlinear propensity score specifications like (60) are typically used in practice, and these effectively create interactions between all values of the covariates and the instrument because $p(1, x) - p(0, x) = \Phi(\pi_1 + \pi_2 + \pi_3 x) - \Phi(\pi_1 + \pi_3 x)$ differs for all values of x .

The Ito et al. (2023) estimates in Fig. 5 are based on a separable specification with a binary instrument. The authors provide evidence that takeup differs heavily by baseline household characteristics, in particular a measure of expected savings from switching to dynamic pricing given historical usage. The incentive has an impact throughout the distribution of household characteristics, leading to wide variation in the propensity score. This is what allows the authors to identify and estimate a flexible MTE curve across a wide range of u . The cost is the separability assumption, which requires the pattern of UHTE to not depend on observables.

Is the cost worth it? Should we be extrapolating at all? Certain target parameters, such as the ATE, depend on the MTE at extreme quantiles of u , and so require some extrapolation whenever these extreme quantiles are not represented in the propensity score, which is common. So whether extrapolation is necessary is a matter of the research question and how much variation there is in the data. The Ito et al. (2023) study provides a good example: the authors observed one incentive policy, but the purpose of their analysis was to compare different incentive policies and estimate an optimal one. There is no way to do this without interpolating and extrapolating.

The central role of extrapolation in IV methods with UHTE means that different specifications should be compared for sensitivity. This is already common practice in applications of MTE. Partial identification analysis provides a formal way to incorporate specification ambiguity by allowing for models that are too rich to generate a single value of the target parameter. Mogstad et al. (2018) develop a partial identification approach for MTE analysis. For applications of this approach see Mogstad et al. (2017), Rose and Shem-Tov (2021), Gulotty and Yu (2023), and Daljord et al. (2023).⁵² A major

⁵¹ The scare quotes are because $P_i/2$ is a function of Z_i and X_i , so not really endogenous. However, it arises in (58) because the MTR depends on u , and the dependence between U_i and $(Y_i(0), Y_i(1))$ is the source of endogeneity.

⁵² Han and Yang (2024) and Marx (2024) consider partial identification approaches that exploit the full independence between the instrument and potential outcomes, rather than the mean-independence considered here and in Mogstad et al. (2018). Kowalski (2023b,a) considers special cases that arise with a binary treatment and binary instrument, derives closed-form bounds that are easy to implement, and applies the bounds to study mammography and the overdiagnosis of breast cancer.

benefit of considering partial identification is that it allows one to harness the identifying content of nonparametric shape restrictions, such as monotonicity or concavity, which often have a clear economic interpretation. The major challenge with partial identification is computation, estimation, and inference; see [Canay and Shaikh \(2017\)](#) and [Molinari \(2020\)](#) for general discussions, and [Shea and Torgovitsky \(2023\)](#) for a discussion in the context of MTE methods.

4.3.5 Unstratified regressions and local instrumental variables

The implied conditional mean for Y_i considered in (49) is stratified in the sense that it conditions on the treatment indicator, D_i . [Heckman and Vytlacil \(2007b\)](#), Section 4.8 call this the selection or control function approach, while [Brinch et al. \(2017\)](#) call it the “separate” approach. An alternative is an unstratified regression where the conditioning on D_i is dropped and the coarser conditional mean of Y_i given only P_i and X_i is used instead. This produces the relationship

$$\mathbb{E}[Y_i|P_i, X_i] = \mathbb{E}[Y_i(0)|X_i] + \underbrace{\int_0^{P_i} \text{MTE}(u, X_i) du}_{\mathbb{E}[D_i(Y_i(1) - Y_i(0))|P_i, X_i]; \text{ see Appendix E.1}}. \quad (62)$$

[Heckman and Vytlacil \(2007b\)](#), Section 4.8 describe approaches based on (62) as “IV approaches” in contrast to control function approaches. We adopt the terminology stratified for (49) and unstratified for (62) because both use the variation in P_i produced by Z_i . The two approaches are more similar than they are different.

[Heckman and Vytlacil \(1999\)](#) observed that the derivative of the unstratified regression (62) identifies the MTE:

$$\text{LIV}(u, x) \equiv \underbrace{\frac{\partial}{\partial u} \mathbb{E}[Y_i|P_i = u, X_i = x]}_{\text{local instrumental variable}} = \text{MTE}(u, x). \quad (63)$$

They describe this derivative as the local instrumental variable (LIV) estimand due to its interpretation as a limiting case of the usual reduced-form-to-first-stage ratio in traditional linear IV estimands. If $u = p(z, x)$ is set to be an observed propensity score value, then

$$\text{LIV}(p(z, x), x) \approx \frac{\mathbb{E}[Y_i|p(Z_i, X_i) = p(z', x), X_i = x] - \mathbb{E}[Y_i|p(Z_i, X_i) = p(z, x), X_i = x]}{\frac{\mathbb{E}[D_i|Z_i = z', X_i = x]}{p(z', x)} - \frac{\mathbb{E}[D_i|Z_i = z, X_i = x]}{p(z, x)}},$$

where the approximation is for $p(z', x) \approx p(z, x)$. Because it is a derivative, LIV(u, x) is only well-defined if there is continuous variation in P_i around u , conditional on $X_i = x$, which requires continuous variation in Z_i . Assuming such variation is available, (63) shows that the MTE is also identified at that point, a relationship that can be used to view the MTE at specific points of evaluation as limiting versions of the LATE. [Carneiro et al. \(2011\)](#) develop a

semiparametric local polynomial estimator of the LIV using [Robinson's \(1988\)](#) approach for partially linear models; see [Cornelissen et al. \(2016\)](#) and [Andresen \(2018\)](#) for more details on implementation.

Continuous instrument variation is a luxury that is not available in many IV applications. Continuous covariate variation can be used as a substitute if the MTE is assumed to be separable so that $MTE(u, x) = m_U(u) + m_X(x)$ for two functions m_U and m_X . Estimation based on the LIV requires one or the other, so may not be applicable or attractive in many situations.

Alternatively, one can start with the unstratified regression (62) and use a linear-in-parameters specification:

$$\mathbb{E}[Y_i(0)|X_i = x] = x'\vartheta_0 \quad \text{and} \quad MTE(u, x) = \sum_{k=1}^{d_\vartheta} \vartheta_k b_k(u, x), \quad (64)$$

where b_k are known basis functions specified by the researcher and ϑ_k are unknown parameters. Notice that in contrast to (48), which parameterized the two treatment arms in the MTR separately, now we are parameterizing their difference—the MTE—as well as the baseline covariate relationship in the untreated state. Substituting these forms into (62) produces

$$\mathbb{E}[Y_i|P_i, X_i] = X'\vartheta_0 + \sum_{k=1}^{d_\vartheta} \vartheta_k B_k \quad \text{where } B_k \equiv \int_0^P b_k(u, X_i) du. \quad (65)$$

Identification is again a matter of whether the Gram matrix for this linear regression is invertible, which requires having sufficient variation in the propensity score P_i , controlling for X_i .

Compared to the stratified regression (49), the unstratified regression (65) exploits less of the observed variation in the data, because it does not condition on D_i . An implication is that more instrument variation is needed for identification when considering comparable specifications. For example, suppose that there are no covariates and assume that the MTE is linear in u , so that

$$\mathbb{E}[Y_i(0)] = \vartheta_0 \quad \text{and} \quad MTE(u) = \vartheta_1 + \vartheta_2 u.$$

A linear MTE is implied by the linear MTR in (51). In principle, it is a weaker parameterization, although as a practical matter it is probably not substantively different.⁵³ However, when substituted into (65), the linear MTE produces a regression that is quadratic in P_i :

$$\mathbb{E}[Y_i|P_i, X_i] = \vartheta_0 + \vartheta_1 P_i + \vartheta_2 \frac{P_i^2}{2}. \quad (66)$$

⁵³ For example suppose that we specify the MTR as $MTR(d|u, x) = \theta_1(1 - d) + \theta_2(1 - d)u + \theta_3d + \theta_4du + \theta_5u^2$, where the quadratic term does not depend on d . The implied MTE is then $(\theta_3 - \theta_1) + (\theta_4 - \theta_2)u$, which is linear in u .

This is in contrast to (52), which was linear in the propensity score, but stratified by treatment arm. Whereas a binary instrument was sufficient for invertibility with the stratified regression, three points of instrument support are needed for the comparable unstratified approach.

4.3.6 Estimation and inference

Estimating θ in the stratified regression requires first estimating the treatment propensity score to obtain estimates $\hat{P}_i \equiv \hat{p}(Z_i, X_i)$ of P_i . Typically one would use a logit or probit model for this purpose so that the \hat{P}_i lie between 0 and 1, but a linear model could also be used. Replacing P_i with \hat{P}_i in the definition of B_{ik} gives an estimate \hat{B}_{ik} of B_{ik} , collected into a vector \hat{B}_i . Then θ can be estimated with ordinary least squares:

$$\hat{\theta} = \left(\sum_{i=1}^n \hat{B}_i \hat{B}' \right)^{-1} \left(\sum_{i=1}^n \hat{B}_i Y_i \right). \quad (67)$$

Estimating ϑ in the unstratified regression (62) proceeds the same way, except that B_{ik} and \hat{B}_{ik} are defined differently and baseline covariates X_i are included additively, so that the regression is of Y_i on X_i and \hat{B}_i .

The parameters θ of the MTR (or ϑ or the MTE) are usually not of ultimate interest; instead we are interested in the target parameter that can be constructed from the MTR (or MTE). Suppose in particular that the target parameter takes the form

$$\tau = \sum_{d \in \{0,1\}} \mathbb{E} \left[\int_0^1 \text{MTR}(d|u, X_i) \omega^*(d|u, Z_i, X_i) du \right], \quad (68)$$

where the weights ω^* are assumed to be identified, but may need to be estimated. Substituting (48) into this expression gives

$$\tau = \sum_{k=1}^K \theta_k b_k^* \quad \text{where } b_k^* \equiv \mathbb{E} \left[\sum_{d \in \{0,1\}} \int_0^1 b_k(d|u, X_i) \omega^*(d|u, Z_i, X_i) du \right]. \quad (69)$$

Each of the b_k^* are identified but need to be estimated if ω^* needs to be estimated, or if b_k or ω^* depend on X_i or Z_i . A natural estimator of \hat{b}_k^* is

$$\hat{b}_k^* \equiv \frac{1}{n} \sum_{i=1}^n \sum_{d \in \{0,1\}} \int_0^1 b_k(d|u, X_i) \hat{\omega}^*(d|u, Z_i, X_i) du, \quad (70)$$

where $\hat{\omega}^*$ is an estimator of ω^* . For example, if the target parameter is the ATT, then

$$\hat{\omega}^*(d|u, Z_i, X_i) = \frac{\mathbb{1}[u \leq \hat{P}_i]}{n^{-1} \sum_{j=1}^n \hat{P}_j}. \quad (71)$$

The final estimate of the target parameter τ is then

$$\hat{\tau} \equiv \sum_{k=1}^{d_\theta} \hat{\theta}_k \hat{b}_k^*. \quad (72)$$

Estimating a target parameter with an unstratified regression uses exactly the same procedure, just that τ must only depend on the MTE, and not the two arms of the MTR separately.

Computing $\hat{\theta}$ and $\hat{\tau}$ requires calculating the integrals in the definitions of B_i and b^* . The `mtefe` package for Stata (Andresen, 2018) and the `ivmte` package for R (Shea and Torgovitsky, 2023) both contain functionality that automates this task.⁵⁴ Given the integration, computation is simply a matter of one logistic or other binary response regression to estimate the propensity score p and one linear regression to estimate the MTR parameters θ . The `ivmte` package for R also contains functionality for implementing the partial identification approach developed by Mogstad et al. (2018), but estimation and inference is more complicated.

The formal asymptotic theory for $\hat{\theta}$ and $\hat{\tau}$ does not appear to have been worked out yet.⁵⁵ There is no reason to expect that these estimators would not be asymptotically normal, but their asymptotic variances will be complicated by the fact that \hat{B}_{ik} and \hat{b}_k^* are estimated in a first step. Simulation evidence by Andresen (2018) bears out the normality and suggests that naive variance calculations that ignore first step estimation error may still be roughly accurate. Bootstrapping the entire procedure will account for this first step estimation error and is easy to do. This already appears to be standard practice among empirical practitioners. All aspects of the estimation procedure are smooth, so there is no reason to expect that the bootstrap would not be consistent (Fang and Santos, 2019), at least assuming that the instruments are sufficiently strong.⁵⁶

These statements apply equally to the stratified and unstratified regressions. Which should be used? One consideration is the amount of instrument variation available, as less is required for the stratified regressions. Another consideration is the target parameter: a stratified approach estimates the MTE, and so will not provide enough information to compute target parameters that depend on the MTR components themselves. This comes up in the Ito et al. (2023) study, where the target parameter considered depends asymmetrically on electricity usage under dynamic and static pricing, which create different marginal surpluses. Assuming that enough variation is available to consider both and that the target parameter depends only on the MTE, the only remaining consideration is presumably statistical precision. Andresen (2018) provides some simulation evidence that suggests the stratified approach tends to lead to more precise estimates. On balance then, stratified MTR approaches seem preferable, although more research on the statistical differences would be useful.

⁵⁴ The `mtefe` package improves on the earlier `margte` package by Brave and Walstrum (2014).

⁵⁵ Carneiro and Lee (2009) and Sasaki and Ura (2023) have derived formal results for semiparametric approaches that are not linear-in-parameters.

⁵⁶ As with linear IV estimators, weak instruments could make the asymptotic approximation a poor description of the finite sample behavior, and would also render the bootstrap inconsistent (e.g. Andrews et al., 2019).

4.3.7 Applications and uses of marginal treatment effects

Applications of MTE methods are widespread and have been proliferating rapidly in the past fifteen years. Table 6 provides a list of some empirical applications of MTE methods. All of these applications use the MTE to investigate patterns of UHTE. These patterns add depth and nuance to the empirical analysis and can sometimes speak to questions about mechanisms.

Another complementary use of MTE methods is for estimating the impacts of explicit policy counterfactuals. The Ito et al. (2023) study is one example. Another example is given by Cornelissen et al. (2018), who study the impact of publicly provided childcare on childrens' outcomes, and then use their MTE estimates to simulate the aggregate impacts of expanding publicly provided childcare. A third example is Mogstad et al. (2017), who use MTE estimates to compare the cost effectiveness of different subsidies for encouraging the use of mosquito nets.

In each case, the authors measure the policy counterfactual using what Heckman and Vytlacil (2001a, 2005) called policy-relevant treatment effects (PRTEs). A PRTE is defined by a hypothetical modification of the instrument and/or propensity score from p and Z_i to p° and Z_i° . This results in a different hypothetical treatment selection,

$$D_i^\circ \equiv \mathbb{1}[U_i \leq p^\circ(Z_i^\circ, X_i)],$$

and so also different realized outcomes,

$$Y_i^\circ \equiv (1 - D_i^\circ)Y_i(0) + D_i^\circ Y_i(1).$$

Let $p^{\circ\circ}$ and $Z_i^{\circ\circ}$ denote some other policy that leads to treatment choices $D_i^{\circ\circ}$ and realized outcomes $Y_i^{\circ\circ}$. Then the PRTE for these two policies is defined as

$$\text{PRTE} \equiv \frac{\mathbb{E}[Y_i^\circ] - \mathbb{E}[Y_i^{\circ\circ}]}{\mathbb{E}[D_i^\circ] - \mathbb{E}[D_i^{\circ\circ}]},$$

which gives the average change in outcomes per net change in treatment participation. An alternative definition omits the denominator and just measures the average change in outcomes (Heckman and Vytlacil, 2001a; Carneiro et al., 2010).⁵⁷ The contrasting policy is often taken to be the baseline status quo, $p^{\circ\circ} = p$, $Z_i^{\circ\circ} = Z_i$, so that $D_i^{\circ\circ} = D_i$ and $Y_i^{\circ\circ} = Y_i$ are the observed treatment and outcomes.

Constructing p° and Z_i° may require some extrapolation or speculation. Cornelissen et al. (2018) consider a policy that takes $p^\circ(Z_i, X_i) = \min\{1.5p(Z_i, X_i), 1\}$ and so increases the likelihood of attendance for every child by one and a half times. As the authors point out, it is not clear what type

⁵⁷ Carneiro et al. (2010) also define a limiting version of the PRTE for small policy changes, which they call the marginal PRTE or MPRTE. The advantage of the MPRTE is that it doesn't require any extrapolation, and so in principle can be estimated nonparametrically.

TABLE 6 Empirical applications of marginal treatment effects.

Labor and human capital	Moffitt (2008), Carneiro et al. (2011), Kaufmann (2014), Carneiro et al. (2016), Joensen and Nielsen (2016), Nybom (2017), Dal Bó et al. (2021), De Groot and Declercq (2021), Gathmann et al. (2021), Heinesen and Stenholz Lange (2022), Westphal et al. (2022), Dutz et al. (2022), Hurnlum et al. (2023)
Development	Mogstad et al. (2017), Bandiera et al. (2020), Berry et al. (2020), Manda et al. (2020), Li et al. (2021a), Mellon Bedi et al. (2021), Sarr et al. (2021)
Health	Basu et al. (2007), Johar and Maruyama (2014), Basu et al. (2014), Alessie et al. (2020), Depalo (2020), Gong et al. (2020), Zeng et al. (2020), Kowalski (2023c,a), Wilding et al. (2023), Gupta et al. (2024)
Family and childhood development	Doyle (2007), Brinch et al. (2017), Cornelissen et al. (2018), Felfe and Lalivé (2018), Priebé (2020), Hojman and Lopez Boo (2022), Liu et al. (2022)
Crime	Doyle (2008), Arnold et al. (2018), Bhuller et al. (2020), Arnold et al. (2022), Baron and Gross (2022), Arbour (2022), Arteaga (2023), Agan et al. (2023), Posebom (2023), Gonçalves and Mello (2023)
Public programs	Maestas et al. (2013), French and Song (2014), Moffitt (2019), Moffitt and Zahn (2019), Aizawa et al. (2023)
Energy	Wang et al. (2020), Li et al. (2021b), Ito et al. (2023)
Other	Galasso et al. (2013) (innovation), Kasahara et al. (2016) (international trade), Daljord et al. (2023) (marketing ⁽²⁾), Coury et al. (2022) (history), Heldring et al. (2022) (history)

Notes: A list of papers that have applied MTE methods to empirical problems. We limit the list to papers that fit the binary treatment setting in Section 4.3.

of concrete intervention would achieve this new level of attendance. Another counterfactual policy intervention the authors consider is a direct increase in their instrument, the number of available childcare seats per capita, from Z_i to $Z_i^* = Z_i + .4$. Here the intervention is more clear, but the impact that this has on attendance depends on $p(Z_i^*, X_i)$, which involves some extrapolation beyond the observed support of Z_i . Two types of extrapolation (or interpolation) are required for evaluating a PRTE of this sort: the effect of changing the instrument on treatment takeup, and the effect of treatment on outcomes for those induced to change their treatment status under the new policy.

The definition of the PRTE is premised on the fundamentals of the environment remaining stable under the new policy, an assumption that [Heckman and Vytlacil \(2005\)](#) describe as policy invariance. The need for policy invariance is not specific to MTE methods; it's a necessity for any sort of counterfactual policy analysis. In the context of the MTE, policy invariance means that the distribution of $(Y_i(0), Y_i(1), U_i)$ remains the same under different policies. There are certainly good reasons to be skeptical of such an assumption. For example, if the childcare expansion entertained by [Cornelissen et al. \(2018\)](#) is achieved by adding poorer quality childcare facilities, then the treatment effect $Y_i(1) - Y_i(0)$ could change, leading to a failure of policy invariance. Again, this is not a drawback of MTE methods or even IV strategies more generally, but rather an inherent limitation of evaluating a policy with a model that doesn't model all possible effects of the policy.

4.4 Binary treatments when monotonicity is violated

The monotonicity condition plays a central role when estimating both LATEs and MTEs. Yet as we saw in [Section 3.4](#), it can be unattractive in some settings, such as in judge designs. Multiple instruments are also difficult to square with the traditional monotonicity condition ([Section 3.5](#)). What can be done in these cases?

The simplest solution is to redefine the instrument in a way that makes the monotonicity condition more plausible. For example, instead of using all judges individually, [Dahl et al. \(2014\)](#) consider estimates based on a binary instrument that defines whether the judge is one of the most strict or one of the least strict, with moderate judges being omitted from the analysis. Monotonicity violations that might occur among a variety of similar judges are perhaps less likely to occur when comparing extreme judges, making the monotonicity condition with the binary instrument more plausible. This point was recently recycled by [Sigstad \(2024b,a\)](#). Binarizing the instrument also makes it easier to assess the impact that violations of monotonicity would have through a sensitivity analysis like the one in [Section 3.4](#).

The same idea can also be applied to multiple instruments. The problem in that case was the difficulty in comparing treatment choice behavior under pairs of instrument values that were not ordered in a natural way, such as

$(Z_{i1}, Z_{i2}) = (0, 1)$ and $(Z_{i1}, Z_{i2}) = (1, 0)$. One solution is to remove these instrument values and only consider $(Z_{i1}, Z_{i2}) = (0, 0)$ and $(1, 1)$, which can be naturally ordered if both Z_{i1} and Z_{i2} are incentives to take treatment. Versions of this idea have been used by Frölich (2007), Goff (2024), and van't Hoff et al. (2024). An alternative is to use only one component of the instrument at a time, conditioning on the rest of the components as covariates. Monotonicity in each instrument separately allows for the estimation of separate LATEs and separate MTE curves, one for each instrument; see Mogstad et al. (2021) for an empirical illustration. Mogstad et al. (2024) show how MTE curves for different instruments can be aggregated in a partial identification framework.

Recording the instrument or conditioning on subcomponents are simple solutions, but they reduce the amount of effective exogenous variation. A more ambitious approach is to design a new selection model that allows for deviations from monotonicity. The major obstacle is identification. With a binary treatment, monotonicity enables identification of the shares of each choice group. Removing monotonicity requires adding a new assumption or allowing for the possibility of partial identification.

Gautier and Hoderlein (2015) and Gautier (2021) consider random coefficient versions of the threshold-crossing model (44) and show that point identification can be obtained under extreme large support assumptions on the available of instrument variation. Ura and Zhang (2024) and Han and Kaido (2024) provide partial identification approaches that are applicable to models that do not satisfy monotonicity, such as a random coefficients model, but the approaches come with some of the familiar computational and statistical challenges of partial identification. Dutz et al. (2022) develop a simple non-monotonic model of survey response and apply it under conditions that lead to either point or partial identification.

Arnold et al. (2022, Section 4) point out that MTE-style regressions of outcomes on propensity scores can still be estimated even if monotonicity does not hold. Assuming that these regressions are correctly parameterized, they can still be extrapolated to estimate unconditional potential outcome means. Their approach effectively replaces low-level behavioral assumptions about monotonicity with higher-level statistical assumptions about the relationship between outcomes and the propensity score. Arnold et al. (2022, Section 5) develop a parametric selection model that does not impose monotonicity and show how to estimate the model, but do not establish identification.

Measurement error in the treatment provides another source of monotonicity violations. Even if monotonicity is satisfied for the correctly measured (but latent) binary treatment variable, misclassification will mean that it is violated for the observed, mismeasured treatment variable. Ura (2018), Calvi et al. (2022), and Tommasi and Zhang (2024) consider identification of the LATE in the presence of this type of measurement error, while Possebom (2023), Acerenza et al. (2023), and Acerenza (2024) consider identification of the MTE. Partial identification emerges in all of these analyses, with the exception of Calvi et al. (2022).

4.5 Ordered treatments

The linear MTE approach for binary treatments extends to ordered treatments with one important caveat: the natural generalization of the threshold-crossing model (44) is no longer equivalent to the monotonicity condition.

4.5.1 Threshold-crossing with multiple treatments

Suppose as in Section 3.6 that the treatment variable takes values d_0, d_1, \dots, d_J arranged in increasing order. Instead of (44), assume that

$$\mathbb{1}[D_i \geq d_j] = \mathbb{1}[V_i \leq \nu(d_j | Z_i, X_i)] \quad \text{for each } j. \quad (73)$$

The function ν now depends on d_j , but the unobservable V_i is the same for all d_j , an important distinction that we will return to ahead. Because D_i is no smaller than d_0 we can set $\nu(d_0 | Z_i, X_i) = +\infty$. It's also convenient to add an artificial value $d_{J+1} > d_J$ with $\nu(d_{J+1} | Z_i, X_i) = -\infty$ to reflect that D_i must always be strictly smaller than d_{J+1} .

Normalizing (73) as in the binary case simplifies it to

$$\mathbb{1}[D_i \geq d_j] = \mathbb{1}\left[\underbrace{F_{V|X}(V_i | X_i)}_{\equiv U_i} \leq \underbrace{F_{V|X}(\nu(d_j | Z_i, X_i) | X_i)}_{p(d_j | Z_i, X_i)}\right] \equiv \mathbb{1}[U_i \leq p(d_j | Z_i, X_i)], \quad (74)$$

where $p(d_j | z, x) \equiv \mathbb{P}[D_i \geq d_j | Z_i = z, X_i = x]$ is a generalization of the propensity score (the “greater than” propensity score or the conditional survival function when viewed as a function of d_j). As in the binary case, $p(d_j | z, x)$ is identified. In the multivalued case, it is decreasing in d_j , with $p(d_0 | z, x) = 1$ and $p(d_{J+1} | z, x) = 0$.

Writing (74) in terms of individual levels makes it look a bit more familiar:

$$D_i = d_0 + \sum_{j=1}^J (d_j - d_0) \underbrace{\mathbb{1}[p(d_{j+1} | Z_i, X_i) < U_i \leq p(d_j | Z_i, X_i)]}_{\mathbb{1}[D_i < d_{j+1} \text{ and } D_i \geq d_j] = \mathbb{1}[D_i = d_j]}. \quad (75)$$

This is an ordered response model (e.g. Greene and Hensher, 2009; Wooldridge, 2010, Chapter 16). The binary threshold-crossing model (5) is recovered by setting $J = 1$, $d_0 = 0$, and $d_1 = 1$, so that $p(d_0 | z, x) = 1$, $p(d_1 | z, x) = \mathbb{P}[D_i = 1 | z, x]$ is the usual binary propensity score, and $p(d_2 | z, x) = 0$.

In the multivalued case, the ordered response model is no longer equivalent to the monotonicity condition. This was shown by Vytlacil (2006) in a follow-up to Vytlacil (2002). The reason can be seen with three values ($J = 2$), a binary instrument, no covariates, and with monotonicity in the direction $D_i(1) \geq D_i(0)$. Monotonicity allows for the choice groups $G_i = (d_0, d_2)$ and $G_i = (d_1, d_1)$ to both exist. In terms of (75), the first group would consist of those individuals with U_i strictly larger than $p(d_1 | 0)$, giving $D_i(0) = d_0$, and weakly smaller than $p(d_2 | 1)$, giving $D_i(1) = d_2$. So for this first group to exist, it must be that $p(d_1 | 0) < p(d_2 | 1)$. On the other hand, the second group consists of

those values of U_i that lie in the intervals $(p(d_2|z), p(d_1|z)]$ for both $z = 0$ and $z = 1$. But if the first group exists, these two intervals must be disjoint:

$$\underbrace{p(d_2|0) < p(d_1|0)}_{U_i \text{ here if } G_i(0)=d_1} \underset{\text{if } G_i=(d_0,d_2) \text{ exists}}{\sim} \underbrace{p(d_2|1) < p(d_1|1)}_{U_i \text{ here if } G_i(1)=d_1}. \quad (76)$$

Intuitively, the ordered response model restricts how much treatment can respond to the instrument: in this case it can either respond a lot ($G_i = (d_0, d_2)$) or not at all ($G_i = (d_1, d_1)$), but not both.

This finding makes sense in the context of reverse engineering linear IV with ordered treatments (Section 3.6). The weights in the Angrist and Imbens (1995) ACR were identified, but only because they combined (“double counted”) multiple choice groups. A counting exercise reveals that the shares of each of the choice groups cannot be point identified. With $J = 2$ and a binary instrument, there are six choice groups consistent with monotonicity: G_i can be (d_0, d_0) , (d_0, d_1) , (d_0, d_2) , (d_1, d_1) , (d_1, d_2) , or (d_2, d_2) . Yet there are only four independent choice probabilities: $\mathbb{P}[D_i = d|Z_i = z]$ for $d = d_0, d_1$, and $z = 0, 1$, with the probability for $D_i = d_2$ being implied by the sum-to-one constraint. Five independent choice group probabilities cannot be uniquely pinned down by four independent choice probabilities. The ordered threshold model (75) effectively rules out an additional choice group a priori, restoring point identification of the group shares.

Is this additional restriction attractive? Vytlacil (2006) shows that it’s not inherent to latent variable notation or even to a threshold-crossing structure. Vytlacil (2006) extends his equivalence result from the binary case to a more flexible class of ordered response models with thresholds that vary according to additional latent variables, which he shows is again equivalent to the monotonicity condition stated with potential choices notation. The number of latent variables in these models makes it clear that they are not point identified without additional distributional structure or extreme assumptions on the available instrument variation (Cunha et al., 2007). We are left with a familiar trilemma: (i) use a selection model that admits point identification but makes potentially restrictive behavioral assumptions; (ii) relax the behavioral assumptions but impose additional parametric structure; or (iii) allow for partial identification. Most empirical applications of forward engineered with ordered treatments have taken the first option and used (75) as the selection model.⁵⁸

⁵⁸ An example of an application of the third option is Goldin et al. (2021), who maintain the usual monotonicity condition. Their partial identification argument is further developed in Vohra and Goldin (2024). Kamat et al. (2024) develop and apply a partial identification approach under a different model of ordered choice that they describe as “latent monotonicity.” Their approach uses a linear-in-parameters framework similar to the one discussed ahead, but without the benefit of a point identified selection model. Arteaga (2023) uses the same latent monotonicity model with additional assumptions that effectively return the problem back to a binary treatment case, to which she then applies MTE methods.

4.5.2 A linear regression formulation

Using (75) as the selection model makes it possible to directly adapt the linear regression formulation of MTE from the binary case.⁵⁹ The marginal treatment response (MTR) is defined as before, except now there are more values of d for its first argument. We again assume that the MTR function has the linear-in-parameters form (48). For notation, let $P_i(d_j) \equiv p(d_j|Z_i, X_i)$ be the d_j -specific greater-than propensity score and collect these scores into $P_i \equiv (P_i(d_1), \dots, P_i(d_J))$. Then (49) can be generalized to

$$\begin{aligned} \mathbb{E}[Y_i|D_i, P_i, X_i] &= \sum_{k=1}^{d_\theta} \theta_k B_{ik} \\ \text{where } B_{ik} &\equiv \sum_{j=0}^J \mathbb{1}[D_i = d_j] \left(\frac{1}{P_i(d_j) - P_i(d_{j+1})} \int_{P_i(d_{j+1})}^{P_i(d_j)} b_k(d_j|u, X_i) du \right). \end{aligned} \quad (77)$$

First step estimates of $P_i(d_j)$ can be constructed by estimating an ordered response model and then used to construct estimates \hat{B}_{ik} of B_{ik} . At that point the story becomes the same as in the binary treatment case: a linear regression of Y_i onto \hat{B}_{ik} to estimate the θ_k 's, which can then be used to estimate a variety of target parameters.

The primary difference with the binary treatment case is that it might also be attractive to parameterize the d dimension of the MTR function when D_i has a cardinal interpretation. For example, the linear specification (51) can be extended so that each treatment value d has its own linear-in- u function, leading to $d_\theta = 2(J + 1)$ parameters. These can be point identified if the binary instrument satisfies $p(d_j|0) \neq p(d_j|1)$ for $j = 1, \dots, J$. A more parsimonious specification could be to interact the levels of d and u , so that there are only four basis functions $(1, d, u, du)$ with four parameters to estimate. Many other types of parameterizations are possible. Point identification is a matter of whether the resulting Gram matrix formed from the B_{ik} variables is invertible, which requires the $P_i(d_j)$ scores to vary sufficiently with Z_i and/or X_i .

Rose and Shem-Tov (2021) apply a linear-in-parameters ordered treatment MTR analysis in their study of the effect of incarceration on recidivism.⁶⁰ The authors' treatment is incarceration measured in months, which can be expected to have an impact on recidivism that is both nonlinear over length and heterogeneous across individuals. The authors use discontinuities in sentencing guidelines as instruments to produce linear IV estimates, but they correctly recognize that the interpretation of these estimates is opaque with both a

⁵⁹ Heckman et al. (2006) and Heckman and Vytlacil (2007b) provide the earliest discussions phrased in terms of local instrumental variable estimands.

⁶⁰ Other recent applications are Cornelissen et al. (2018) and Rivera (2023).

multivalued treatment and a multivalued instrument (not to mention covariates). Even if weakly causal, the linear IV estimates smear UHTE and nonlinearities into a single hard-to-interpret number. The authors provide suggestive evidence of both UHTE and nonlinearities by showing that their linear IV estimates change substantially when using different instruments or when including nonlinear treatment terms.

[Fig. 6](#) reproduces a central empirical finding from [Rose and Shem-Tov \(2021, Figure 6\)](#). The outcome is an indicator for any reincarceration within five years of sentencing. The authors assume that the MTR function is a fifth degree polynomial in u that is additively separable between u and x , but make no assumptions about how the MTR varies across d . Because the MTR is so flexible in u , the authors proceed as if it (and any target parameters generated from it) are potentially only partially identified and extend the partial identification framework of [Mogstad et al. \(2018\)](#) to the ordered treatment case. While some of their bound estimates are wide, many are quite narrow, including the ones in [Fig. 6](#), which are essentially point estimates.

The left panel of [Fig. 6](#) plots estimates of the counterfactual probability of recidivism (being reincarcerated) if counterfactually not incarcerated, conditional on sentence length. The strong upward trend indicates strong selection patterns: judges assign longer sentences to individuals more likely to offend if not incarcerated. The right panel shows that the effects of a hypothetical two year sentence are large and negative. The effect of the two year sentence increases dramatically with an individual's actual sentence length, signaling

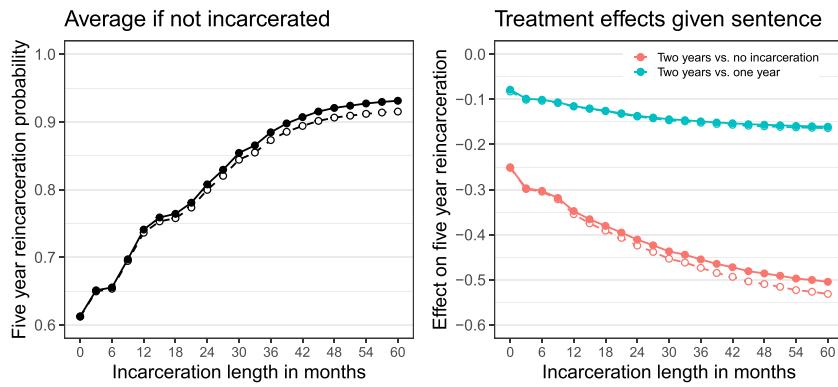


FIG. 6 Marginal treatment effect estimates from [Rose and Shem-Tov \(2021\)](#). Notes: Authors' reproduction of Figure 6 of [Rose and Shem-Tov \(2021\)](#). We thank Evan Rose for providing the necessary data. The solid lines are upper bound estimates and the dotted lines are lower bound estimates. The left-hand panel reports estimates of $\mathbb{E}[Y_i(0)|D_i = d]$, where Y_i is an indicator for any reincarceration within five years and D_i is incarceration length in months. The right-hand panel reports estimates of $\mathbb{E}[Y_i(24) - Y_i(0)|D_i = d]$ in red and $\mathbb{E}[Y_i(24) - Y_i(12)|D_i = d]$ in blue.

important treatment effect heterogeneity. On top of the heterogeneity, the right panel shows evidence of nonlinearity, with the effect being driven by the first year of the hypothetical two year sentence.

Rose and Shem-Tov (2021) point out that accounting for the type of heterogeneity and nonlinearity visible in Fig. 6 is important when considering sentencing policy. They use their estimates to consider the impacts of a budget-neutral change in sentencing that increases the rate of incarceration while reducing the length of longer sentences. As expected from Fig. 6, they find that this type of reallocation can produce large reductions in the rate and duration of reincarceration.

Linear IV estimates are not up to the task of evaluating this type of nuanced policy counterfactual. If viewed as correctly-specified, a linear IV estimate for a specification that is linear in treatment mechanically produces no aggregate effects from reallocating sentence lengths. A reverse engineering interpretation allows the coefficient on the single linear treatment variable to be interpreted as a non-negatively weighted average across different treatment intensities (Section 3.6), but knowing the components in this weighted average is what's needed for evaluating a reallocation of sentence lengths. Adding in a nonlinear treatment term to the linear IV specification puts one in a reverse engineering no-mans land of no known results other than to assume constant effects. But the assumption of constant effects is both implausible a priori and strongly at odds with the right-hand panel of Fig. 6. Ignoring treatment effect heterogeneity would overstate the benefit from reducing sentence lengths by not accounting for the higher effect of incarceration on more severe offenders.

4.5.3 Continuous treatments

In some cases it might be reasonable to think of the treatment as being ordered and continuous. The model analogous to (75) for a continuous treatment can be written as

$$D_i = \nu(Z_i, X_i, V_i), \quad (78)$$

where ν is an unknown function that is assumed to be strictly decreasing in V_i , which is still a single scalar unobservable, and still assumed to satisfy full exogeneity together with the potential outcomes. It is again possible to normalize V_i by replacing it with $U_i \equiv F_{V|X}(V_i|X_i)$, a transformation that can be absorbed into the definition of ν (Matzkin, 2003). Upon doing so, we see that ν is identified by the survival function

$$p(d|Z_i, X_i) \equiv \mathbb{P}[D_i \geq d|Z_i, X_i] = \mathbb{P}[U_i \leq \nu^{-1}(Z_i, X_i, d)|Z_i, X_i] = \nu^{-1}(Z_i, X_i, d), \quad (79)$$

where ν^{-1} is the inverse of ν in its V_i argument. The selection model is then

$$D_i = p^{-1}(U_i|Z_i, X_i) = \nu(Z_i, X_i, U_i), \quad (80)$$

which is just a function of a scalar uniform random variable and point identified objects, as before. The MTR $m(d|u, x)$ is defined the same way as before, except now the first argument can take a continuum of values.

There is an econometrics literature that analyzes models of continuous treatments together with a selection equation like (80).⁶¹ [Imbens and Newey \(2009\)](#) show that if no assumptions are placed on the MTR, then extreme instrument variation is necessary for point identification of $\mathbb{E}[Y_i(d)] = \mathbb{E}[\text{MTR}(d|U_i, X_i)]$. [Torgovitsky \(2015, 2017\)](#) shows that the strong assumption of rank invariance across different potential outcomes (e.g. [Heckman et al., 1997](#)) enables nonparametric point identification of average and quantile potential outcomes even with only a binary instrument. Viewed in terms of the MTR function, these results are fully nonparametric in d , u , and x . They represent polar cases that are likely to be unattractive for most applications.

Imposing some parametric assumptions seems reasonable. [Masten and Torgovitsky \(2016\)](#) show that if

$$\text{MTR}(d|u, x) = \rho_0(u) + \rho_1(u)d + \rho_2(u)'x, \quad (81)$$

then the functions $\rho_0(u)$, $\rho_1(u)$, and $\rho_2(u)$ can be identified for all u with only a binary instrument, implying identification of the entire MTR function. [Masten and Torgovitsky \(2014\)](#) analyze a kernel-based linear regression estimator, which can be implemented with the Stata command `ivcrc` ([Benson et al., 2022](#)); see [Gollin and Udry \(2021\)](#) and [Carrillo et al. \(2023\)](#) for empirical applications. [Florens et al. \(2008\)](#) consider a r th degree polynomial specification that omits covariates but includes an additional unknown function of d :

$$\text{MTR}(d|u) = \rho_0(u) + \rho_1(u)d + \cdots + \rho_r(u)d^r + \bar{\rho}(d). \quad (82)$$

Their identification results require continuous instrument variation and they do not consider estimation.

[Chernozhukov et al. \(2020\)](#) and [Newey and Stouli \(2021\)](#) extend the analysis of both [Florens et al. \(2008\)](#) and [Masten and Torgovitsky \(2016\)](#) to allow for more general specifications of the MTR, as well as quantile counterparts. These include the linear-in-parameters form of the MTR (48), which should be particularly attractive and flexible for applications. If we let $P_i \equiv p(D_i|Z_i, X_i)$, which is equal to U_i by (80), then

$$\mathbb{E}[Y_i|D_i, P_i, X_i] = \sum_{k=1}^{d_0} \theta_k B_{ik} \quad \text{where } B_{ik} \equiv b_k(D_i|P_i, X_i). \quad (83)$$

⁶¹ The selection equation is traditionally assumed to be strictly increasing in U_i rather than strictly decreasing, but this is not material. What matters is the invertibility created by strict monotonicity. The literature also typically uses nonseparable models for Y_i rather than potential outcomes; we have translated the notation and findings to the concept of an MTR so as to stick with the potential outcomes notation.

Implementation proceeds as before: estimate p , now using distribution or quantile regression (e.g. Chernozhukov et al., 2013) to construct \hat{B}_{ik} , then regress Y_i onto \hat{B}_{ik} to estimate the θ_k 's. Relative to (81) and (82), the linear-in-parameters specification allows for parameterizations in both the u and d dimensions. This can be used to lessen the demands on instrument variation.

4.5.4 Selection models that do not allow for heterogeneity

The selection equations (44), (75), and (78) for the binary, ordered discrete, and continuous cases all have a single unobservable, but one that enters the equation non-additively. This is important. As we noted, for multivalued cases the selection models are not equivalent to the monotonicity condition, which generally requires additional latent variables (Vytlačil, 2006). However, they do still allow for unobserved heterogeneity in how the instrument affects treatment choice.

In contrast, Heckman and Vytlačil (1998) and Wooldridge (1997, 2003, 2008) consider linear and additive selection models like

$$D_i = \nu_0 + \nu_1 Z_i + V_i, \quad (84)$$

where V_i is assumed to be mean-independent of Z_i . The authors show that under this condition and the linear-in-treatment specification (81), the linear IV estimand is equal to $\mathbb{E}[\rho_1(U_i)]$, and so the average partial effect $\mathbb{E}[\text{MTR}(d|U_i) - \text{MTR}(d'|U_i)]$ is identified for any pairs d and d' .

This result comes at a high cost. While (84) may look like the usual statistical first stage regression (7), the regression only ensures that V_i and Z_i are orthogonal, which is weaker than mean-independence. Imposing mean-independence requires thinking of (84) as a selection model that describes choices under counterfactual manipulation of the instrument. It is a particularly restrictive selection model because it implies that the effect of Z_i on D_i is ν_1 —constant for all individuals—so that there is no unobserved heterogeneity in the effect of the instrument on the treatment. This seems like an unacceptably asymmetric assumption given the motivating goal of allowing for UHTE in treatment effects.

4.6 Unordered treatments

The major obstacle in extending these ideas to unordered treatments is identification of the selection equation. Suppose that D_i takes one of J unordered values d_0, d_1, \dots, d_J , and that selection follows the discrete choice model

$$D_i = \arg \max_{d_j \in \{d_0, d_1, \dots, d_J\}} \nu(d_j | Z_i, X_i) + U_{ij}, \quad (85)$$

where ν is again an unknown function and U_{ij} are unobservables, with $\nu(d_0 | z, x) = 0$ and $U_{i0} = 0$ as a normalization. Relative to the binary and ordered cases, there are now multiple unobservables, one for each choice,

which we collect as the vector $U_i \equiv (U_{i1}, \dots, U_{iJ})$. The familiar econometric interpretation of (85) views the arguments of the argmax as indirect utilities for choosing option d_j , with the observed choice D_i being the one with the highest utility. These indirect utilities can differ with the instrument, the observed covariates, and the unobservables.

The definition of the MTR extends immediately with the change that now u is a vector, not a scalar. We can still consider a linear-in-parameters specification like (48). As one example, [Kline and Walters \(2016\)](#) consider a case with three choices ($J = 2$) and assume that

$$\text{MTR}(d_j | \underbrace{u_1, u_2}_u, x) = \rho_0(d_j)'x + \rho_1(d_j)u_1 + \rho_2(d_j)u_2, \quad (86)$$

where ρ_j are unknown coefficients that are different for each treatment state d_j . This MTR specification can be written in the linear basis form with nine components by including indicators for each treatment state. Where things become difficult is the following step: what does (86) imply about the conditional mean of the observed outcome? It can still be related to the MTR via

$$\mathbb{E}[Y_i(d) | D_i = d_j, Z_i = z, X_i] = \mathbb{E}[\text{MTR}(d_j | U_i, x) | \underbrace{U_i \in \mathcal{U}^*(d_j | z, X_i)}_{\text{set of } U_i \text{ for which } d_j \text{ is optimal}}, X_i], \quad (87)$$

where $\mathcal{U}^*(d_j | z, x)$ is the subset of U_i for which d_j is the maximizer of (85). But evaluating this expression further requires knowing something about the distribution of U_i , even if the MTR function is assumed to have a linear-in-parameters form.

This problem is the same issue that arose for binary treatments without monotonicity ([Section 4.4](#)). Counterfactual choice probabilities—let alone the distribution of U_i —are not point identified in a traditional discrete choice model like (85) without parametric assumptions or extreme instrument variation ([Tebaldi et al., 2023](#)). This is in contrast to the models for ordered choice considered in the previous section, all of which admitted nonparametric point identification of choice probabilities. The uniform normalization that produced U_i came from folding the unknown distribution of the original latent variable V_i into the definition of the MTR, permitting focus on a single unknown object. These nonparametric simplifications are not available for the unordered case, at least not with a model like (85).

One path is to embrace the need for parameterization and leverage insights from the well-developed literature on discrete choice (e.g. [Train, 2009](#)). A pioneering early example is [Dubin and McFadden \(1984\)](#), who used a multinomial logit model for the selection equation and also made an assumption like (86) to derive a linear-in-parameters expression for the conditional mean.⁶² See

⁶² This is again an example of a control function approach ([Heckman and Robb, 1985; Vella, 1998; Wooldridge, 2015](#)), in contrast to fully parameterizing the entire model and basing estimation off of the likelihood function, which requires stronger assumptions. See [Geweke et al. \(2003\)](#) for an example of this type of approach for unordered treatments.

[Abdulkadiroğlu et al. \(2020\)](#) for a recent application of their approach. [Kline and Walters \(2016\)](#) replaced the logit with a multinomial probit that allows for correlation between U_{ij} and U_{ik} . They derived the resulting expression for the observed conditional outcome mean when the MTR is given by (86). The result looks like a multivariate generalization of the inverse Mills' ratio expression (56). [Kline and Walters \(2016\)](#) show that if there is only a binary instrument that affects the utility of one option, then additional variation in choice probabilities due to covariates is needed for identification if the MTR is separable between x and u , as in (86). [Hull \(2020\)](#) uses a similar approach to estimate hospital quality.

These types of parametric distributional assumptions may of course be unattractive, or at least not sufficiently easy to flexibly modify. [Dahl \(2002\)](#) replaced the explicit parametric distributional assumptions with higher-level index-sufficiency assumptions. [Heckman and Pinto \(2018\)](#), [Lee and Salanié \(2023\)](#), and [Navjeevan et al. \(2023\)](#) consider selection models for unordered treatments that are more restrictive than (85), which can lead to point identification of average treatment effects for certain choice groups either nonparametrically or with a little bit of added parametric structure ([Pinto, 2022](#)).⁶³ [Lee and Salanié \(2023\)](#) and [Kamat \(2024\)](#) consider partial identification. All of these approaches are relatively tailored and may require a fair amount of case-specific work to be applied.

As usual, extreme amounts of instrument variation can solve these difficulties. [Heckman et al. \(2008\)](#) showed that instruments that drive choice probabilities to one effectively reduce the problem back to the binary treatment setting. [Lee and Salanié \(2018\)](#) showed that natural generalizations of the local instrumental variable argument extend to the unordered treatment setting for a variety of choice models *if* these models are point identified, which typically requires extreme instrument variation (or parametric assumptions). Also as usual, instruments with extreme variation don't exist in practice.

[Mountjoy \(2022\)](#) shows how continuous but local (not extreme) instrument variation can be used nonparametrically with unordered treatments. Mountjoy's argument is based on having choice-specific instruments in (85).⁶⁴ In his application with $J = 2$ representing the choice of two-year or four-year college, this means that $\nu(d_1|Z_{i1}, Z_{i2}, X_i) = \nu(d_1|Z_{i1}, X_i)$ and $\nu(d_2|Z_{i1}, Z_{i2}, X_i) = \nu(d_2|Z_{i2}, X_i)$. [Mountjoy \(2022\)](#) argues that this can be satisfied with separate two- and four-year distance instruments. He then shows that marginal shifts in these instruments can be used to separately identify marginal treatment effects

⁶³ See [Navjeevan et al. \(2023\)](#) and [Xie \(2024\)](#) for estimation methods that incorporate covariates gracefully.

⁶⁴ Mountjoy shows that the full structure of (85) is not necessary, although it nicely captures the key elements. [Heckman and Pinto \(2018\)](#) and [Loeser \(2023\)](#) provide equivalence results relating discrete choice models of treatment selection to restrictions on their implied potential choices.

of d_2 relative to d_0 and of d_4 relative to d_0 for those on the margin of indifference between these choices. Identifying the ν function or the distribution of U_i can be bypassed because of the assumption that each instrument affects only one choice. [Humphries et al. \(2023b\)](#) show that if this assumption is dropped, then one can still apply Mountjoy's argument by estimating ν and then using the estimated $\nu(d_j|Z_i, X_i)$ as themselves choice-specific instruments.

Unordered choice problems can also arise out of dynamic or multistage decision problems. Versions of the above ideas have been applied to these settings as well. For examples, see [Heckman et al. \(2016, 2018\)](#), [Walters \(2018\)](#), and [Humphries et al. \(2023a\)](#).

4.7 No selection model

All of the reverse and forward engineering approaches for allowing UHTE discussed so far have maintained a selection model. This is certainly not innocuous. The behavioral assumptions imposed by a selection model, such as the monotonicity condition, can sometimes be unattractive. The full exogeneity condition that is invariably imposed with a selection model places strong requirements on the instrument, as we noted in [Section 2.5](#). In this section, we consider an approach due to [Manski \(1990\)](#) that allows for UHTE but does not impose a selection model.

4.7.1 Manski-Robins and IV intersection bounds

Suppose that outcome exogeneity is satisfied, so that $\mathbb{E}[Y_i(d)|Z_i = z] = \mathbb{E}[Y_i(d)]$ for all d and z . Assume that the treatment is binary, for simplicity.⁶⁵ Then

$$\begin{aligned} \mathbb{E}[Y_i(1)] &= \underbrace{\mathbb{E}[Y_i(1)|Z_i = z]}_{\text{outcome exogeneity}} = \underbrace{\mathbb{E}[Y_i|D_i = 1, Z_i = z]p(z)}_{\text{identified}} \\ &\quad + \underbrace{\mathbb{E}[Y_i(1)|D_i = 0, Z_i = z]}_{\text{not directly identified}}(1 - p(z)). \end{aligned}$$

The only term on the right-hand side that is not identified by the data is the counterfactual treated mean for those in the untreated state. Assume that it lies in $[y_{lb}, y_{ub}]$. This assumption could be based either on the logical support of Y_i , in which case it's not restrictive, or it could be based on substantive restrictions about a reasonable range for the conditional mean of Y_i . Substituting these bounds for the unidentified counterfactual gives upper and lower bounds on

⁶⁵The following argument applies equally well for non-binary treatments, but the bounds will tend to be wider.

$\mathbb{E}[Y_i(1)]$ that depend on z . The tightest bounds are then found by taking the largest lower bound and the smallest upper bound across z :

$$\begin{aligned}\mathbb{E}[Y_i(1)] &\in \max_z \mathbb{E}[Y_i|D_i = 1, Z_i = z]p(z) + y_{lb}(1 - p(z)), \\ &\min_z \mathbb{E}[Y_i|D_i = 1, Z_i = z]p(z) + y_{ub}(1 - p(z)).\end{aligned}\tag{88}$$

A symmetric set of bounds can be derived for the untreated mean, $\mathbb{E}[Y_i(0)]$. Bounds for the ATE are then formed by taking the difference of the bounds for the potential outcome means.

This argument was first considered without an instrument (Z_i deterministic) by [Manski \(1989\)](#) and [Robins \(1989\)](#), a result often described as the “worst-case” bounds, although that phrase is a bit misleading, so we will describe these as the Manski-Robins bounds.⁶⁶ [Manski \(1990, 1994\)](#) observed that an instrument allows for the construction of the IV bounds in (88), which are often described as “intersection bounds” due to their max-min structure. The intersection bounds only depend on the outcome exogeneity assumption. They do not require full exogeneity nor any behavioral assumptions about selection, such as the monotonicity condition.

4.7.2 Empirical illustration

[Table 7](#) reports estimates of Manski-Robins and IV intersection bounds using data from [Gelbach \(2002\)](#), who estimated the impact of public school availability on maternal labor supply. The sample is restricted to mothers whose youngest child was five years old in 1980. The treatment D_i is an indicator for whether the mother’s five-year-old was enrolled in public school. The outcome Y_i is an indicator for whether the mother was employed in the previous year. Gelbach instruments for D_i with indicators Z_i for the quarter when the five-year-old was born, an instrument that is relevant because of age-at-entry rules for public kindergartens.

The top portion of [Table 7](#) reports two sets of estimated Manski-Robins bounds on the untreated means. In the first set of bounds, we take these to be the logically possible values for a binary variable of $y_{lb} = 0$ and $y_{ub} = 1$. In the second set of bounds, we make the substantive (potentially incorrect) assumption that counterfactual employment probabilities lie between $y_{lb} = .4$ and $y_{lb} = .8$, compared to estimated conditional employment probabilities $\mathbb{E}[Y_i|D_i = d, Z_i = z]$ that range between .49 and .71 over different values of d and z .

The first three rows of [Table 7](#) do not condition on the instrument. The unconditional treatment propensity is .632, so bounds on the treated mean are

⁶⁶All bounds are achieved at the “worst case.”

TABLE 7 Manski-Robins and instrumental variable bounds in Gelbach (2002).

	$p(z)$	Logical		Substantive	
		$(y_{lb} = 0, y_{ub} = 1)$	$(y_{lb} = 0.4, y_{ub} = 0.8)$	LB	UB
<i>Manski-Robins bounds</i>					
$\mathbb{E}[Y_i(0)]$.632	.275	.908	.528	.781
$\mathbb{E}[Y_i(1)]$.632	.425	.793	.572	.719
$\mathbb{E}[Y_i(1) - Y_i(0)]$		-.482	.518	-.209	.191
$\mathbb{E}[Y_i(0) Z_i = 75: Q1]$.313	.496	.809	.621	.746
$\mathbb{E}[Y_i(0) Z_i = 74: Q4]$.553	.344	.897	.565	.787
$\mathbb{E}[Y_i(0) Z_i = 74: Q3]$.793	.159	.952	.476	.793
$\mathbb{E}[Y_i(0) Z_i = 74: Q2]$.834	.127	.961	.461	.795
$\mathbb{E}[Y_i(1) Z_i = 75: Q1]$.313	.192	.879	.467	.742
$\mathbb{E}[Y_i(1) Z_i = 74: Q4]$.553	.355	.802	.534	.712
$\mathbb{E}[Y_i(1) Z_i = 74: Q3]$.793	.545	.752	.628	.711
$\mathbb{E}[Y_i(1) Z_i = 74: Q2]$.834	.582	.748	.648	.715
<i>Instrumental variable bounds</i>					
$\mathbb{E}[Y_i(0)]$.496	.809	.621	.746
$\mathbb{E}[Y_i(1)]$.582	.748	.648	.711
$\mathbb{E}[Y_i(1) - Y_i(0)]$		-.227	.252	-.098	.090

Notes: Sample analog estimates of the components of (88), the symmetric expressions for $\mathbb{E}[Y_i(0)]$, and bounds on the ATE formed from the difference. The data is the sample of 10,932 single mothers whose youngest child was five years old in 1980. The outcome variable is an indicator for employment in 1979. The values of Z_i indicate the birth quarter of this child. heteroskedasticity-robust standard errors (not shown) for $\mathbb{E}[Y_i|D_i = d, Z_i = z]$ are smaller than .02 for all values of d and z and for $p(z)$ are smaller than .01 for all values of z .

narrower than on the untreated mean. The implied bounds on the ATE are quite wide, even when placing substantive prior bounds on counterfactual employment probabilities. These bounds do not make use of the assumption that the instrument satisfies outcome exogeneity.

The subsequent rows report Manski-Robins bounds that condition on the instrument. These are just conditional-on- Z_i versions of the first three rows. The propensity score varies with the conditioning value of the instrument, leading to variation in the width of the bounds. The bounds for the untreated conditional mean are narrowest for the youngest children (75:Q1), who are

least likely to be in public kindergarten. For the treated conditional mean, they are narrowest for the oldest children (74:Q2). Outcome exogeneity allows bounds for the youngest and oldest children to be combined through the intersection bounds in (88). The result is shown in the final three rows together with the implied bounds on the ATE.

While narrower than the unconditional Manski-Robins bounds, both the logical and substantive IV intersection bounds on the ATE are still quite wide. As a point of reference, an uncontrolled OLS estimate is $-.076$ (standard error $.009$), which increases to $-.013$ (SE: $.008$) when adding state fixed effects and demographic controls (Gelbach, 2002, Table 7, columns (1)–(2)). An uncontrolled linear IV estimate is $.036$ (SE: $.022$), which increases to $.040$ (SE: $.020$) in Gelbach's preferred linear IV specification that includes controls (Gelbach, 2002, Table 7, column (3)).⁶⁷ The logical IV bounds on the ATE by contrast range from $-.227$ to $.252$. The substantive bounds of $-.098$ to $.090$ are considerably narrower, but are still consistent with negative or positive effects larger than any of Gelbach's estimates. This is despite the relatively wide variation in the propensity score between younger and older children from $.313$ to $.834$. That the IV bounds are inconclusive about the ATE even in a setting with this type of propensity score variation is probably why they are not often used in practice.⁶⁸

Inconclusive does not mean useless. The bounds shown in Table 7 are sharp, meaning the best possible given the assumption of outcome exogeneity. So, far from being useless, they indicate the central role played by making additional assumptions. A linear IV estimate based on a binarized version of the instrument that groups the two earliest and two latest quarters is $.034$ (SE: $.024$). If we assume away UHTE, then this estimate is a consistent estimate of the ATE. If we allow for UHTE, then the most we can conclude about the ATE is that it is contained within the bounds given in Table 7. In this way, the IV bounds quantify the empirical importance of assuming constant treatment effects.

4.7.3 *The role of a selection model*

By the same reasoning, the IV bounds also quantify the empirical importance of using a selection model. Under full exogeneity and monotonicity, the

⁶⁷ The bounds in Table 7 do not control for covariates. Controlling for covariates in a parsimonious way is a challenge for this type of bounding analysis.

⁶⁸ Applications of IV bounds can be found in Pepper (2000), Siddique (2013), and Shurtz et al. (2022). A more commonly-applied bounding approach uses an assumption that Manski and Pepper (2000) call monotone instrumental variables (MIV). The MIV assumption is that $\mathbb{E}[Y_i(d)|Z_i = z]$ is increasing or decreasing in z , instead of constant (both increasing and decreasing) as under outcome exogeneity. Bounds based on MIV can be compelling tools for causal inference, but the name MIV is perhaps a misnomer, as the whole point of the assumption is that the variables used as Z_i no longer need to be excluded and exogenous; monotone covariates would be an equally appropriate description. For applications of MIV, see Kreider and Pepper (2007), Blundell et al. (2007), Kreider et al. (2012), and De Haan and Leuven (2020).

binarized linear IV estimate of .034 is a consistent estimate of the LATE.⁶⁹ Comparing the LATE estimate to the ATE bounds measures how different the treatment effect for compliers could be from the treatment effect for the overall population. In this case, a positive LATE is consistent with an ATE that is positive and considerably larger, or negative and equally large. Similarly, using the selection model together with a linear MTE extrapolation produces an estimated ATE of .025 (SE.022). Selecting this number from the ATE bounds depends on the validity of the selection model and the extrapolation of the MTE.

The absence of a selection model helps clarify why one can be so useful. Modeling how counterfactual objects relate to observable ones is the fundamental challenge in all causal inference. For IV with a binary treatment, the counterfactual object is the conditional mean $\mathbb{E}[Y_i(1)|D_i = 0, Z_i = z]$. Bounding this object to $[y_{lb}, y_{ub}]$ is one simple model. For developing a more involved model, one effectively has two components to work with: the potential treatment arm, $Y_i(1)$ vs. $Y_i(0)$, or the conditioning event, $D_i = 0, Z_i = z$ vs. $D_i = d, Z_i = z$ for different values of d and z . One option with the former is to assume that $Y_i(1) \geq Y_i(0)$, an assumption Manski (1997) described as monotone treatment response. Apart from that, there is little scope for additional assumptions that still allow for UHTE, at least with a binary treatment.⁷⁰ That leaves modeling the conditioning event of the treatment and instrument value, which means modeling the treatment selection process.

Selections models become a necessity when the goal is to evaluate a policy change that affects treatment choice. The Cornelissen et al. (2018) evaluation of publicly provided childcare and the Ito et al. (2023) study of dynamic pricing provide two clear examples (see Section 4.3.7). There is little hope for conclusive inference about these types of policy counterfactuals without imposing assumptions on how the instrument affects treatment selection.

⁶⁹ Perhaps surprisingly, the ATE bounds in Table 7 are still sharp even after imposing monotonicity and full exogeneity (Balke and Pearl, 1993,1997; Heckman and Vytlacil, 2001b; Bai et al., 2024). To be more precise, these selection model assumptions have testable implications, as explored by the literature cited in Section 2.7, but if the testable implications are satisfied, then the sharp bounds on the ATE are the same as they are without the selection model. This result applies more generally to the entire marginal distributions of potential outcomes (Kitagawa, 2009,2021), but not to the joint distribution (Kamat, 2021). It breaks down when additional assumptions are added; see Flores and Chen (2018) for a survey and Shaikh and Vytlacil (2011), Bhattacharya et al. (2012), Huber et al. (2015), and Machado et al. (2019) for some specific results.

⁷⁰ One option that has been explored is rank invariance (Chernozhukov and Hansen, 2005), which can provide point identification without a selection model, albeit with a more complicated relevance condition on how the instrument affects the treatment. Vuong and Xu (2017) show how to combine rank invariance with the usual monotonicity condition to point identify the entire marginal distributions of potential outcomes under a standard instrument relevance condition.

4.8 Summary of forward engineering

In this section we've discussed some forward engineering approaches for incorporating UHTE into IV models. These run the gamut from the always-possible option of assuming there is no UHTE, to estimating LATEs directly, to extrapolating MTEs for binary or multivalued treatments, to bounding analyses that use only the most essential properties of an IV. What unites all of these approaches as forward engineering is that take the target parameter as the focus and design an estimator suitable for estimating it. What divides them are the target parameters they focus on and the assumptions they use to estimate those target parameters.

[Fig. 7](#) provides a familiar graphical tool for comparing these different approaches. It depicts a stylized production possibilities frontier for empirical research using IV with a binary treatment, with the attendant trade-off between assumptions and conclusions that [Manski \(2003\)](#) has described as the Law of Decreasing Credibility. At one corner is the assumption of no UHTE, the strongest assumption that we have considered in this chapter, under which the assumed-to-be-constant treatment effect is point identified under the classical linear IV assumptions. In the other corner are Manski-Robins and IV intersection bounds that do not impose full exogeneity or a selection model, but provide only a bound on the ATE. Selection models allow one to explore the area between these two corners, either at a single point with an unconditional LATE, or at multiple points by extrapolating an MTE curve under additional parametric or shape restrictions.

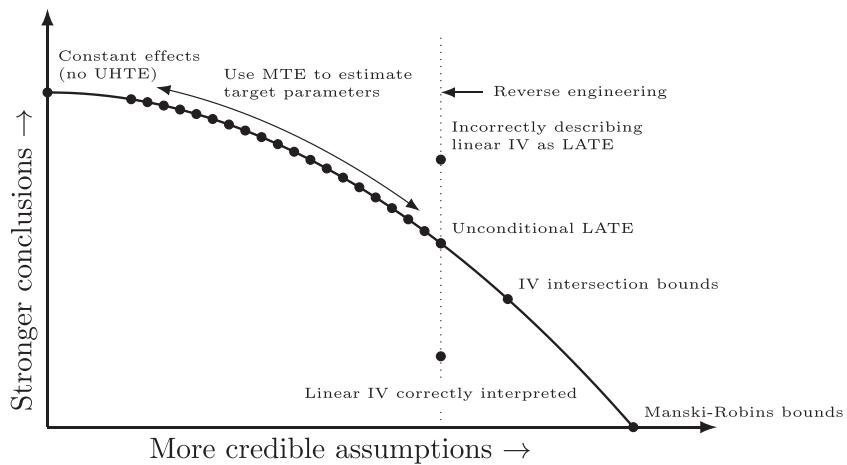


FIG. 7 The empirical production possibility frontier for IV methods. Notes: Two primary trade-offs involved in producing empirical research with a binary treatment.

Our two-dimensional figure can't capture the many dimensions on which empirical approaches differ. What is a stronger conclusion, a bound on an ATE or a point estimate of a LATE? How does one trade-off statistical precision with a broader and more ambitious target parameter? What is the appropriate value for a researcher's time, and how should this value be priced against the quality considerations of the empirical research they produce?

Yet we think that the idea of an empirical production function still nicely describes some useful points that should seem uncontroversial to most economists. Locating inside the frontier is suboptimal: instead of estimating a difficult-to-interpret statistically-weighted average of LATEs, it's possible under essentially the same assumptions to estimate an unconditional LATE using the estimators in [Section 4.2](#). Locating outside of the frontier is not possible: using a linear IV estimator as if it estimates an ATE, or even an unconditional LATE, might be typographically possible to put on paper, but it is logically incoherent under the baseline LATE assumptions. Locating on the frontier is the goal, but reasonable people can disagree about which point on the frontier they prefer.

5 Recommendations for practice

In this section we distill the discussion of this chapter into concrete recommendations for practitioners. We organize our recommendations around three steps:

Three steps for incorporating UHTE into an IV analysis

1. Assess the likely role of UHTE.
2. Reverse engineer, cautiously.
3. Forward engineer estimates of interpretable target parameters.

Throughout, we take it for granted that the instrument satisfies exclusion and outcome exogeneity. Supporting these assumptions is an important part of a compelling IV analysis, with or without UHTE.

5.1 Step 1: Assess the likely role of UHTE

All of the complications, subtleties, and caveats discussed in this chapter vanish under the assumption that there is no UHTE. Our view is that UHTE is probably a generic feature of economic environments. But it still makes sense to assess the source of UHTE and its likely magnitude before embarking on an econometric quest that would be much simpler if it could be convincingly assumed away.

The first task is to think about the nature of the treatment and outcome. Why could treatment effects vary? Is there any reason to think that treatment effects *would not* vary? We expect that the answer to this first question will be

negative in almost all cases: heterogeneous treatment effects cannot be ruled out on a priori grounds alone. For example, if the outcome Y_i is binary or discrete, treatment effects are almost necessarily heterogeneous.⁷¹

In order to create difficulties, however, the UHTE needs to be systematic, in the sense of being correlated with treatment choice. How plausible is it to assume that the UHTE is asystematic? Answering this question requires considering the source of endogeneity for which the IV is intended as a remedy. Does the treatment represent the choice that some economic agent is making? If so, is it possible to write down a plausible model in which the treatment choice is endogenous, but the agent makes their choice without regard to its possible effects on the outcome variable? A compelling positive answer may make it reasonable to assume that there is no UHTE.

To defend such an assumption, it seems like a reasonable exercise to consider *observed* heterogeneity in treatment effects (OHTE). Given an assumption of no UHTE, patterns of OHTE can be estimated with linear IV by interacting the treatment variable with pre-determined covariates, such as sociodemographic characteristics. It's possible for there to be UHTE but no OHTE, or for there to be OHTE but no UHTE. Nevertheless, it seems reasonable to support any assumption of no UHTE with compelling evidence that there is also no OHTE. Compelling evidence in practice would mean precisely estimated zeros for the interactions between treatment and background characteristics.

Assessing the likely role of UHTE

- Is there any a prior reason to believe that treatment effects are constant?
- Is there a reasonable behavioral model under which there is UHTE, but economic agents choose treatment without taking it into account?
- Is there compelling evidence on the lack of OHTE?
 - If the answer is “no” to all of these questions, then proceed to step two. Otherwise, ask yourself the following question:
- Are you willing to maintain an explicit assumption that there is no UHTE?
 - If the answer is “yes,” then use linear IV estimators, and be sure to include the explicit assumption of no UHTE when describing your empirical results. If the answer is “no,” then proceed to step two.

5.2 Step 2: Reverse engineer with caution

Our discussion in [Section 3](#) landed on the conclusion that reverse engineered interpretations of linear IV estimates are often not applicable. However, there are important cases in which these interpretations do apply. Reverse engineering arguments can be successful in these cases, but the necessary assumptions need to be assessed carefully and the interpretations reported accurately.

⁷¹ If the outcome is binary and the treatment is also binary, then $Y_i(1) - Y_i(0)$ can take three values: –1, 0, or 1. The assumption that treatment effects are constant implies that the average treatment effect across the population or within any subgroup must also be –1, 0, or 1.

The first task is determining which setting in [Table 3](#) is applicable and how this interacts with the assumptions about the selection process. In particular, the number of values that the treatment and instrument take is crucial for assessing whether the standard monotonicity condition is reasonable. The most favorable cases are when both the treatment and instrument are binary or ordered. Unordered treatments require careful thinking about the appropriate selection model. Unordered instruments require careful thinking about whether the standard monotonicity condition is likely to be violated. Unfortunately, there are currently no appealing general-purpose alternatives to the monotonicity condition. Average monotonicity has been suggested recently ([Frandsen et al., 2023](#)), but as our discussion in [Section 3.4](#) showed, it is generally no easier to justify than the usual monotonicity condition, except on the narrow technical grounds that it is mathematically weaker. Full exogeneity is required for reverse engineering in all settings.

The second task is determining whether the linear IV specification actually delivers a weakly causal interpretation under full exogeneity and an appropriate monotonicity condition. The primary concern here is satisfying the rich covariates condition. If the instrument is independent of covariates, then rich covariates is automatically satisfied. Otherwise, the covariate specification needs to be scrutinized. The [Ramsey \(1969\)](#) RESET test for a linear regression of Z_i on X_i is the primary tool for doing so; a rejection of the RESET test is a rejection of rich covariates and so also a rejection of the null hypothesis that the linear IV specification produces a weakly causal estimand. If evidence is found that the estimand is not weakly causal, then there's little point in proceeding to its interpretation. One can try adjusting the specification by hand or using machine learning tools to help select the specification, as illustrated in [Section 4.2.3](#).

Assuming that the estimand is weakly causal, the third task is giving it a more concrete interpretation. What factors determine the weights? Which subgroups receive the most weight? The least weight? What counterfactual would the estimand describe? How would one describe the counterfactual in a sentence, or explain it in words to a colleague? If these tasks are hard for the researcher, it suggests that the interpretation of the estimand is also going to be difficult for the consumer of the research as well. In some cases, such as with ordered treatments, there may be multiple competing interpretations to choose from or discuss jointly. Binarizing the instrument can ease the interpretation challenges created by having multiple complier groups.

The fourth task is the simplest but most important: clearly and transparently communicate the interpretation of the estimand and the assumptions on which the interpretation rests. Be both correct and honest. Incorrectly describing an estimand as “the LATE” is no better than incorrectly describing the exclusion or exogeneity assumptions. Both errors amount to logically incoherent descriptions of causal inference.

The four tasks of compelling reverse engineering can be remembered with the acronym JOSH: .

The JOSH method for reverse engineering

- Judge the setting.
- Obtain a weakly causal interpretation.
- Scrutinize the interpretation.
- Honestly communicate to the audience.

5.3 Step 3: Forward engineer estimates of interpretable target parameters

Valiant efforts at implementing the JOSH method of reverse engineering will still end in failure if the setting is simply too complicated to obtain a weakly causal interpretation. Our survey of reverse engineering suggests that this will often be so, as even moderate departures from the baseline LATE setting can make it difficult to obtain a weakly causal interpretation. Even when it is possible to reverse engineer a weakly causal interpretation, an honest description of this interpretation may be convoluted, unclear, or have only a loose connection to the motivating research question. These are all reasons for pursuing forward engineering as a complement to reverse engineering.

Forward engineering requires choosing some target parameters. Which target parameters are useful necessarily depends on the context and the researcher's motivation. What can be said about any given target parameter depends on what assumptions are made. This reflects the necessary trade-off captured in the empirical production frontier (Fig. 7). We recommend that researchers explore this frontier by reporting estimates of interesting target parameters under several different sets of assumptions. The linear-in-parameters specification that we used throughout much of Section 4 makes this relatively easy to do, at least if one sticks to point identified settings. Considering partial identification provides further flexibility for exploring the empirical production frontier but raises the difficulty of implementation.

Some target parameters will necessarily be easier to draw conclusions about than others. The unconditional LATE is often point identified nonparametrically while the ATE seldom is. We don't see this as a good reason to omit target parameters, either because they are interesting but too difficult, or easy but not particularly interesting. It is hard to disagree with [Imbens](#)'s (2010, pp. 414–15) advice to report both estimates of nonparametrically identified quantities, like LATEs, together with estimates of target parameters with higher "external validity." As part of this, [Imbens](#) (2010) emphasizes being clear about the degree to which estimates of these different quantities depend on different assumptions, another point that is hard to disagree with.

Our recommendation for forward engineering is to embrace these two points: estimate LATEs and estimate other target parameters that are relevant to the empirical question. Through it all, be clear and upfront about the role of the maintained assumptions, an important part of which is reporting estimates under different sets of assumptions. This recommendation shouldn't be controversial. It comes with a cost of more difficult implementation. How large of a cost this is depends on the setting.

Forward engineering is now relatively low cost for binary or multivalued treatments. Estimating unconditional LATEs or the unconditional ACR with propensity score weighting is simply a matter of estimating a logistic regression. Estimating MTE curves involves estimating a binary or ordered logistic regression together with some properly-specified linear regressions. Software is available in both R and Stata to streamline either task, although the MTE software is currently limited to binary treatments. Even applying machine learning methods like DDML is relatively low cost, albeit potentially demanding computationally.

Deviating from binary or ordered treatments to unordered treatments raises the cost of forward engineering considerably, and at current puts one into lesscharted methodological territory. There are a number of successful empirical examples of forward engineering that one can try to follow as a guide (see [Section 4.6](#)), however implementation will often require bespoke analysis and coding. This is not a reflection of the difficulty of forward engineering, but rather the difficulty of unordered treatments, a case for which there are also few meaningful reverse engineering results. The same comments apply to the binary or ordered treatment case without the usual monotonicity condition.

6 Conclusion

The literature on including UHTE for IV methods now spans several decades and has been recognized in two Nobel prize awards to three scholars. Reflections on this work have often focused on the question of whether LATEs are interesting quantities. See for example the exchange between [Deaton \(2010\)](#), [Imbens \(2010\)](#), and [Heckman and Urzua \(2010\)](#).

Our review suggests that this question is a bit of a red herring. As we showed in [Section 3](#), the LATE result is so specialized that empirical researchers, who by and large use linear IV, often *aren't actually estimating LATEs*. That's not a problem with the LATE, it's a problem with the practice of what we've described as reverse engineering: starting with an estimator and working backwards to an interpretation. The concept of a LATE is amenable to reverse engineering in a couple of stylized baseline cases, but by and large the types of complications usually found in empirical work invalidate simple interpretations of linear IV estimates as LATEs, or even as "weakly causal" estimands. We expect that the same type of fragility will also be found in other applications of reverse engineering once researchers start to consider how different forms of misspecification interact with each other.⁷²

The obvious alternative is also the oldest one: work forward, instead of backward. As we showed in [Section 4](#), there are now many well-developed and relatively low-cost tools that can be used to forward engineer estimators that

⁷² As one example, [Blandhol et al. \(2022\)](#) show that Angrist's (1998) interpretation of the OLS estimand under selection on observables collapses unless the propensity score has implicitly been correctly specified.

estimate specific target parameters with clear interpretations, including LATEs. These estimators do not rely on fundamentally different assumptions, they simply make the assumptions harder to hide and easier to adjust than when reverse engineering. Many of the challenges about modeling selection that arise in forward engineering also arise in reverse engineering. But solving them is easier without self-imposing the straightjacket of the linear IV estimator.

Appendix A Potential outcomes or latent variables? It's just notation ...

Start with a latent variable model of form (3) with $Y_i = f(D_i, \epsilon_i)$. Let the potential outcomes be defined as $Y_i(d) \equiv f(d, \epsilon_i)$. If treatment states take values in a finite set \mathcal{D} , then

$$Y_i = f(D_i, \epsilon_i) = \sum_{d \in \mathcal{D}} \mathbb{I}[D_i = d] f(d, \epsilon_i) \equiv \sum_{d \in \mathcal{D}} \mathbb{I}[D_i = d] Y_i(d) = Y_i(D_i).$$

So, starting with a latent variable model we have constructed potential outcomes that generate the same observed outcome, Y_i . The assumption that \mathcal{D} is finite is just to preserve the familiarity of the summation; if \mathcal{D} is infinite, then consider only the first and final equalities.

Conversely, suppose that there are potential outcomes $Y_i(d)$ for each treatment state in a set \mathcal{D} , which we again begin by assuming is finite and enumerated as $\mathcal{D} = \{d_0, d_1, \dots, d_J\}$. Let $\epsilon_{ij} \equiv Y_i(d_j)$ and $\epsilon_i \equiv (Y_i(d_0), Y_i(d_1), \dots, Y_i(d_J))$. Then take

$$f(D_i, \epsilon_i) \equiv \sum_{j=0}^J \mathbb{I}[D_i = d_j] \epsilon_{ij}.$$

This implies that

$$Y_i = \sum_{j=0}^J \mathbb{I}[D_i = d_j] Y_i(d_j) \equiv \sum_{j=0}^J \mathbb{I}[D_i = d_j] \epsilon_{ij} \equiv f(D_i, \epsilon_i).$$

So, starting with potential outcomes, we have constructed a latent variable model that generates the same observed outcome, Y_i . The assumption that \mathcal{D} is finite is again unimportant; if \mathcal{D} were infinite then ϵ_i would represent the random function $d \mapsto Y_i(d)$, and f would have domain that includes these random functions.

Appendix B Definition of a weakly causal estimand

Blandhol et al. (2022) introduced the definition of a weakly causal estimand as a way of extending the logic of a non-negatively weighted average to estimands that might not be expressible as a weighted average at all. To see how this could arise, consider an OLS estimator of Y_i on D_i and covariates X_i , and let β denote the estimand corresponding to the coefficient on D_i . Suppose that $D_i \in \{0, 1\}$ is binary and that $(Y_i(0), Y_i(1))$ is independent of D_i conditional on X_i . Then a bit of Frisch-Waugh-Lovell algebra shows that

$$\begin{aligned}\beta &= \mathbb{E}[\omega_0(X_i)\mathbb{E}[Y_i(0)|X_i]] + \mathbb{E}[\omega_1(X_i)\mathbb{E}[Y_i(1)|X_i]] \\ \text{where } \omega_d(x) &= \begin{cases} \mathbb{E}[(D_i - X'\delta)^{-1}(x'\delta)(1 - \mathbb{E}[D_i|X_i = x])], & \text{if } d = 0 \\ \mathbb{E}[(D_i - X'\delta)^{-1}\mathbb{E}[D_i|X_i = x](1 - x'\delta)], & \text{if } d = 1 \end{cases} \\ \text{with } \delta &\equiv \mathbb{E}[X_i X']^{-1} \mathbb{E}[X_i D_i].\end{aligned}\tag{B1}$$

In general, there is no way to rewrite (B1) as a weighted average of treatment effects, like (9), because $\omega_0(x) \neq -\omega_1(x)$ except when $\mathbb{E}[D_i|X_i] = X'\delta$. The notion that is captured by analyzing the sign of the weights for an estimand with a form like (9) becomes more complicated for an estimand that does not have that form, like β in (B1).

Here we will develop that notion in a way that is a bit more abstract than in [Blandhol et al. \(2022\)](#) and also allows for unordered treatments. Let F denote a distribution for $(\{Y_i(d)\}_{d \in \mathcal{D}}, \{D_i(z)\}_{z \in \mathcal{Z}}|X_i, Z_i)$, where \mathcal{D} and \mathcal{Z} are the set of values that the treatment and instrument take, respectively. Each F generates a distribution $\chi(F)$ of observables (Y_i, D_i, X_i, Z_i) through the definition of potential outcomes and the distribution of (X_i, Z_i) , which is not modeled, and is viewed as part of the definition of χ . An estimand is a function that takes a distribution of observables such as $\chi(F)$ and maps it into a number or vector of numbers $\tau(\chi(F))$.⁷³

Let \mathcal{F} denote the subset of F that satisfy a set of assumptions. For example, in an IV context with full exogeneity, \mathcal{F} only includes F for which the joint distribution of $Y_i(d)$ and $D_i(z)$ across all d and z is independent of Z_i , conditional on X_i . Usually, \mathcal{F} will also include further assumptions, such as the monotonicity condition, which rules out F that allow for certain types of choice groups, such as defiers.

Let $\mathcal{F}^\diamond \subseteq \mathcal{F}$ denote a subset of F in \mathcal{F} that share a property that we want to be reflected in the estimand, $\tau(\chi(F))$. Let \mathcal{T}^\diamond be a set of values for $\tau(\chi(F))$ that reflect this property. We say that the estimand τ is faithful to the pair $(\mathcal{F}^\diamond, \mathcal{T}^\diamond)$ if

$$F \in \mathcal{F}^\diamond \Rightarrow \tau(\chi(F)) \in \mathcal{T}^\diamond.\tag{B2}$$

For example, suppose D_i is ordered, taking values d_0, d_1, \dots, d_J . Then \mathcal{F}_+^\diamond could be the subset of F in \mathcal{F} for which $\mathbb{E}_F[Y_i(d_j) - Y_i(d_{j-1})|G_i = g, X_i = x]$ is non-negative for all $j \geq 1$, all g , and all x , where \mathbb{E}_F denotes expectation taken under F , and G_i is the usual group notation derived from $\{D_i(z)\}_{z \in \mathcal{Z}}$. The set \mathcal{T}_+^\diamond could be the set of non-negative numbers. Then τ is faithful to $(\mathcal{F}_+^\diamond, \mathcal{T}_+^\diamond)$ if $\tau(\chi(F)) \geq 0$ whenever $F \in \mathcal{F}$ is such that all treatment effects are non-negative.

For the ordered treatment case, [Blandhol et al. \(2022\)](#) say that τ is weakly causal if it is faithful to both $(\mathcal{F}_+^\diamond, \mathcal{T}_+^\diamond)$ and $(\mathcal{F}_-^\diamond, \mathcal{T}_-^\diamond)$, where \mathcal{F}_-^\diamond is the subset of \mathcal{F} for which all treatment effects are non-positive and \mathcal{T}_-^\diamond is the set of non-positive numbers. Given (B2), this means that (i) if all treatment effects are non-negative, then the estimand is also non-negative, and (ii) if all treatment effects are non-positive, then the estimand is also non-positive. This is the same notion that is usually captured by the sign of the weights being non-negative in a

⁷³ In what follows, τ could also just be some quantity that is determined by $\chi(F)$; it need not necessarily be an estimand in the sense of being the limit of some estimator.

decomposition like (9), but formalized in a way that can also be applied to estimands like (B1) that do not have a weighted average decomposition.⁷⁴

We can use this framework to extend the definition of a weakly causal estimand to the unordered case by choosing different pairs $(\mathcal{F}^\diamond, \mathcal{T}^\diamond)$. Let $\mathcal{F}_{0 \rightarrow \ell,+}^\diamond$ denote the subset of \mathcal{F} for which the contrast $\mathbb{E}_F[Y_i(d_\ell) - Y_i(d_0)|G_i = g, X_i = x]$ between a particular treatment state d_ℓ and a base state d_0 is non-negative for all g and x , ignoring the other treatment states. Keep \mathcal{T}_+ defined as the set of non-negative numbers. Let $\mathcal{F}_{0 \rightarrow \ell,-}^\diamond$ and \mathcal{T}_- be defined symmetrically for the non-positive case. Then an estimand τ is weakly causal for the treatment contrast between d_ℓ and d_0 if it is faithful to both $(\mathcal{F}_{0 \rightarrow \ell,+}^\diamond, \mathcal{T}_+)$ and $(\mathcal{F}_{0 \rightarrow \ell,-}^\diamond, \mathcal{T}_-)$.

Notice that because $\mathcal{F}^\diamond \subseteq \mathcal{F}$, whether an estimand is faithful to $(\mathcal{F}^\diamond, \mathcal{T}^\diamond)$ depends on the maintained assumptions on \mathcal{F} . For example, an estimand may not be weakly causal if \mathcal{F} includes F that do not satisfy a monotonicity condition, but might become weakly causal when \mathcal{F} is restricted to only include F that do satisfy a monotonicity condition. The nature of (B2) means that making \mathcal{F} a smaller set makes it easier for an estimand to be faithful to any given $(\mathcal{F}^\diamond, \mathcal{T}^\diamond)$ pair. The literature on reverse engineering can be seen as an effort to choose τ and \mathcal{F} in a way that ensures τ is weakly causal.

Appendix C Deriving the average causal response and an alternative decomposition

We first derive (22), which to our knowledge has only appeared in the literature for the case when $d_j = j$ are the integers. Start by decomposing the outcome as

$$\begin{aligned} Y_i &= \sum_{j=0}^J \mathbb{I}[D_i = d_j] Y_i(d_j) \\ &= \sum_{j=0}^J (\mathbb{I}[D_i \geq d_j] - \mathbb{I}[D_i \geq d_{j+1}]) Y_i(d_j) = Y_i(d_0) + \sum_{j=1}^J \mathbb{I}[D_i \geq d_j] (Y_i(d_j) - Y_i(d_{j-1})), \end{aligned} \quad (\text{C1})$$

where the final equality follows from a change of variables and d_{J+1} in the summand when $j = J$ can be interpreted as any value larger than d_J . Taking the

⁷⁴To show that an estimand is not weakly causal it suffices to find an $F \in \mathcal{F}_+^\diamond$ for which $\tau(\chi(F)) < 0$. For example, to show that (B1) is not weakly causal, one can construct an F such that $\mathbb{E}_F[Y_i(1) - Y_i(0)|X_i = x] \geq 0$ for all x , but with $\mathbb{E}_F[Y_i(0)|X_i = x]$ chosen to make $\beta < 0$. This is always possible when $\omega_0(x) \neq -\omega_1(x)$, because one can write β as

$$\beta = \mathbb{E}[(\omega_0(X_i) + \omega_1(X_i))\mathbb{E}_F[Y_i(0)|X_i]] + \mathbb{E}[\omega_1(X_i)(\mathbb{E}_F[Y_i(1) - Y_i(0)|X_i])]. \quad (\text{B3})$$

Even if $\omega_1(x) \geq 0$ for all x , the fact that $\omega_0(x) \neq -\omega_1(x)$ for all x means that $\beta = \tau(\chi(F))$ depends on some $\mathbb{E}_F[Y_i(0)|X_i = x]$, and therefore can change signs as F varies across \mathcal{F}_+^\diamond .

difference in conditional expectation with respect to z and applying full exogeneity then gives

$$\begin{aligned} & \mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] \\ &= \sum_{j=1}^J \mathbb{E}[\mathbb{I}[D_i(1) \geq d_j] - \mathbb{I}[D_i(0) \geq d_j]](Y_i(d_j) - Y_i(d_{j-1})). \end{aligned} \quad (\text{C2})$$

Under the monotonicity condition, the event that $D_i(1) < d_j$ and $D_i(0) \geq d_j$ has probability zero, so the difference in indicators only takes the value zero or one with positive probability. Conditioning on the event that it is one leaves

$$\begin{aligned} & \mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0] \\ &= \sum_{j=1}^J \mathbb{P}[D_i(1) \geq d_j > D_i(0)]\mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})|D_i(1) \geq d_j > D_i(0)]. \end{aligned}$$

The same algebraic argument can be applied to the denominator of the Wald estimand with Y_i replaced by $D_i = \sum_{j=0}^J \mathbb{I}[D_i = d_j]d_j$, yielding

$$\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] = \sum_{j=1}^J \mathbb{P}[D_i(1) \geq d_j > D_i(0)](d_j - d_{j-1}).$$

Taking the ratio gives the right-hand side of (22), which is equal to the simple IV estimand when the instrument is binary (recall (13)).

To derive the alternative group-based decomposition (23), let $\mathcal{G}(j) \equiv \{(g(0), g(1)) \in \mathcal{G}: g(1) \geq d_j > g(0)\}$ denote the subset of groups represented by the conditioning event in the unit causal response (22), recalling that \mathcal{G} is the set of groups who could have non-zero probability under the monotonicity condition. Then (22) can be written as

$$\frac{\mathbb{C}[Y_i, Z_i]}{\mathbb{C}[D_i, Z_i]} = \sum_{j=1}^J \frac{\mathbb{P}[G_i \in \mathcal{G}(j)]}{\sum_{\ell=1}^J \mathbb{P}[G_i \in \mathcal{G}(\ell)](d_\ell - d_{\ell-1})}\mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})|G_i \in \mathcal{G}(j)].$$

Focusing on the numerator, write

$$\begin{aligned} & \sum_{j=1}^J \mathbb{P}[G_i \in \mathcal{G}(j)]\mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})|G_i \in \mathcal{G}(j)] \\ &= \sum_{j=1}^J \mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})\mathbb{I}[G_i \in \mathcal{G}(j)]] \\ &= \sum_{j=1}^J \sum_{g \in \mathcal{G}} \mathbb{I}[g(1) \geq d_j > g(0)]\mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})\mathbb{I}[G_i = g]] \\ &= \sum_{g \in \mathcal{G}} \sum_{j=g(0)+1}^{g(1)} \mathbb{E}[Y_i(d_j) - Y_i(d_{j-1})|G_i = g]\mathbb{P}[G_i = g] \\ &= \sum_{g \in \mathcal{G}} \mathbb{P}[G_i = g]\mathbb{E}[Y_i(g(1)) - Y_i(g(0))|G_i = g]. \end{aligned}$$

The same argument applied to the denominator yields

$$\sum_{\ell=1}^J \mathbb{P}[G_i \in \mathcal{G}(\ell)](d_\ell - d_{\ell-1}) = \sum_{g' \in \mathcal{G}} \mathbb{P}[G_i = g'](g'(1) - g'(0)).$$

Expression (23) follows after multiplying and dividing each term in the numerator by $g(1) - g(0)$.

Appendix D Estimating the average causal response with covariates

Suppose that D_i is multivalued and ordered, taking values d_0, d_1, \dots, d_J , as in Section 3.6. Define $W_i(z) \equiv Y_i(D_i(z))$ as before. The second half of (36) did not depend on how many values D_i takes, so we still have:

$$\frac{\mathbb{E}[W_i(1) - W_i(0)]}{\mathbb{E}[D_i(1) - D_i(0)]} = \frac{\mathbb{E}[\mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i]]}{\mathbb{E}[\mathbb{E}[D_i|Z_i = 1, X_i] - \mathbb{E}[D_i|Z_i = 0, X_i]]},$$

But the interpretation of the left-hand side of this equality changes now that D_i takes multiple ordered values. Write $W_i(z) \equiv Y_i(D_i(z))$ like (C1):

$$W_i(z) = Y_i(d_0) + \sum_{j=1}^J \mathbb{1}[D_i(z) \geq d_j](Y_i(d_j) - Y_i(d_{j-1})).$$

Then $\mathbb{E}[W_i(1) - W_i(0)]$ matches the expression (C1) for the Z_i -differenced conditional mean of Y_i derived when Z_i was unconditionally exogenous:

$$\mathbb{E}[W_i(1) - W_i(0)] = \sum_{j=1}^J \mathbb{E}[(\mathbb{1}[D_i(1) \geq d_j] - \mathbb{1}[D_i(0) \geq d_j])(Y_i(d_j) - Y_i(d_{j-1}))].$$

The rest of the derivation in Appendix C shows that the right-hand side is the numerator of the ACR. The argument for the denominator follows the same logic with $W_i(z) \equiv Y_i(D_i(z))$ replaced by simply $D_i(z) = \sum_{j=0}^J \mathbb{1}[D_i(z) = d_j]d_j$.

Appendix E Derivations for marginal treatment effects

This appendix contains some derivations relevant binary marginal treatment effects for binary treatments.

E.1 Derivations of weighting expressions

Consider the ATT expression given in (47) and Table 5. We derive this by iterating expectations:

$$\begin{aligned}
\mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1] &= \mathbb{E}\left[(Y_i(1) - Y_i(0)) \frac{\mathbb{I}[U_i \leq p(Z_i, X_i)]}{\mathbb{P}[D_i = 1]}\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y_i(1) - Y_i(0)|U_i, X_i, Z_i] \frac{\mathbb{I}[U_i \leq p(Z_i, X_i)]}{\mathbb{P}[D_i = 1]}\right] \\
&= \mathbb{E}\left[(\text{MTR}(1|U_i, X_i) - \text{MTR}(0|U_i, X_i)) \frac{\mathbb{I}[U_i \leq p(Z_i, X_i)]}{\mathbb{P}[D_i = 1]}\right] \\
&= \mathbb{E}\left[\int_0^1 (\text{MTR}(1|u, X_i) - \text{MTR}(0|u, X_i)) \frac{\mathbb{I}[u \leq p(Z_i, X_i)]}{\mathbb{P}[D_i = 1]} du\right].
\end{aligned}$$

The third equality used full exogeneity and the definition of the MTR. The fourth equality iterated expectation on U_i given X_i and Z_i and used the normalization that the conditional distribution of U_i is uniformly distributed.

The linear-in-parameters representation (49) for the conditional mean of the observed outcome is derived like this:

$$\begin{aligned}
\mathbb{E}[Y_i|D_i = 1, P_i = \bar{p}, X_i] &= \mathbb{E}[Y_i(1)|U_i \leq \bar{p}, P_i = \bar{p}, X_i] \\
&= \mathbb{E}[Y_i(1)|U_i \leq \bar{p}, X_i] \\
&= \mathbb{E}[\mathbb{E}[Y_i(1)|U_i, X_i]|U_i \leq \bar{p}, X_i] \\
&= \mathbb{E}[\text{MTR}(1|U_i, X_i)|U_i \leq \bar{p}, X_i] = \frac{1}{\bar{p}} \int_0^{\bar{p}} \text{MTR}(1|u, X_i) du,
\end{aligned}$$

where the second equality uses full exogeneity that the fact that $P_i \equiv p(Z_i, X_i)$ is a function of Z_i and X_i . A symmetric derivation for the untreated arm gives

$$\mathbb{E}[Y_i|D_i = 0, P_i = \bar{p}, X_i] = \frac{1}{(1 - \bar{p})} \int_{-\bar{p}}^1 \text{MTR}(0|u, X_i) du.$$

Putting them together and applying the linear-in-parameters assumption (48) produces (49):

$$\begin{aligned}
\mathbb{E}[Y_i|D_i, P_i, X_i] &= \frac{(1 - D_i)}{(1 - P_i)} \int_{-P_i}^1 \text{MTR}(0|u, X_i) du + \frac{D_i}{P_i} \int_0^{P_i} \text{MTR}(1|u, X_i) du \\
&= \frac{(1 - D_i)}{(1 - P_i)} \int_{-P_i}^1 \sum_{k=1}^{d_\theta} \theta_k b_k(0|u, X_i) du + \frac{D_i}{P_i} \int_0^{P_i} \sum_{k=1}^{d_\theta} \theta_k b_k(1|u, X_i) du \\
&= \sum_{k=1}^{d_\theta} \theta_k B_{ik},
\end{aligned}$$

where B_{ik} is as defined in (49).

Lastly, the expression (62) used to derive the LIV follows from

$$\begin{aligned}\mathbb{E}[D_i(Y_i(1) - Y_i(0))|P_i, X_i] &= \mathbb{E}[\mathbb{I}[U_i \leq P_i](Y_i(1) - Y_i(0))|P_i, X_i] \\ &= \mathbb{E}[\mathbb{I}[U_i \leq P_i]\text{MTE}(U_i, X_i)|P_i, X_i] = \int_0^{P_i} \text{MTE}(U_i, X_i) du,\end{aligned}$$

again making use of the normalization that U_i is uniform given Z_i and X_i , and so also given P_i and X_i .

E.2 The normal selection model

This section provides detail on the derivation of (56). With the pre-normalization selection equation the propensity score satisfies

$$p(z) \equiv \mathbb{P}[D_i = 1|Z_i = z] = \mathbb{P}[V_i \leq \nu(z)] = \Phi(\nu(z)),$$

noting that V_i has mean zero and variance one. Consider $b_4(d|u) = d\Phi^{-1}(u)$. Then

$$B_{i4} = D_i \frac{1}{P_i} \int_0^{P_i} \Phi^{-1}(u) du \quad \text{where } P_i = \Phi(\nu(Z_i)).$$

A change of variables to $v \equiv \Phi^{-1}(u)$ leads to a mean of a standard normal truncated at $\Phi^{-1}(P_i)$:

$$B_{i4} = D_i \int_{-\infty}^{\Phi^{-1}(P_i)} v \frac{\phi(v)}{\Phi(\Phi^{-1}(P_i))} dv.$$

This can be expressed in terms of the inverse Mills' ratio (e.g [Hansen, 2022a](#), pg. 116), so that

$$B_{i4} = -D_i \frac{\phi(\Phi^{-1}(P_i))}{\Phi(\Phi^{-1}(P_i))} \equiv -D_i \lambda(\Phi^{-1}(P_i)).$$

The corresponding term for the untreated arm is derived similarly.

E.3 Saturated MTR specifications reproduce the LATE

We give a simple and general justification of the finding in [Brinch et al. \(2017\)](#) and [Kline and Walters \(2019\)](#) that in saturated settings even a misspecified MTR reproduces the usual LATE. For simplicity, suppose that Z_i is binary and there are no covariates. Consider $\mathbb{E}[Y_i|Z_i = z]$ as a target parameter. An MTR function parameterized by θ implies the following values for this target parameter:

$$\mathbb{E}_\theta[Y_i|Z_i = z] = \mathbb{E}_\theta[Y_i(0)] + \mathbb{E}_\theta[\mathbb{I}[U_i \leq p(z)](Y_i(1) - Y_i(0))],$$

where \mathbb{E}_θ indicates expectation taken under the assumption that the MTR function is parameterized by θ . When the MTR follows the linear-in-parameters specification (48),

$$\mathbb{E}_\theta[Y_i|Z_i = z] = \sum_{k=1}^{d_\theta} \theta_k \int_0^1 b_k(0|u) du + \theta_k \int_0^{p(z)} (b_k(1|u) - b_k(0|u)) du.$$

The LATE formed from θ is then

$$\frac{\mathbb{E}_\theta[Y_i|Z_i = 1] - \mathbb{E}_\theta[Y_i|Z_i = 0]}{p(1) - p(0)} = \sum_{k=1}^{d_\theta} \theta_k \int_0^1 (b_k(1|u) - b_k(0|u)) \frac{\mathbb{I}[p(0) < u \leq p(1)]}{p(1) - p(0)} du,$$

which matches the expression given in Table 5, with $\underline{u} = p(0)$, $\bar{u} = p(1)$, and $MTR(d|u) = \sum_{k=1}^{d_\theta} \theta_k b_k(d|u)$.

Given this observation, it suffices to show that if the implied regression specification (49) is saturated, then $\mathbb{E}_m[Y_i|Z_i = 1] = \mathbb{E}[Y_i|Z_i = 1]$ even if the MTR function m is misspecified. Saturation implies that for all values of d and z ,

$$\mathbb{E}[Y_i|D_i = d, Z_i = z] = \mathbb{E}[Y_i|D_i = d, P_i = p(z)] = \sum_{k=1}^{d_\theta} \theta_k B_k(d, p(z)), \quad (\text{E1})$$

where

$$B_k(d, p(z)) \equiv \left(\frac{1-d}{1-p(z)} \right) \int_{p(z)}^1 b_k(0|u) du + \frac{d}{p(z)} \int_0^{p(z)} b_k(1|u) du.$$

So, in particular,

$$\begin{aligned} \mathbb{E}[Y_i|Z_i = z] &= \mathbb{E}[Y_i|D_i = 0, Z_i = z](1 - p(z)) + \mathbb{E}[Y_i|D_i = 1, Z_i = z]p(z) \\ &= \sum_{k=1}^{d_\theta} \theta_k (B_k(0, p(0))(1 - p(0)) + B_k(1, p(1))p(1)) \\ &= \sum_{k=1}^{d_\theta} \theta_k \left(\int_{p(z)}^1 b_k(0|u) du + \int_0^{p(z)} b_k(1|u) du \right) \\ &= \sum_{k=1}^{d_\theta} \theta_k \int_0^1 b_k(0|u) du + \sum_{k=1}^{d_\theta} \theta_k \int_0^{p(z)} (b_k(1|u) - b_k(0|u)) du \\ &= \mathbb{E}_\theta[Y_i|Z_i = z]. \end{aligned}$$

Note that the same argument works for the sample as well: a saturated specification will continue to satisfy the sample analog of (E1) with $\mathbb{E}[Y_i|D_i = d, Z_i = z]$ and $p(z)$ replaced by their conditional sample means, and with θ_k replaced by its estimator. The argument holds outside of a regression context and doesn't depend on the linear-in-parameters specification. The driving observation is just that in a saturated model, even an incorrectly-specified MTR function that is fit to the data will produce the same estimates of $\mathbb{E}[Y_i|D_i = d, Z_i = z]$ for all values of d and z , and so will also produce the same estimate of the LATE.

References

- Abadie, A., 2003. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 113, 231–263.
- Abdulkadiroğlu, A., Pathak, P.A., Schellingen, J., Walters, C.R., 2020. Do parents value school effectiveness? *American Economic Review* 110, 1502–1539.

- Acerenza, S., 2024. Partial identification of marginal treatment effects with discrete instruments and misreported treatment*. *Oxford Bulletin of Economics and Statistics* 86, 74–100.
- Acerenza, S., Ban, K., Kédagni, D., 2023. Marginal treatment effects with a misclassified treatment.
- Acerenza, S., Possebom, V., Sant'Anna, P.H.C., 2024. Was javert right to be suspicious? Unpacking treatment effect heterogeneity of alternative sentences on time-to-recidivism in Brazil.
- Agan, A., Doleac, J.L., Harvey, A., 2023. Misdemeanor prosecution. *The Quarterly Journal of Economics* 138, 1453–1505.
- Ahrens, A., Hansen, C.B., Schaffer, M.E., Wiemann, T., 2023. Ddml: double/debiased machine learning in Stata.
- Ahrens, A., Hansen, C.B., Schaffer, M.E., Wiemann, T., 2024a. Ddml: double/debiased machine learning.
- Ahrens, A., Hansen, C.B., Schaffer, M.E., Wiemann, T., 2024b. Model averaging and double machine learning.
- Aizawa, N., Mommaerts, C., Rennane, S.L., 2023. Firm accommodation after disability: labor market impacts and implications for social insurance.
- Alessie, R.J.M., Angelini, V., Mierau, J.O., Viluma, L., 2020. Moral hazard and selection for voluntary deductibles. *Health Economics* 29, 1251–1269.
- Alvarez, L.A., Toneto, R., 2024. The interpretation of 2SLS with a continuous instrument: a weighted LATE representation. *Economics Letters* 237, 111658.
- Andresen, M.E., 2018. Exploring marginal treatment effects: flexible estimation using Stata. *The Stata Journal: Promoting communications on statistics and Stata* 18, 118–158.
- Andresen, M.E., Huber, M., 2021. Instrument-based estimation with binarised treatments: issues and tests for the exclusion restriction. *The Econometrics Journal* 24, 536–558.
- Andrews, I., Stock, J.H., Sun, L., 2019. Weak instruments in instrumental variables regression: theory and practice. *Annual Review of Economics* 11, 727–753.
- Angrist, J.D., 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica* 66, 249–288.
- Angrist, J.D., 2022. Empirical strategies in economics: illuminating the path from cause to effect. *Econometrica* 90, 2509–2539.
- Angrist, J.D., Evans, W.N., 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. *The American Economic Review* 88, 450–477.
- Angrist, J.D., Fernández-Val, I., 2013. EXTRAPOLATE-ing: external validity and overidentification in the LATE framework. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), *Advances in Economics and Econometrics*. Cambridge University Press, pp. 401–434.
- Angrist, J.D., Graddy, K., Imbens, G.W., 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies* 67, 499–527.
- Angrist, J.D., Imbens, G.W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90, 431–442.
- Angrist, J.D., Imbens, G.W., Krueger, A.B., 1999. Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14, 57–67.
- Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.
- Angrist, J.D., Krueger, A.B., 1999. Chapter 23 Empirical Strategies in Labor Economics. vol. 3. Elsevier, pp. 1277–1366 Part A.

- Angrist, J.D., Pathak, P.A., Walters, C.R., 2013. Explaining charter school effectiveness. *American Economic Journal: Applied Economics* 5, 1–27.
- Angrist, J.D., Pischke, J.-S., 2009. *Mostly Harmless Econometrics: An Empiricistas Companion*. Princeton University Press.
- Arbour, W., 2022. Can recidivism be prevented from behind bars? Evidence from a behavioral program. Working Paper.
- Arnold, D., Dobbie, W., Hull, P., 2022. Measuring racial discrimination in bail decisions. *American Economic Review* 112, 2992–3038.
- Arnold, D., Dobbie, W., Yang, C.S., 2018. Racial bias in bail decisions. *The Quarterly Journal of Economics* 133, 1885–1932.
- Arteaga, C., 2023. Parental incarceration and children's educational attainment. *Review of Economics and Statistics* 105, 1394–1410.
- Bai, Y., Huang, S., Moon, S., Shaikh, A.M., Vytlacil, E.J., 2024. On the identifying power of monotonicity for average treatment effects.
- Bailey, M.J., Goodman-Bacon, A., 2015. The war on poverty's experiment in public medicine: community health centers and the mortality of older Americans. *American Economic Review* 105, 1067–1104.
- Baker, S.G., Lindeman, K.S., 1994. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine* 13, 2269–2278.
- Balke, A., Pearl, J., 1993. Nonparametric bounds on causal effects from partial compliance data. Technical Report R-199, University of California, Los Angeles, Computer Science Department.
- Balke, A., Pearl, J., 1997. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association* 92, 1171–1176.
- Balla-Elliott, D., 2023. Identifying causal effects in information provision experiments.
- Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., Sulaiman, M., 2020. Women's empowerment in action: evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics* 12, 210–259.
- Baron, E.J., Gross, M., 2022. Is there a foster care-to-prison pipeline? Evidence from quasi-randomly assigned investigators.
- Basu, A., Heckman, J.J., Navarro-Lozano, S., Urzua, S., 2007. Use of instrumental variables in the presence of heterogeneity and self-selection: an application to treatments of breast cancer patients. *Health Economics* 16, 1133–1157.
- Basu, A., Jena, A.B., Goldman, D.P., Philipson, T.J., Dubois, R., 2014. Heterogeneity in action: the role of passive personalization in comparative effectiveness research. *Health Economics* 23, 359–373.
- Becker, G., 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, first ed. National Bureau of Economic Research, Inc.
- Becker, G.S., 1967. *Human Capital and the Personal Distribution of Income: An Analytical Approach*, 1. Institute of Public Administration.
- Behaghel, L., Crepon, B., Gurgand, M., 2013. Robustness of the encouragement design in a two-treatment randomized control trial.
- Bekker, P.A., 1994. Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.
- Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C., 2017. Program evaluation and causal inference with high-dimensional data. *Econometrica* 85, 233–298.
- Benson, D., Masten, M.A., Torgovitsky, A., 2022. Ivrc: an instrumental-variables estimator for the correlated random-coefficients model. *The Stata Journal: Promoting Communications on Statistics and Stata* 22, 469–495.

- Berry, J., Fischer, G., Guiteras, R., 2020. Eliciting and utilizing willingness to pay: evidence from field trials in Northern Ghana. *Journal of Political Economy* 128, 1436–1473.
- Bhattacharya, J., Shaikh, A.M., Vytlacil, E., 2012. Treatment effect bounds: an application to Swan-Ganz catheterization. *Journal of Econometrics* 168, 223–243.
- Bhuller, M., Dahl, G.B., Løken, K.V., Mogstad, M., 2020. Incarceration, recidivism, and employment. *Journal of Political Economy* 128, 1269–1324.
- Bhuller, M., Sigstad, H., 2024. 2SLS with multiple treatments. *Journal of Econometrics* 242, 105785.
- Björklund, A., Moffitt, R., 1987. The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics* 69, 42–49.
- Blandhol, C., Bonney, J., Mogstad M., Torgovitsky, A., 2022. When is TSLS actually LATE? Tech. Rep. w29709, National Bureau of Economic Research, Cambridge, MA.
- Blandhol, C., Bonney, J., Mogstad, M., Torgovitsky, A., 2024. When is TSLS actually LATE? Tech. Rep. w29709, National Bureau of Economic Research, Cambridge, MA.
- Blundell, R., Dias, M.C., 2009. Alternative approaches to evaluation in empirical microeconomics. *The Journal of Human Resources* 44, 565–640.
- Blundell, R., Gosling, A., Ichimura, H., Meghir, C., 2007. Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica* 75, 323–363.
- Brave, S., Walstrum, T., 2014. Estimating marginal treatment effects using parametric and semiparametric methods. *The Stata Journal: Promoting communications on statistics and Stata* 14, 191–217.
- Briggs, J., Sachs, G., Caplin , A., Leth-Petersen, S., Tonetti, C., 2024. Identification of marginal treatment effects using subjective expectations.
- Brinch, C.N., Mogstad, M., Wiswall, M., 2012. Beyond LATE with a discrete instrument, Working paper.
- Brinch, C.N., Mogstad, M., Wiswall, M., 2017. Beyond LATE with a discrete instrument. *Journal of Political Economy* 125, 985–1039.
- Calonico, S., Cattaneo, M.D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82, 2295–2326.
- Calvi, R., Lewbel, A., Tommasi, D., 2022. LATE with missing or mismeasured treatment. *Journal of Business & Economic Statistics* 40, 1701–1717.
- Canay, I.A., Shaikh, A.M., 2017. Practical and theoretical advances in inference for partially identified models. In: Honore, B., Pakes, A., Piazzesi, M., Samuelson, L. (Eds.), *Advances in Economics and Econometrics*. Cambridge University Press, pp. 271–306.
- Card, D., 1993. Using geographic variation in college proximity to estimate the return to schooling, Tech. rep., National Bureau of Economic Research.
- Card, D., 1995. Using geographic variation in college proximity to estimate the return to schooling. In: Christofides, L.N., Grant, K.E., Swidinsky, R. (Eds.), *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*. University of Toronto Press, Toronto, pp. 201–222.
- Card, D., 1999. Chapter 30 The Causal Effect of Education on Earnings. Elsevier, pp. 1801–1863 vol. Volume 3, Part A.
- Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. *Econometrica* 69, 1127–1160.
- Card, D., Lee, D.S., Pei, Z., Weber, A., 2015. Inference on causal effects in a generalized regression kink design. *Econometrica* 83, 2453–2483.

- Carneiro, P., Heckman, J.J., Vytlacil, E., 2010. Evaluating marginal policy changes and the average effect of treatment for individuals at the margin. *Econometrica* 78, 377–394.
- Carneiro, P., Heckman, J.J., Vytlacil, E.J., 2011. Estimating marginal returns to education. *American Economic Review* 101, 2754–2781.
- Carneiro, P., Lee, S., 2009. Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. *Journal of Econometrics* 149, 191–208.
- Carneiro, P., Lokshin, M., Umapathi, N., 2016. Average and marginal returns to upper secondary schooling in Indonesia. *Journal of Applied Econometrics* 32, 16–36.
- Carr, T., Kitagawa, T., 2023. Testing instrument validity with covariates.
- Carrillo, P., Donaldson, D., Pomeranz, D., Singhal, M., 2023. Misallocation in firm production: a nonparametric analysis using procurement lotteries, Tech. Rep. w31311, National Bureau of Economic Research.
- Cattaneo, M.D., Keele, L., Titiunik, R., Vazquez-Bare, G., 2016. Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics* 78, 1229–1248.
- Chen, X., 2007. Chapter 76 Large sample sieve estimation of semi-nonparametric models. In: Heckman, J.J., Leamer, E.E. (Eds.), *Handbook of Econometrics*. Elsevier, pp. 5549–5632 , Volume 6, Part 2.
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal* 21, C1–C68.
- Chernozhukov, V., Fernández-Val, I., Melly, B., 2013. Inference on counterfactual distributions. *Econometrica* 81, 2205–2268.
- Chernozhukov, V., Fernández-Val, I., Newey, W., Stouli, S., Vella, F., 2020. Semiparametric estimation of structural functions in nonseparable triangular models. *Quantitative Economics* 11, 503–533.
- Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. *Econometrica* 73, 245–261.
- Chyn, E., Frandsen, B., Leslie, E.C., 2024. Examiner and judge designs in economics: a practitioner's guide.
- Cornelissen, T., Dustmann, C., Raute, A., Schönberg, U., 2016. From LATE to MTE: alternative methods for the evaluation of policy interventions. *Labour Economics* 41, 47–60.
- Cornelissen, T., Dustmann, C., Raute, A., Schönberg, U., 2018. Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy* 126, 2356–2409.
- Coury, M., Kitagawa, T., Shertzer, A., Turner, M., 2022. The value of piped water and sewers: evidence from 19th century Chicago.
- Cunha, F., Heckman, J.J., Navarro, S., 2007. The identification and economic content of ordered choice models with stochastic thresholds. *International Economic Review* 48, 1273–1309.
- Cunningham, S., 2021. *Causal Inference: The Mixtape*. Yale University Press.
- Dahl, G.B., 2002. Mobility and the return to education: testing a roy model with multiple markets. *Econometrica* 70, 2367–2420.
- Dahl, G.B., Kostøl, A.R., Mogstad, M., 2014. Family welfare cultures. *The Quarterly Journal of Economics* 129, 1711–1752.
- DalBó, E., Finan, F., Li, N.Y., Schechter, L., 2021. Information technology and government decentralization: experimental evidence from Paraguay. *Econometrica* 89, 677–701.
- Daljord, Ø., Mela, C.F., Roos, J.M.T., Sprigg, J., Yao, S., 2023. The design and targeting of compliance promotions. *Marketing Science* 42, 866–891.

- de Chaisemartin, C., 2017. Tolerating defiance? Local average treatment effects without monotonicity. *Quantitative Economics* 8, 367–396.
- de Chaisemartin, C., D'Haultfoeuille, X., 2018. Fuzzy differences-in-differences. *The Review of Economic Studies* 85, 999–1028.
- de Chaisemartin, C., Lei, Z., 2023. More robust estimators for instrumental-variable panel designs, with an application to the effect of imports from China on US Employment.
- De Groot, O., Declercq, K., 2021. Tracking and specialization of high schools: heterogeneous effects of school choice. *Journal of Applied Econometrics* 36, 898–916.
- De Haan, M., Leuven, E., 2020. Head start and the distribution of long-term education and labor market outcomes. *Journal of Labor Economics* 38, 727–765.
- Deaton, A., 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* 48, 424–455.
- Depalo, D., 2020. Explaining the causal effect of adherence to medication on cholesterol through the marginal patient. *Health Economics* (n/a).
- Dinkelman, T., 2011. The effects of rural electrification on employment: new evidence from South Africa. *American Economic Review* 101, 3078–3108.
- Dobbie, W., Goldin, J., Yang, C.S., 2018. The effects of pre-trial detention on conviction, future crime, and employment: evidence from randomly assigned judges. *American Economic Review* 108, 201–240.
- Dobbie, W., Song, J., 2015. Debt relief and debtor outcomes: measuring the effects of consumer bankruptcy protection. *American Economic Review* 105, 1272–1311.
- Donald, S.G., Hsu, Y.-C., Lieli, R.P., 2014. Testing the unconfoundedness assumption via inverse probability weighted estimators of (L)ATT. *Journal of Business & Economic Statistics* 32, 395–415.
- Doyle Jr., J.J., 2007. Child protection and child outcomes: measuring the effects of foster care. *The American Economic Review* 97, 1583–1610.
- Doyle Jr., J.J., 2008. Child protection and adult crime: using investigator assignment to estimate causal effects of foster care. *Journal of Political Economy* 116, 746–770.
- Dubin, J.A., McFadden, D.L., 1984. An econometric analysis of residential electric appliance holdings and consumption. *Econometrica* 52, 345.
- Dutz, D., Greenstone, M., Hortaçsu, A., Lacouture, S., Mogstad, M., Roumis, D., Shaikh, A.M., Torgovitsky, A., Van Dijk, W., 2023a. Selection bias in voluntary random testing: evidence from a COVID-19 antibody study. *AEA Papers and Proceedings* 113, 562–566.
- Dutz, D., Greenstone, M., Hortaçsu, A., Lacouture, S., Mogstad, M., Shaikh, A.M., Torgovitsky, A., van Dijk, W., 2023b. Representation and hesitancy in population health research: evidence from a COVID-19 antibody study.
- Dutz, D., Huitfeldt, I., Lacouture, S., Mogstad, M., Torgovitsky, A., van Dijk, W., 2022. Selection in surveys: using randomized incentives to detect and account for nonresponse bias, Tech. Rep. w29549, National Bureau of Economic Research, Cambridge, MA.
- Eisenhauer, P., Heckman, J.J., Vytlacil, E., 2015. The generalized Roy model and the cost-benefit analysis of social programs. *Journal of Political Economy* 123, 413–443.
- Fang, Z., Santos, A., 2019. Inference on directionally differentiable functions. *The Review of Economic Studies* 86, 377–412.
- Felfe, C., Lalivé, R., 2018. Does early child care affect children's development? *Journal of Public Economics* 159, 33–53.
- Florens, J.P., Heckman, J.J., Meghir, C., Vytlacil, E., 2008. Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica* 76, 1191–1206.

- Flores, C.A., Chen, X., et al., 2018. Average Treatment Effect Bounds with an Instrumental Variable: Theory and Practice. Springer.
- Frandsen, B., Lefgren, L., Leslie, E., 2023. Judging judge fixed effects. *American Economic Review* 113, 253–277.
- Frangakis, C.E., Rubin, D.B., 2002. Principal stratification in causal inference. *Biometrics* 58, 21–29.
- French, E., Song, J., 2014. The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy* 6, 291–337.
- Frölich, M., 2007. Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics* 139, 35–75.
- Galasso, A., Schankerman, M., Serrano, C.J., 2013. Trading and enforcing patent rights. *The RAND Journal of Economics* 44, 275–312.
- Garen, J., 1984. The returns to schooling: a selectivity bias approach with a continuous choice variable. *Econometrica* 52, 1199.
- Gathmann, C., Vonnahme, C., Busse, A., Kim, J., 2021. Marginal Returns to Citizenship and Educational Performance. RWI, DE.
- Gautier, E., 2021. Relaxing monotonicity in endogenous selection models and application to surveys. In: Daouia, A., Ruiz-Gazen, A. (Eds.), *Advances in Contemporary Statistics and Econometrics*. Springer International Publishing, Cham, pp. 59–78.
- Gautier, E., Hoderlein, S., 2015. A triangular treatment effect model with random coefficients in the selection equation, arXiv:1109.0362 [math, stat].
- Gelbach, J.B., 2002. Public schooling for young children and maternal labor supply. *The American Economic Review* 92, 307–322.
- Geweke, J., Gowrisankaran, G., Town, R.J., 2003. Bayesian inference for hospital quality in a selection model. *Econometrica* 71, 1215–1238.
- Goff, L., 2024. A vector monotonicity assumption for multiple instruments. *Journal of Econometrics* 241, 105735.
- Goldberger, A.S., 1983. Abnormal selection bias. In: Karlin, S., Amemiya, T., Goodman, L.A. (Eds.), *Studies in Econometrics, Time Series, and Multivariate Statistics*. Academic Press, pp. 67–84.
- Goldin, J., Lurie, I.Z., McCubbin, J., 2021. Health insurance and mortality: experimental evidence from taxpayer outreach. *The Quarterly Journal of Economics* 136, 1–49.
- Goldsmith-Pinkham, P., Hull, P., Kolesár, M., 2024. Contamination bias in linear regressions.
- Gollin, D., Udry, C., 2021. Heterogeneity, measurement error, and misallocation: evidence from African agriculture. *Journal of Political Economy* 129, 1–80.
- Gonçalves, F.M., Mello, S., 2023. Police discretion and public safety.
- Gong, J., Lu, Y., Xie, H., 2020. The average and distributional effects of teenage adversity on long-term health. *Journal of Health Economics* 71, 102288.
- Goodman, J., Gurantz, O., Smith, J., 2020. Take two! SAT retaking and college enrollment gaps. *American Economic Journal: Economic Policy* 12, 115–158.
- Goodman-Bacon, A., 2021. Difference-in-Differences With Variation In Treatment Timing. *Journal of Econometrics* 225, 254–277.
- Greene, W.H., Hensher, D.A., 2009. *Modeling Ordered Choices*. Cambridge University Press.
- Griliches, Z., 1977. Estimating the returns to schooling: some econometric problems. *Econometrica* 45, 1–22.
- Gronau, R., 1974. Wage comparisons—a selectivity bias. *Journal of Political Economy* 82, 1119–1143.

- Gulotty, R., Yu, A.Z., 2023. Must watch propaganda: the marginal treatment effect of foreign media among always-takers. *Political Science Research and Methods* 1–18.
- Gupta, A., Howell, S.T., Yannelis, C., Gupta, A., 2024. Owner incentives and performance in healthcare: private equity investment in nursing homes. *The Review of Financial Studies* 37, 1029–1077.
- Hahn, J., Kuersteiner, G., Santos, A., Willigrod, W., 2024. Overidentification in shift-share designs.
- Hahn, J., Todd, P., Van der Klauuw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69, 201–209.
- Han, S., Kaido, H., 2024. Set-valued control functions.
- Han, S., Yang, S., 2024. A computational approach to identification of treatment effects for policy evaluation. *Journal of Econometrics* 240, 105680.
- Hansen, B., 2022a. Probability and Statistics for Economists. Princeton University Press.
- Hansen, B.E., 2022b. Econometrics. Princeton University Press.
- Hansen, C., Hausman, J., Newey, W., 2008. Estimation with many instrumental variables. *Journal of Business & Economic Statistics* 26, 398–422.
- Heckman, J., 1974. Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–694.
- Heckman, J., 1997. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *The Journal of Human Resources* 32, 441–462.
- Heckman, J., Hohmann, N., Smith, J., Khoo, M., 2000. Substitution and dropout bias in social experiments: a study of an influential social experiment. *The Quarterly Journal of Economics* 115, 651–694.
- Heckman, J., Vytlacil, E., 1998. Instrumental variables methods for the correlated random coefficient model: estimating the average rate of return to schooling when the return is correlated with schooling. *The Journal of Human Resources* 33, 974–987.
- Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*.
- Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica* 47, 153–161.
- Heckman, J.J., 2010. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature* 48, 356–398.
- Heckman, J.J., Honoré, B.E., 1990. The empirical content of the roy model. *Econometrica* 58, 1121–1149.
- Heckman, J.J., Humphries, J.E., Veramendi, G., 2016. Dynamic treatment effects. *Journal of Econometrics* 191, 276–292.
- Heckman, J.J., Humphries, J.E., Veramendi, G., 2018. Returns to education: the causal effects of education on earnings, health, and smoking. *Journal of Political Economy* 126, S197–S246.
- Heckman, J.J., Lalonde, R.J., Smith, J.A., 1999. Chapter 31 The Economics and Econometrics of Active Labor Market Programs. Elsevier, pp. 1865–2097 vol. Volume 3, Part A.
- Heckman, J.J., Pinto, R., 2018. Unordered monotonicity. *Econometrica* 86, 1–35.
- Heckman, J.J., Robb, R., 1985. Alternative methods for evaluating the impact of interventions. In: Heckman, J.J., Singer, B. (Eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge University Press.
- Heckman, J.J., Smith, J., Clements, N., 1997. Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. *The Review of Economic Studies* 64, 487–535.
- Heckman, J.J., Urzua, S., 2010. Comparing IV with structural models: what simple IV can and cannot identify. *Journal of Econometrics* 156, 27–37.

- Heckman, J.J., Urzua, S., Vytlacil, E., 2006. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics* 88, 389–432.
- Heckman, J.J., Urzua, S., Vytlacil, E., 2008. Instrumental variables in models with multiple outcomes: the general unordered case. *Annales daEconomie et de Statistique* 91/92, 151–174.
- Heckman, J.J., Vytlacil, E., 2001a. Policy-relevant treatment effects. *The American Economic Review* 91, 107–111.
- Heckman, J.J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. *Econometrica* 73, 669–738.
- Heckman, J.J., Vytlacil, E.J., 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4730–4734.
- Heckman, J.J., Vytlacil, E.J., 2001b. Instrumental variables, selection models, and tight bounds on the average treatment effect. In: Lechner, M., Pfeiffer, F. (Eds.), *Econometric Evaluations of Active Labor Market Policies in Europe*. Physica, Heidelberg and Berlin.
- Heckman, J.J., Vytlacil, E.J., 2007a. Chapter 70 Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation. In: Heckman, J.J., Leamer, E.E. (Eds.), *Handbook of Econometrics*. Elsevier, pp. 4779–4874. Volume 6, Part 2.
- Heckman, J.J., Vytlacil, E.J., 2007b. Chapter 71 Econometric evaluation of social programs, Part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In: Heckman, J.J., Leamer, E.E. (Eds.), *Handbook of Econometrics*. Elsevier, pp. 4875–5143. Volume 6, Part 2.
- Heiler, P., 2022. Efficient covariate balancing for the local average treatment effect. *Journal of Business & Economic Statistics* 40, 1569–1582.
- Heinesen, E., Hvid, C., Kirkebøen, L., Leuven, E., Mogstad, M., 2022. Instrumental variables with unordered treatments: theory and evidence from returns to fields of study.
- Heinesen, E., Stenholte Lange, E., 2022. Vocational versus general upper secondary education and earnings. *Journal of Human Resources* 0221–11497R2.
- Heldring, L., Robinson, J.A., Vollmer, S., 2022. The economic effects of the english parliamentary enclosures.
- Hirano, K., Imbens, G.W., Rubin, D.B., Zhou, X.-H., 2000. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 1, 69–88.
- Hojman, A., LopezBoo, F., 2022. Public childcare benefits children and mothers: evidence from a nationwide experiment in a developing country. *Journal of Public Economics* 212, 104686.
- Hong, H., Nekipelov, D., 2010. Semiparametric efficiency in nonlinear LATE models. *Quantitative Economics* 1, 279–304.
- Huber, M., Laffers, L., Mellace, G., 2015. Sharp IV bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance. *Journal of Applied Econometrics* 32, 56–79.
- Huber, M., Mellace, G., 2014. Testing instrument validity for LATE identification based on inequality moment constraints. *Review of Economics and Statistics* 97, 398–411.
- Hull, P., 2020. Estimating hospital quality with quasi-experimental data, SSRN Scholarly Paper ID 3118358, Social Science Research Network, Rochester, NY.
- Humlum, A., Munch, J., Rasmussen, M., 2023. What works for the unemployed? Evidence from quasi-random caseworker assignments.
- Humphries, J.E., Joensen, J.S., Veramendi, G.F., 2023a. Complementarities in high school and college investments. SSRN Electronic Journal.

- Humphries, J.E., Ouss, A., Stavreva, K., Stevenson, M.T., van Dijk, W., 2023b. Conviction, incarceration, and recidivism: understanding the revolving door.
- Imbens, G.W., 2010. Better LATE than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature* 48, 399–423.
- Imbens, G.W., 2015. Matching methods in practice: three examples. *Journal of Human Resources* 50, 373–419.
- Imbens, G.W., 2022. Causality in econometrics: choice vs chance. *Econometrica* 90, 2541–2566.
- Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. *Econometrica* 62, 467–475.
- Imbens, G.W., Newey, W.K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77, 1481–1512.
- Imbens, G.W., Rubin, D.B., 1997. Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies* 64, 555–574.
- Ito, K., Ida, T., Tanaka, M., 2023. Selection on welfare gains: experimental evidence from electricity plan choice. *American Economic Review* 113, 2937–2973.
- Joensen, J.S., Nielsen, H.S., 2016. Mathematics and gender: heterogeneity in causes and consequences. *The Economic Journal* 126, 1129–1163.
- Johar, M., Maruyama, S., 2014. Does coresidence improve an elderly parent's health? Does coresidence improve an elderly patient's health? *Journal of Applied Econometrics* 29, 965–983.
- Kamat, V., 2021. On the identifying content of instrument monotonicity.
- Kamat, V., 2024. Identifying the effects of a program offer with an application to head start. *Journal of Econometrics* 240, 105679.
- Kamat, V., Norris, S., Pecenco, M., 2024. Conviction, incarceration, and policy effects in the criminal justice system.
- Kane, T.J., Rouse, C.E., 1995. Labor-market returns to two- and four-year college. *The American Economic Review* 85, 600–614.
- Kang, J.D.Y., Schafer, J.L., 2007. Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data 22, 523–539.
- Kasahara, H., Liang, Y., Rodrigue, J., 2016. Does importing intermediates increase the demand for skilled workers? Plant-level evidence from Indonesia. *Journal of International Economics* 102, 242–261.
- Kaufmann, K.M., 2014. Understanding the income gradient in college attendance in mexico: the role of heterogeneity in expected returns: income gradient in college attendance. *Quantitative Economics* 5, 583–630.
- Kédagni, D., Mourifié, I., 2020. Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika* 107, 661–675.
- Kirkeboen, L.J., Leuven, E., Mogstad, M., 2016. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics* 131, 1057–1111.
- Kitagawa, T., 2009. Identification region of the potential outcome distributions under instrument independence, Cemmap working paper.
- Kitagawa, T., 2015. A test for instrument validity. *Econometrica* 83, 2043–2063.
- Kitagawa, T., 2021. The identification region of the potential outcome distributions under instrument independence. *Journal of Econometrics* 225, 231–253.
- Kline, P., Santos, A., 2013. Sensitivity to missing data assumptions: theory and an evaluation of the U.S. wage structure: sensitivity to missing data assumptions. *Quantitative Economics* 4, 231–267.
- Kline, P., Walters, C.R., 2016. Evaluating public programs with close substitutes: the case of head start*. *The Quarterly Journal of Economics* 131, 1795–1848.

- Kline, P., Walters, C.R., 2019. On heckits, LATE, and numerical equivalence. *Econometrica* 87, 677–696.
- Kling, J.R., 2006. Incarceration length, employment, and earnings. *American Economic Review* 96, 863–876.
- Kolesár, M., 2013. Estimation in an instrumental variables model with treatment effect heterogeneity.
- Kowalski, A.E., 2023a. Behaviour within a clinical trial and implications for mammography guidelines. *The Review of Economic Studies* 90, 432–462.
- Kowalski, A.E., 2023b. How to examine external validity within an experiment. *Journal of Economics & Management Strategy* 32, 491–509.
- Kowalski, A.E., 2023c. Reconciling seemingly contradictory results from the oregon health insurance experiment and the Massachusetts health reform. *Review of Economics and Statistics* 105, 646–664.
- Kowalski, A., 2016. Doing more when you're running LATE: applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments, NBER Working paper 22363.
- Kreider, B., Pepper, J.V., 2007. Disability and employment. *Journal of the American Statistical Association* 102, 432–441.
- Kreider, B., Pepper, J.V., Gundersen, C., Jolliffe, D., 2012. Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported. *Journal of the American Statistical Association* 107, 958–975.
- Lee, D.S., 2008. Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics* 142, 675–697.
- Lee, K., Miguel, E., Wolfram, C., 2019. Experimental evidence on the economics of rural electrification. *Journal of Political Economy* 128, 1523–1565.
- Lee, S., 2018. A consistent variance estimator for 2SLS when instruments identify different LATEs. *Journal of Business & Economic Statistics* 36, 400–410.
- Lee, S., Salanié, B., 2018. Identifying effects of multivalued treatments. *Econometrica* 86, 1939–1963.
- Lee, S., Salanié, B., 2023. Treatment effects with targeting instruments.
- Leung, P., O'Leary, C., 2020. Unemployment insurance and means-tested program interactions: evidence from administrative data. *American Economic Journal: Economic Policy* 12, 159–192.
- Lewis, H.G., 1974. Comments on selectivity biases in wage comparisons. *Journal of Political Economy* 82, 1145–1155.
- Li, H., Liu, Y., Zhao, X., Zhang, L., Yuan, K., 2021a. Estimating effects of cooperative membership on farmers' safe production behaviors: evidence from the rice sector in China. *Environmental Science and Pollution Research* 28, 25400–25418.
- Li, M., Jin, T., Liu, S., Zhou, S., 2021b. The cost of clean energy transition in rural China: Evidence based on marginal treatment effects. *Energy Economics* 97, 105167.
- Liu, S., Jin, T., Li, M., Zhou, S., 2022. Fertility policy adjustments and female labor supply: estimation of marginal treatment effect using Chinese census data. *Journal of Human Resources*.
- Loeser, J., 2023. Modeling selection into unordered treatments: an equivalence result, Working Paper.
- Machado, C., Shaikh, A.M., Vytlacil, E.J., 2019. Instrumental variables and the sign of the average treatment effect. *Journal of Econometrics*.

- MacCurdy, T., Chen, X., Hong, H., 2011. Flexible estimation of treatment effect parameters. *American Economic Review* 101, 544–551.
- Maestas, N., Mullen, K.J., Strand, A., 2013. Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt. *The American Economic Review* 103, 1797–1829.
- Manda, J., Alene, A.D., Tufa, A.H., Abdoulaye, T., Kamara, A.Y., Olufajo, O., Boukar, O., Manyong, V.M., 2020. Adoption and ex-post impacts of improved cowpea varieties on productivity and net returns in Nigeria. *Journal of Agricultural Economics* 71, 165–183.
- Manski, C., 1994. The selection problem. *Advances in Econometrics*, Sixth World Congress, vol. 1. pp. 143–170.
- Manski, C.F., 1989. Anatomy of the selection problem. *The Journal of Human Resources* 24, 343–360.
- Manski, C.F., 1990. Nonparametric bounds on treatment effects. *The American Economic Review* 80, 319–323.
- Manski, C.F., 1997. Monotone treatment response. *Econometrica* 65, 1311–1334.
- Manski, C.F., 2003. Partial Identification of Probability Distributions. Springer.
- Manski, C.F., 2007. Partial identification of counterfactual choice probabilities. *International Economic Review* 48, 1393–1410.
- Manski, C.F., Pepper, J.V., 2000. Monotone instrumental variables: with an application to the returns to schooling. *Econometrica* 68, 997–1010.
- Marshall, J., 2016. Coarsening bias: how coarse treatment measurement upwardly biases instrumental variable estimates. *Political Analysis* 24, 157–171.
- Martinez-Iriarte, J., Sun, Y., 2024. Identification and estimation of unconditional policy effects of an endogenous binary treatment: an unconditional MTE approach.
- Marx, B.M., Turner, L.J., 2019. Student loan nudges: experimental evidence on borrowing and educational attainment. *American Economic Journal: Economic Policy* 11, 108–141.
- Marx, P., 2024. Sharp bounds in the latent index selection model. *Journal of Econometrics* 238, 105561.
- Masten, M.A., Poirier, A., 2020. Inference on breakdown frontiers. *Quantitative Economics* 11, 41–111.
- Masten, M.A., Poirier, A., 2021. Salvaging falsified instrumental variable models. *Econometrica* 89, 1449–1469.
- Masten, M.A., Torgovitsky, A., 2014. Instrumental variables estimation of a generalized correlated random coefficients model, cemmap working paper 02/14.
- Masten, M.A., Torgovitsky, A., 2016. Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics* 98, 1001–1005.
- Matzkin, R.L., 2003. Nonparametric estimation of nonadditive random functions. *Econometrica* 71, 1339–1375.
- Mellon Bedi, S., Azzarri, C., HundieKotu, B., Kornher, L., vonBraun, J., 2021. Scaling-up agricultural technologies: who should be targeted? *European Review of Agricultural Economics* jbab0 54.
- Miyaji, S., 2024a. Instrumented difference-in-differences with heterogeneous treatment effects.
- Miyaji, S., 2024b. Two-way fixed effects instrumental variable regressions in staggered DID-IV designs.
- Moffitt, R., 2008. Estimating marginal treatment effects in heterogeneous populations. *Annales d'Economie et de Statistique* 239–261.
- Moffitt, R.A., 2019. The marginal labor supply disincentives of welfare reforms, Working Paper 26028, National Bureau of Economic Research.

- Moffitt, R.A., Zahn, M.V., 2019. The marginal labor supply disincentives of welfare: evidence from administrative barriers to participation.
- Mogstad, M., Santos, A., Torgovitsky, A., 2017. Using instrumental variables for inference about policy relevant treatment parameters, NBER Working Paper.
- Mogstad, M., Santos, A., Torgovitsky, A., 2018. Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica* 86, 1589–1619.
- Mogstad, M., Torgovitsky, A., 2018. Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics* 10.
- Mogstad, M., Torgovitsky, A., Walters, C.R., 2021. The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review* 111, 3663–3698.
- Mogstad, M., Torgovitsky, A., Walters, C.R., 2024. Policy evaluation with multiple instrumental variables. *Journal of Econometrics*, 105718.
- Molinari, F., 2020. Microeconomics with partial identification, arXiv:2004.11751 [econ].
- Mountjoy, J., 2022. Community colleges and upward mobility. *American Economic Review* 112, 2580–2630.
- Mourifié, I., Wan, Y., 2016. Testing local average treatment effect assumptions. *The Review of Economics and Statistics* 99, 305–313.
- Mueller-Smith, M., 2015. The criminal and labor market impacts of incarceration, Working Paper, 59.
- Navjeevan, M., Pinto, R., Santos, A., 2023. Identification and estimation in a class of potential outcomes models.
- Newey, W., Stouli, S., 2021. Control variables, discrete instruments, and identification of structural functions. *Journal of Econometrics* 222, 73–88.
- Newey, W.K., 1994. The asymptotic variance of semiparametric estimators. *Econometrica* 62, 1349–1382.
- Noack, C., 2021. Sensitivity of LATE estimates to violations of the monotonicity assumption.
- Nobel Committee, 2021. Scientific background on the sweriges riksbank prize in economic sciences in Memory of Alfred Nobel 202, 48.
- Norris, S., Pecenco, M., Weaver, J., 2021. The Effects Of Parental And Sibling Incarceration: Evidence from Ohio. *American Economic Review* 111, 2926–2963.
- Nyblom, M., 2017. The distribution of lifetime earnings returns to college. *Journal of Labor Economics* 000-000.
- Ogburn, E.L., Rotnitzky, A., Robins, J.M., 2015. Doubly Robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77, 373–396.
- Pearl, J., 2011. Principal stratification - a goal or a tool? *The International Journal of Biostatistics* 7, 1–13.
- Pepper, J.V., 2000. The intergenerational transmission of welfare receipt: a nonparametric bounds analysis. *The Review of Economics and Statistics* 82, 472–488.
- Permutt, T., Hebel, J.R., 1989. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics* 45, 619–622.
- Phillips, G.D.A., Hale, C., 1977. The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review* 18, 219–228.
- Pinto, R., 2022. Beyond intention-to-treat: using the incentives of moving to opportunity to identify neighborhood effects, Working Paper.
- Poirier, A., Słoczyński, T., 2024. Quantifying the internal validity of weighted estimands.
- Possebom, V., 2023. Crime and mismeasured punishment: marginal treatment effect with misclassification. *The Review of Economics and Statistics* 1–42.

- Priebe, J., 2020. Quasi-experimental evidence for the causal link between fertility and subjective well-being. *Journal of Population Economics* 33, 839–882.
- Puhani, P., 2000. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* 14, 53–68.
- Ramsey, J.B., 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society: Series B (Methodological)* 31, 350–371.
- Rivera, R., 2023. Release, detail, or surveil?.
- Robins, J.M., 1989. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health Service Research Methodology: A Focus on AIDS* 113–159.
- Robins, J.M., Greenland, S., 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3, 143.
- Robins, J.M., Greenland, S., 1996. Identification of causal effects using instrumental variables: comment. *Journal of the American Statistical Association* 91, 456–458.
- Robinson, P.M., 1988. Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.
- Rose, E.K., Shem-Tov, Y., 2021. How does incarceration affect reoffending? Estimating the dose-response function. *Journal of Political Economy* 000–000.
- Rose, E.K., Shem-Tov, Y., 2023. On recoding ordered treatments as binary indicators.
- Roy, A.D., 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3, 135–146.
- Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688–701.
- Sargan, J.D., 1958. The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393–415.
- Sarr, M., BezabihAyele, M., Kimani, M.E., Ruhinduka, R., 2021. Who benefits from climate-friendly agriculture? The marginal returns to a rainfed system of rice intensification in Tanzania. *World Development* 138, 105160.
- Sasaki, Y., Ura, T., 2023. Estimation and inference for policy relevant treatment effects. *Journal of Econometrics* 234, 394–450.
- Shaikh, A.M., Vytlacil, E.J., 2011. Partial identification in triangular systems of equations with binary dependent variables. *Econometrica* 79, 949–955.
- Shea, J., Torgovitsky, A., 2023. Ivtmte: an R package for extrapolating instrumental variable estimates away from compliers*. *Observational Studies* 9, 1–42.
- Shurtz, I., Eizenberg, A., Alkalay, A., Lahad, A., 2022. Physician workload and treatment choice: the case of primary care. *The RAND Journal of Economics* 53, 763–791.
- Siddique, Z., 2013. Partially identified treatment effects under imperfect compliance: the case of domestic violence. *Journal of the American Statistical Association* 108, 504–513.
- Sigstad, H., 2024a. Marginal treatment effects and monotonicity.
- Sigstad, H., 2024b. Monotonicity among judges: evidence from judicial panels and consequences for judge IV designs. *SSRN Electronic Journal*.
- Singh, R., Sun, L., 2024. Double robustness for complier parameters and a semi-parametric test for complier characteristics. *The Econometrics Journal* 27, 1–20.
- Słoczyński, T., 2020. When should we (not) interpret linear IV estimands as LATE?.
- Słoczyński, T., 2022. Interpreting OLS estimands when treatment effects are heterogeneous: smaller groups get larger weights. *The Review of Economics and Statistics* 104, 501–509.
- Słoczyński, T., Uysal, S.D. Wooldridge, J.M., 2022. Doubly Robust estimation of local average treatment effects using inverse probability weighted regression adjustment.

- Słoczyński, T., Uysal, S.D., Wooldridge, J.M., 2024. Abadieas kappa and weighting estimators of the local average treatment effect. *Journal of Business & Economic Statistics* 1–14.
- Słoczyński, T., Wooldridge, J.M., 2018. A general double robustness result for estimating average treatment effects. *Econometric Theory* 34, 112–133.
- Splawa-Neyman, J., Dabrowska, D.M., Speed, T.P., 1990. On the application of probability theory to agricultural experiments. essay on principles. Section 9. *Statistical Science* 5, 465–472.
- Stevenson, M.T., 2018. Distortion of justice: how the inability to pay bail affects case outcomes. *The Journal of Law, Economics, and Organization* 34, 511–542.
- Sun, B., Tan, Z., 2022. High-dimensional model-assisted inference for local average treatment effects with instrumental variables. *Journal of Business & Economic Statistics* 40, 1732–1744.
- Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics* 225, 175–199.
- Sun, Z., 2023. Instrument validity for heterogeneous causal effects. *Journal of Econometrics* 237, 105523.
- Swanson, S.A., Hernán, M.A., 2014. Think globally, act globally: an epidemiologist's perspective on instrumental variable estimation. *Statistical Science* 29, 371–374.
- Tan, Z., 2006. Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association* 101, 1607–1618.
- Tebaldi, P., Torgovitsky, A., Yang, H., 2023. Nonparametric estimates of demand in the california health insurance exchange. *Econometrica* 91, 107–146.
- Theil, H., 1971. *Principles of Econometrics*. John Wiley & Sons.
- Thornton, R.L., 2008. The demand for, and impact of, learning HIV status. *American Economic Review* 98, 1829–1863.
- Tommasi, D., Zhang, L., 2024. Bounding program benefits when participation is misreported. *Journal of Econometrics* 238, 105556.
- Torgovitsky, A., 2015. Identification of nonseparable models using instruments with small support. *Econometrica* 83, 1185–1197.
- Torgovitsky, A., 2017. Minimum distance from independence estimation of nonseparable instrumental variables models. *Journal of Econometrics* 199, 35–48.
- Train, K.E., 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Ura, T., 2018. Heterogeneous treatment effects with mismeasured endogenous treatment. *Quantitative Economics* 9, 1335–1370.
- Ura, T., Zhang, L., 2024. Policy relevant treatment effects with multidimensional unobserved heterogeneity.
- Uysal, S.D., 2011. Three essays on doubly robust estimation methods.
- van't Hoff, N., Lewbel, A., Mellace, G., 2024. Limited monotonicity and the combined compliers LATE, Boston College Working Papers in Economics.
- Vella, F., 1998. Estimating models with sample selection bias: a survey. *The Journal of Human Resources* 33, 127–169.
- Vohra, V., J. Goldin, J., 2024. Identifying the cumulative causal effect of a non-binary treatment from a binary instrument.
- Vuong, Q., Xu, H., 2017. Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics* 8, 589–610.
- Vytlačil, E., 2002. Independence, monotonicity, and latent index models: an equivalence result. *Econometrica* 70, 331–341.
- Vytlačil, E., 2006. Ordered discrete-choice selection models and local average treatment effect assumptions: equivalence, nonequivalence, and representation results. *The Review of Economics and Statistics* 88, 578–581.

- Wald, A., 1940. The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics* 11, 284–300.
- Walters, C.R., 2018. The demand for effective charter schools. *Journal of Political Economy* 126, 2179–2223.
- Wang, W., Ida, T., Shimada, H., 2020. Default effect versus active decision: evidence from a field experiment in Los Alamos. *European Economic Review* 128, 103498.
- Westphal, M., Kamhöfer, D.A., Schmitz, H., 2022. Marginal college wage premiums under selection into employment. *The Economic Journal* 132, 2231–2272.
- Wilding, A., Munford, L., Sutton, M., 2023. Estimating the heterogeneous health and well-being returns to social participation. *Health Economics* 32, 1921–1940.
- Willis, R.J., Rosen, S., 1979. Education and self-selection. *Journal of Political Economy* 87, S7–S36.
- Windmeijer, F., 2019. Two-stage least squares as minimum distance. *The Econometrics Journal* 22, 1–9.
- Wooldridge, J.M., 1997. On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters* 56, 129–133.
- Wooldridge, J.M., 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters* 79, 185–191.
- Wooldridge, J.M., 2008. Instrumental variables estimation of the average treatment effect in correlated random coefficient models. In: Millimet, D., Smith, J., Vytlacil, E. (Eds.), *Modeling and Evaluating Treatment Effects in Econometrics*. Else.
- Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Wooldridge, J.M., 2015. Control function methods in applied econometrics. *Journal of Human Resources* 50, 420–445.
- Xie, H., 2024. Efficient and Robust estimation of the generalized LATE model. *Journal of Business & Economic Statistics* 42, 1053–1065.
- Yau, L.H.Y., Little, R.J., 2001. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* 96, 1232–1244.
- Zeng, S., Li, F., Ding, P., 2020. Is being an only child harmful to psychological health?: Evidence from an instrumental variable analysis of Chinaas one-child policy. *Journal of the Royal Statistical Society Series A: Statistics in Society* 183, 1615–1635.

Chapter 2

Firm wage effects[☆]

Patrick Kline*

University of California, Berkeley, USA

*Corresponding author. e-mail address: pkline@berkeley.edu

Chapter Outline

1 Background	119	4.3 Clustering approaches	156
2 What sorts of firms pay high wages?	120	4.4 How variable are worker and firm effects?	159
2.1 Productivity, worker flows, and firm size	121	5 Regressing firm effects on observables	161
2.2 Entry, reallocation, and dynamics	122	5.1 One step vs two	162
2.3 Sorting, outsourcing, and displacement	123	5.2 Variance estimation	163
2.4 Industry structure and amenities	124	5.3 Revisiting the firm size wage premium	164
3 The AKM model	125	6 Hiring origins and state dependence	166
3.1 An edgy interpretation of firm effects	127	6.1 Structural interpretation	167
3.2 Evaluating the AKM restrictions	135	6.2 Testable restrictions	168
3.3 Causality	140	6.3 It ain't where you're from, it's where you're at	169
4 Variance decomposition	146	6.4 Information and conduct	170
4.1 Limited mobility bias	147	7 Conclusion	171
4.2 Cross-fitting and bias correction	148	Appendix: Covariance between person and firm effects	172
		References	176

Nearly a century of empirical study supports the view that employers offer different wages for identical work. Fueled by the dissemination of linked employer-employee datasets, a rapidly advancing literature seeks to quantify

☆ This is the first part of a larger chapter on the topic of “wage setting power” that was initially prepared for the Handbook of Labor Economics conference in Berlin, which was generously funded by the Rockwool Foundation Berlin (RFBerlin). Based on discussions with the editors it was determined to be better for pedagogical reasons to publish the two parts as separate chapters. I thank David Card, Raffaele Saggio, Ben Scuderi, Isaiah Andrews, and Sophie Sun for helpful comments on an earlier draft that substantially improved the paper. Jordan Cammarota provided outstanding research assistance on this project. Many of the ideas in this chapter grew out of past conversations with collaborators including David Card, Raffaele Saggio, and Mikkel Sølvsten.

the role of firms in generating wage inequality using high dimensional fixed effects methods. This paper provides an overview of the literature on firm wage effects, summarizing the evidence base that has been accumulated on which firms pay high wages, their contribution to inequality, and econometric issues that arise in working with models of firm wage fixed effects.

The paper begins with a survey of early empirical investigations of firm and industry components of wage dispersion. [Slichter \(1950\)](#) and [Stigler \(1962\)](#) pioneered the measurement of wage dispersion across employers, providing estimates of the variability of posted wages within narrowly defined job categories and the variability of wage offers within the same worker. Generations later, [Krueger and Summers \(1988\)](#) used the panel structure of large surveys to study the wage changes accompanying worker mobility between industries, concluding that substantial across industry dispersion is present in average pay for equivalent work. These findings renewed interest in deviations from competitive labor market models and foreshadowed many of the economic and econometric debates surrounding the use of fixed effects methods today ([Katz et al., 1989](#); [Murphy and Topel, 1990](#); [Gibbons and Katz, 1992](#)). A related literature on firm size wage premia and intra-industry dispersion documented sizable wage differences across firms and plants in the same industry ([Brown and Medoff, 1989](#); [Brown et al., 1990](#); [Groshen, 1991](#); [Cappelli and Chaquin, 1991](#)). [Abowd et al. \(1999\)](#)'s landmark study provided a unified framework for studying these phenomena by applying high dimensional fixed effects methods to matched employer-employee data.

A large empirical literature has refined and extended many of the conclusions from [Abowd et al. \(1999\)](#)'s paper. Five notable patterns stand out from this literature. First, consistent with standard job ladder models, firm wage fixed effects have been found to be positively related to proxies of firm productivity, firm size, and revealed preference measures of firm desirability ([Card et al., 2016](#); [Bloom et al., 2018](#); [Sorkin, 2018](#); [Crane et al., 2023](#)). Second, high wage firms tend to employ high wage workers, men, and workers with greater educational attainment ([Card et al., 2013, 2016](#)). Third, firm wage effects are highly temporally persistent ([Lachowska et al., 2023](#); [Engbom et al., 2023](#)), and a handful of studies suggest changes in labor market institutions can alter the mix of firm effects in an economy ([Card et al., 2013](#); [Dustmann et al., 2022](#)). Fourth, high wage firms are more likely to “fissure”, or outsource jobs, and to conduct mass layoffs, both of which may indicate that firms face horizontal equity constraints in wage setting ([Goldschmidt and Schmieder, 2017](#); [Bertheau et al., 2023](#)). Fifth, the latest research suggests that the most productive firms also provide the best amenities, aligning with revealed preference evidence that high wage firms tend to be more desirable than low wage firms ([Sorkin, 2018](#); [Lamadon et al., 2022](#); [Sorkin, 2022](#); [Roussille and Scuderi, 2023](#); [Maestas et al., 2023](#); [Lehmann, 2023](#); [Caldwell et al., 2024b](#)).

Delving into the econometric assumptions underlying many of these studies, we review “the AKM model”: a two-way fixed effects model of wage

determination allowing for unrestricted worker-firm sorting patterns. After discussing the standard identification requirements of two-way fixed effects estimators in matched employer-employee data, a graph theoretic interpretation of the model is introduced where firms serve as vertices and the wage changes between employers constitute directed edges. The restrictions that the AKM model places on these edges are explained, highlighting the special role played by cycles in the mobility network. These restrictions yield a complex mapping between wage changes and firm effects; however, pruning the mobility graph to a spanning tree yields a just-identified set of firm effects with a particularly simple structure. The plausibility of the AKM model restrictions is then evaluated empirically in a benchmark dataset. After accounting for noise in the edge specific wage changes, I find that the least squares estimates provide a remarkably accurate (albeit imperfect) summary of the wage changes associated with moving between particular pairs of firms.

Building on the graph theoretic interpretation, I introduce non-parametric assumptions that endow the wage changes accompanying worker mobility with a causal interpretation. Difficulties arise with aggregating these causal effects into a global ranking of firm wage levels. Least squares estimates of firm effects are shown to rely on “indirect contrasts” involving mobility between other firm pairs than those under consideration, a phenomenon that has been found to also arise in other settings with multiple treatments (Goldsmith-Pinkham et al., 2022). Indirect contrasts can be avoided when the network is pruned to a tree but least squares estimates of firm effects do not automatically allow global comparison of wage levels across firms without further assumptions. The section concludes by proposing an assumption that ensures a transitive ranking of firm wage levels and discussing how this assumption might be usefully weakened in future research.

Abowd et al. (1999) proposed a now canonical variance decomposition of log wages into components attributable to worker and firm heterogeneity and sorting. Plugging estimated fixed effects into variance decompositions has long been understood to generate important biases (Krueger and Summers, 1988; Andrews et al., 2008). I review approaches to circumventing these biases, including the cross-fitting based bias-correction of Kline et al. (2020) and recently proposed clustering methods that assume the firm heterogeneity possesses a lower dimensional structure (Bonhomme et al., 2019, 2023). Cross-fitting approaches require a substantial amount of worker mobility, which researchers typically enforce by pruning to a set of “leave-out connected” firms. A simple approach to bounding the influence of this pruning step on the estimand is proposed and applied to a well known benchmark dataset. I also discuss imputation strategies that can be used to address concerns about biases arising either from pruning or neglected serial correlation. An empirical investigation suggests that the selection biases associated with pruning and serial correlation are likely minimal in large administrative datasets.

Reviewing the empirical literature on bias corrected variance decompositions, I argue that interest ought to center on the magnitude of these variance components themselves rather than variance shares, which are difficult to compare across datasets with different intrinsic noise levels. Bias corrected estimates of the economic magnitude of the variability in firm fixed effects are typically sizable, relative both to the dispersion in person effects and to the effect sizes of human capital interventions. A review of recent studies yields estimated standard deviations of firm fixed effects ranging from 15 to 60 log points, with estimates in the US and European countries clustering around 20 log points. In line with a growing literature on labor market misallocation (e.g., [Hsieh and Klenow, 2009](#)), dispersion in firm effects appears to be most pronounced in the least developed countries. Investigating the factors driving this relationship between dispersion and development is a fruitful area for future research.

A virtue of fixed effects methods is that the estimates can be shared with other research teams who can explore other hypotheses about the relationship between the latent effects and observables. I review the logic of “two-step” regressions of estimated fixed effects on observables, contrasting it with one-step approaches predicated on stronger random effects assumptions. While second step regressions of estimated firm fixed effects on firm and worker level covariates are unbiased, inference is complicated by correlation across the fixed effects estimates, a problem that is well understood theoretically but has largely been ignored in applied work. [Kline et al. \(2020\)](#) proposed an approach to obtaining heteroscedasticity robust standard errors reflecting the uncertainty stemming from the error underlying the linear fixed effects model. I illustrate this approach with an application to the firm size wage premium, which is found to vary in complex ways across Italian regions. Naive two-step standard errors, of the sort that currently pervade the empirical literature, are found to significantly underestimate the true uncertainty present in averages of firm fixed effects in this example.

Finally, I discuss connections between the AKM model and the influential class of search models pioneered by [Postel-Vinay and Robin \(2002a,b\)](#). While these “sequential auction” models have traditionally been assessed based on their ability to jointly explain job mobility and wage dynamics within firm matches, we discuss the theory’s implications for hiring wages. [Di Addario et al. \(2023\)](#) showed that a simple linear specification allowing fixed effects for hiring origins nests the reduced form of hiring wages in the sequential auction model of [Bagger et al. \(2014\)](#). Dispersion in these hiring origin fixed effects can be viewed as capturing a contribution of search frictions (or equivalently, “luck”) to wage inequality. While recent evidence suggests that hiring origins are less influential than these models predict, bilateral competition between firms undoubtedly plays an important role in wage determination for some types of jobs. I discuss the importance for future work of allowing departures from the full information benchmark underpinning canonical variants of this competition framework and conclude with some directions for future research on the econometrics and economics of firm wage setting.

1 Background

Economists have long been aware that employers differ in the pay offered to equivalent workers. [Slichter \(1950\)](#) showed in survey data that the hourly wages of narrowly defined manual occupations varied widely across employers in Boston. Studying industry data from the 1950 Economic Census, he found that industry value added and profits were important drivers of average pay, leading him to conclude that managerial practices were an important determinant of industry pay setting. A decade later, [Stigler \(1962\)](#) collected data on the job offers of business school graduates. In one of the earliest analyses of matched employer-employee data, he documented that within occupational categories, the dispersion of wage offers across companies was of the same order of magnitude as dispersion of wage offers within individual. Moreover, these company pay differences were found to be persistent across years. He concluded from this evidence that wage dispersion for equivalent workers “is of the order of magnitude of 5–10 % even in so well organized a market as that of college graduates at a single university” ([Stigler, 1962](#), p. 96).

Generations later, [Krueger and Summers \(1988\)](#) examined the extent to which industry differences in pay reflected the sorting of high ability workers to high paying sectors. Using the 1984 Current Population Survey, they fit linear models with worker quality controls and industry fixed effects, finding a bias corrected standard deviation across two-digit industries of industry fixed effects in wages of 14 log points and a standard deviation in total compensation of roughly 18 log points. To account for unobserved differences in worker quality, they fit longitudinal models to the 1984 displaced workers survey, finding that including worker fixed effects had little impact on estimates of one digit industry fixed effects, suggesting a limited role for selection on unobserved worker quality. Corroborating this view, [Gibbons and Katz \(1992\)](#) found sizable industry wage differentials even after restricting to transitions induced by mass layoffs or plant closures. A large literature debated the interpretation of these findings and whether they can be attributed to compensating differentials, efficiency wages, or employer learning ([Katz et al., 1989; Murphy and Topel, 1990; Holzer et al., 1991; Gibbons et al., 2005](#)).

Several authors also studied wage differences between establishments and firms of different size ([Oi and Idson, 1999](#)). [Brown and Medoff \(1989\)](#), [Brown et al. \(1990\)](#), and [Oi and Idson \(1999\)](#) showed that larger firms, and larger plants within large firms, paid higher wages. Studying worker switches between establishments again confirmed that these differences were generally not attributable to unobserved worker characteristics. Adjustments for workplace amenities were also found to have little impact on the firm size wage premium. Corroborating evidence from [Groshen \(1991\)](#) and [Cappelli and Chauvin \(1991\)](#) documented large wage dispersion across establishments within industry that could not be explained by differences in measured human

capital. These intra-industry employer differentials were shown to be comparable in magnitude to inter-industry wage differences.

Seeking to unify these findings, [Abowd et al. \(1999\)](#) – henceforth, AKM – studied employer wage differences in large administrative panels from France and the United States featuring worker and firm identifiers. In what may have been the first high dimensional regression in labor economics, they fit linear models allowing a separate fixed effect for each worker and each firm, along with firm specific trends intended to capture heterogeneity in firm seniority trajectories. AKM found that estimated firm wage effects varied substantially across firms and were correlated with observable measures of firm productivity. However, the estimates suggested that worker and firm fixed effects were only modestly positively correlated and that industry and firm size wage premia were largely accounted for by differences in person effects. Unfortunately, shortly after their study was published, subsequent work revealed that some of these empirical conclusions were artifacts of an inaccurate approximation to the full least squares solution ([Abowd et al., 2002, 2003](#)).

Despite these early stumbles, the work of [Abowd et al. \(1999\)](#) heralded an important transition in empirical labor economics towards interest in the development of econometric methods for the study of matched employer-employee data. While the literature on panel data econometrics traditionally treated fixed effects as nuisance parameters ([Chamberlain, 1984](#)), AKM viewed these effects as objects of direct interest. This perspective permeates the literature today. Rather than focus attention on the relationship between wages and a handful of observable firm characteristics such as size, sector, or productivity, labor economists now routinely apply fixed effects estimators to enormous administrative datasets in an attempt to “let the data speak” directly about which employers offer high or low wages. The relationship between employer wage fixed effects and low dimensional worker and firm observables can then be scrutinized in a second step, perhaps even by a different research team. While similar transitions from structured to unstructured data analysis have occurred in many other areas of empirical economics – see the chapter in this Handbook by [Walters \(2024\)](#) for some examples – the change has arguably been most dramatic in the literature on wage determination, where it has long been understood that wages vary meaningfully across employers in ways that are difficult to capture with the worker and firm characteristics measured in standard datasets.

2 What sorts of firms pay high wages?

Before delving into the econometrics of fixed effects models, it is useful to provide an overview of what has been learned about the types of firms that offer high wages from empirical research utilizing matched employer-employee data. This body of work has refined our empirical understanding of

traditional regularities such as the firm size and industry wage premiums, while also offering new insights into how labor market institutions, outsourcing practices, and job displacement contribute to wage inequality.

2.1 Productivity, worker flows, and firm size

The empirical literature finds that firm wage fixed effects are strongly associated both with observable measures of firm productivity and desirability. AKM's original study documented that firm wage effects were positively correlated with value added per worker and capital share. An updated analysis by [Abowd et al. \(2012\)](#) utilizing exact least squares solutions finds qualitatively similar patterns in more recent panels of French and US administrative data. Using Portuguese data on hourly wages merged to firm accounting data from Bureau Van Dijk, [Card et al. \(2016\)](#) documented that firm wage effects exhibit a "hockey stick" like relationship with log value added per worker, exhibiting a slope of essentially zero at very low levels of value added followed by a nearly constant elasticity relationship at higher levels. Subsequent work documents similar nonlinearities in Germany ([Bruns, 2019](#)), France ([Coudin et al., 2018](#)), Canada ([Li et al., 2023](#)), Hungary ([Boza and Reizer, 2024](#)), and Italy ([Di Addario et al., 2023](#)). Possible explanations for the hockey stick shape include the presence of binding wage floors that prohibit very low firm effects, the existence of a "competitive fringe" of less productive firms that engage in essentially competitive wage setting, and non-classical measurement error in value added per worker.

[Sorkin \(2018\)](#) devised a revealed preference measure of firm desirability based on the idea that a desirable firm hires workers from other desirable firms. The proposed measure, which is motivated by a wage posting model in the spirit of [Burdett and Mortensen \(1998\)](#), involves applying the Google PageRank algorithm ([Page et al., 1999](#)) to the network of job to job flows. [Sorkin \(2018\)](#) reports that his measure of firm desirability exhibits a correlation of roughly 0.54 with firm wage effects derived from quarterly earnings in Longitudinal Employer Household Dynamics (LEHD) data. [Crane et al. \(2023\)](#) also use LEHD data to show that firm wage fixed effects are strongly positively related to the "poaching rank" index of [Bagger and Lentz \(2019\)](#), which provides another revealed preference measure of firm desirability consistent with a class of sequential auction models that will be discussed below.

Firm wage fixed effects have been shown to be positively related to firm size and negatively related to quit rates ([Card et al., 2013; Bassier et al., 2022](#)). [Bloom et al. \(2018\)](#) study the changing nature of the firm size wage premium by fitting separate fixed effects models to the US Social Security Administration's Master Earning File in each of three time periods: 1980–1986, 1994–2000, and 2007–2013. In the first period, firm wage fixed effects are monotonically increasing in firm size, with an enormous 55 log point gap in average firm effects between companies with 15,000 or more employees and those with 1–10 employees. In later periods, the relationship between wages

and firm size grows more concave. In the final 2007–2013 sample, monotonicity appears to break down, with mean firm fixed effects estimated to be slightly higher among firms with 1000–2500 employees than at the largest firms. The pay gap between the largest and smallest firms falls to roughly 22 log points in this period. To date, little evidence is available regarding whether similar transitions have occurred in other countries.

2.2 Entry, reallocation, and dynamics

The distribution of firm effects has been shown to respond to changes in labor market institutions. [Card et al. \(2013\)](#) fit separate models to four overlapping 6–7 year intervals of German data spanning the period from 1985 to 2009. They find that the variance of firm wage effects roughly doubles over the course of their study. Most of the growth in dispersion of firm effects occurs in the latter two intervals, a period that saw a rapid liberalization of the German labor market. Analyzing cohorts of firms, they find that within cohort inequality in firm wage effects is roughly stable over time but newer cohorts of firms are more unequal.¹ Tying these cohort trends to the breakdown of the German collective bargaining system, they document that firms not covered by bargaining agreements are more likely to exhibit very low wage fixed effects.

[Song et al. \(2019\)](#) conduct a similar “rolling AKM” analysis making use of US social security records over the period 1978–2013. While they find that inequality increased dramatically across firms over this period, firm effect variances were surprisingly stable, suggesting the rise in between firm inequality was a consequence of increased worker-firm sorting. This discrepancy between the German and US results may have to do with differences in the institutional environment of these labor markets. The US has enjoyed a relatively stable regulatory environment over the period studied by [Song et al. \(2019\)](#), while post-unification Germany faced enormous pressure on its sectoral bargaining system that plausibly paved the way for the entry of very low wage firms ([Dustmann et al., 2014](#)). [Dustmann et al. \(2022\)](#) show that the enactment of a German minimum wage led low wage workers to reallocate to firms with higher wage fixed effects, and that German regions differentially exposed to the minimum wage hike experienced an increase in the average AKM fixed effect of surviving establishments.

The temporal stability of the firm effect variances among cohorts of German firms documented by [Card et al. \(2013\)](#) suggests that firm wage effects are persistent. [Lachowska et al. \(2023\)](#) used hourly wage data derived from Washington state UI records to measure this persistence more carefully. They estimate unrestricted firm fixed effects over pairs of adjacent years, yielding a sequence of fixed effects for each firm. Fitting an AR1 model to these estimates, they find a bias corrected autocorrelation of firm wage effects of 0.98.

¹ [Sorkin and Wallskog \(2023\)](#) find a similar pattern in US data, albeit without controlling for person effects.

Contemporaneous work by Engbom et al. (2023) finds that projecting firm wage effects derived from 8 year intervals onto the effects derived from pooling 32 years of Swedish wage data yields a slope of roughly 0.95, suggesting that wage fixed effects are highly stable among long lived firms.

2.3 Sorting, outsourcing, and displacement

High wage firms employ high wage workers. This pattern has been repeatedly documented in the form of positive bias corrected correlations between worker and firm fixed effects (Andrews et al., 2008; Kline et al., 2020; Bonhomme et al., 2023). However, the pattern is usually evident (albeit attenuated) from uncorrected estimates fit to population level administrative records. Card et al. (2013) and Song et al. (2019) both find that the uncorrected correlation between worker and firm fixed effects has increased in recent decades. Observable worker characteristics are also predictive of firm effects. Low wage firms tend to disproportionately employ women (Card et al., 2016), immigrants (Dostie et al., 2023), minorities (Gerard et al., 2021), younger workers (Kline et al., 2020), and workers with lower educational attainment (Card et al., 2013). Low wage firms are also typically intensive in jobs involving low wage occupations (Card et al., 2013; Goldschmidt and Schmieder, 2017) and tend to exhibit less complex job hierarchies (Huitfeldt et al., 2023).

Goldschmidt and Schmieder (2017) find that German firms with high wage fixed effects are more likely to outsource workers in food services, cleaning, security, and logistics (FCSL) occupations. One interpretation of this pattern is that firms face horizontal equity constraints making it difficult to tailor wages to the match surplus of individual workers. Consistent with this view, they estimate separate firm fixed effects for FCSL and non-FCSL workers at each employer and find that firms paying 10 % higher wages to non-FCSL workers tend to pay FCSL worker roughly 8 % higher wages (Goldschmidt and Schmieder, 2017, Figure A-8).² Rather than share a large wage premium with workers at all layers of the organization, firms tend to spin off jobs lying outside their area of core competency in order to economize on wage costs. In the wake of an outsourcing event, measured as a setting where many FCSL workers move from a “mother” establishment to the same “daughter” establishment specializing in FCSL services, the wage of outsourced workers plummet. This drop turns out to be almost entirely explained by the low fixed effects of establishments specializing in FCSL services. Goldschmidt and Schmieder (2017) argue that the growth of firms specializing in FCSL services is an important driver of German inequality consistent with the firm cohort patterns documented by Card et al. (2013).

In line with the German evidence on outsourcing, Lachowska et al. (2020) document in Washington state UI records that firms in the top quintile of firm

² Conducting a similar exercise in Argentine data, Drenik et al. (2023) find that firms paying regular workers 10 % higher wages pay temporary workers roughly 5 % higher wages.

fixed effects account for a disproportionate share of displaced workers. However, they find that 70 % of displaced workers move to employers with similar or better firm effects despite suffering wage losses. As a result, they estimate that firm fixed effects account for only 17 % of the earnings losses associated with job displacement; however, this share rises to roughly two thirds among the workers who move to lower wage employers upon displacement. [Schmieder et al. \(2023\)](#) find in German administrative records that nearly all of the average daily wage losses associated with displacement are explained by differences in firm effects. [Bertheau et al. \(2023\)](#) study a harmonized panel of seven European countries and find that between 35 % (in Spain) to 100 % (in Portugal) of the daily wage losses of job displacement after five years are explained by the loss of firm fixed effects. In a longer working paper ([Bertheau et al., 2022](#)), they conjecture that this variation across countries may be attributable to the intensity of active labor market policies, which they show turns out to strongly predict the magnitude of country specific wage losses. Like [Lachowska et al. \(2020\)](#), [Bertheau et al. \(2023\)](#) find in all seven countries that job displacement is most common among firms with estimated firm fixed effects in the top quintiles.

2.4 Industry structure and amenities

A headline finding of [Abowd et al. \(1999\)](#)'s original study was that industry wage differentials are largely explained by person effects. This conclusion turned out to have been driven by the computational method used in their analysis to approximate the least squares solution in the largest samples of firms ([Abowd et al., 2002](#)). Subsequent analysis of early LEHD data from four states found substantial differences in average firm effects across sectors ([Abowd et al., 2003](#), Table 11). [Sorkin \(2018, Table V\)](#) finds in a broader LEHD dataset comprised of large employers in 27 states that four digit industry codes account for roughly 45 % of the variation in firm fixed effects.

More recently, [Card et al. \(2024\)](#) analyze LEHD covering all 50 states for the years 2010–2018. They find that roughly one third of the variance in firm wage effects is explained by four digit NAICS industry codes. Remarkably, the average industry premiums are nearly identical for workers who have, and have not, obtained a college degree. They estimate that the highest paying industry is coal mining, while the lowest paying industry is drinking places. Perhaps surprisingly, their industry wage premia estimates turn out to be positively correlated with production function based estimates of industry wage markdowns from [Yeh et al. \(2022\)](#), which may indicate that variation in industry averages of firm wage effects reflect productivity more than market power.

[Card et al. \(2013\)](#) find in German data that between industry dispersion of firm effects rose between 1985 and 2009 and that high wage workers increasingly sort to high wage industries. In contrast, [Haltiwanger et al. \(2024\)](#), fitting AKM models to three intervals of LEHD data covering the period 1996–2018, find that the contribution of industry averages of firm wage fixed

effects to wage inequality has been relatively stable. Like [Card et al. \(2013\)](#), however, they find that the sorting of high wage workers to high wage industries increased substantially.

[Sorkin \(2018, Table V\)](#) reports that nearly half of the variation in his flows based measure of firm desirability is between 4 digit industries. He argues that elevated wages in sectors such as mining primarily reflect compensating differentials. Relating the firm wage effects to measures of firm desirability, he concludes that as much as two thirds of the variation in firm wage fixed effects could reflect compensating differentials. Subsequent work by [Lamadon et al. \(2022\)](#) concurs that compensating differentials are an important determinant of firm wage fixed effects; however, they also find that high wage firms tend to have the best amenities. This view is corroborated by [Sockin \(2022\)](#), who documents that higher wage firms list more job amenities in job advertisements. Consistent with this view, [Maestas et al. \(2023\)](#) find that adjusting for valuations of observed amenities derived from stated preference experiments actually widens inter-industry wage differentials.

An emerging consensus is that the most desirable firms tend to offer both the highest wages and the best amenities, making firms with large wage fixed effects highly desirable on average. [Roussille and Scuderi \(2023\)](#) provide revealed preference evidence from an online job board for software engineers that higher wage firms offer better observed and unobserved amenities. Similar conclusions are reached by [Lehmann \(2023\)](#) and [Lagos \(2019\)](#) utilizing administrative records from Austria and Brazil, respectively. [Caldwell et al. \(2024b\)](#) provide survey evidence from German workers that perceptions of the wages available at other firms are strongly correlated both with firm effect estimates from administrative data and with workers' perceptions of the non-wage amenities at those firms. See [Mas \(2024\)](#) for a comprehensive analysis of the recent literature on compensating differentials.

3 The AKM model

The fixed effects model considered by [Abowd et al. \(1999\)](#) can be written:

$$Y_{it} = \alpha_i + \psi_{j(i,t)} + X'\beta + \varepsilon_{it}, \quad (1)$$

where Y_{it} is the logarithm of worker i 's wages in year t and $j(i, t) \in \{1, \dots, J\} \equiv [J]$ is a function returning the identity of the firm employing worker i in year t . In their original application to an unbalanced panel of French administrative data, J was on the order of five hundred thousand, two million workers were studied, and the panel consisted of roughly five million person-year observations. Subsequent work has considered much larger samples. For instance, [Song et al. \(2019\)](#) fit models with over 79 million person effects and 5.8 million firm effects to a five year panel with 220 million person-year observations. To avoid notational clutter, it will be useful to restrict attention to the case where the panel is balanced in what follows such that $t \in \{1, \dots, T\} \equiv [T]$.

The person effect α_i is a portable component of wages that a worker can take with them to other employers. This parameter can capture skills, as well as

a worker's reputation, bargaining prowess, or discrimination at the market level. The firm effect ψ_j is a non-portable component of wages enjoyed only when a worker is employed at firm j . This effect can be a function of both the firm's productivity, some of which is shared with the worker in the form of higher wages, and its unobserved amenities, which may yield compensating differentials. The firm effect may also reflect the degree to which effort is monitorable at the firm, which can generate variation in efficiency wages (Shapiro and Stiglitz, 1984; Akerlof and Yellen, 1990). The vector X_{it} includes year fixed effects and measures of labor market experience.³

The time varying error ε_{it} captures innovations to the portable component of the worker's wage along with any measurement errors. These errors are assumed to obey a strict exogeneity restriction, requiring that $\mathbb{E}[\varepsilon_{it} | \mathbf{j}(i, s) = j, X_{is} = x] = 0$ for all workers $i \in \{1, \dots, N\} \equiv [N]$, all time periods $(s, t) \in [T]^2$, and all possible firm assignments $j \in [J]$ and covariate values $x \in \mathcal{X}$. From a statistical perspective, ε_{it} provides the "noise" that creates slippage between firm effect estimates and the true fixed effects. Thinking carefully about how to account for this noise is the core contribution of much of the recent econometrics literature studying these models.

The strict exogeneity condition embeds both the requirement that worker mobility between firms is not driven by time varying wage fluctuations (often described as "exogenous mobility") and that the mapping from worker and firm heterogeneity to expected log wages is additively separable. However, it does not restrict, in any way, the joint distribution of worker and firm effects. Therefore, workers may sort to firms based on any function of their own α_i and the vector ψ of firm wage effects. The pairing of (1) with the strict exogeneity restriction has come to be known as "the AKM model" and I will follow convention in using this eponym as a shorthand. It is worth noting, however, that closely related assumptions are now employed in several literatures exploiting the switching of units between groups (e.g., Finkelstein et al., 2016; Chetty and Hendren, 2018).

In the AKM model, movements between firms reveal differences in firm wage setting. In the case where $T=2$, for any two firms $j \neq k$ between which workers move, we have

$$\mathbb{E}[Y_{i2} - Y_{i1} | \mathbf{j}(i, 1) = j, \mathbf{j}(i, 2) = k, X_{i1}, X_{i2}] = \psi_k - \psi_j + (X_{i2} - X_{i1})'\beta. \quad (2)$$

As Abowd et al. (2002) detail, the firm effect levels are only identified up to a constant within the largest "connected set" of employers: that is, the set of firms connected, directly or indirectly, via worker moves. Intuitively, if there are two collections of firms between which workers never move, the difference in their wage levels will not be identified. A single restriction on the firm effects – typically a normalization that one of them is zero – within each

³ See Card et al. (2018) for discussion of identification issues posed by introducing age and year effects. In their original study, Abowd et al. (1999) included firm specific seniority trends, which introduces additional identification challenges that I will not consider here.

connected set is required for the design matrix of worker and firm dummies to have full rank, enabling least squares estimation of (1).

In the German social security records analyzed by Card et al. (2013), the largest connected set captured around 97–98 % of person year observations depending on the period analyzed. These shares can be lower when studying subpopulations. For example, fitting models separately by gender to Portuguese data, Card et al. (2016) find that the largest connected set comprises 88 % of person-year observations for male workers and 91 % of observations for female workers. In both settings, the wage distributions and worker characteristics in the largest connected set tend to be similar to those in the broader population.

Our discussion so far of the connectedness and normalization requirements for estimation of the firm effects has been a bit vague. The next subsection delves deeper into these subjects by providing a graph theoretic interpretation of the AKM model. I focus there on the properties of the *mobility network*, defined as a directed graph where vertices correspond to firms and edges represent worker moves between firms. At the cost of some additional notation, this network based lens will allow us to develop an interpretation of the AKM model as a restricted model of “edge effects.” This interpretation motivates a corresponding representation of the least squares estimator of firm effects as a linear combination of estimated edge effects. A closely related representation was explored by Jochmans and Weidner (2019). My exposition differs from theirs primarily in clarifying how the presence of cycles in the mobility network influence the algebraic mapping between the wage changes of movers and the firm fixed effects estimates. Section 3.2 investigates the extent to which the restrictions motivating the firm fixed effects estimator are satisfied in a benchmark dataset. Section 3.3 discusses causal interpretations of edge and firm effects, concluding with some directions for future research.

3.1 An edgy interpretation of firm effects

We begin with some definitions. A *graph* is a collection of vertices and edges joining those vertices. The graph we are considering is *directed*, which means that each edge starts at one vertex and ends at another. Here, the vertices correspond to the set of firms $[J]$. An *edge* is an ordered pair of vertices $(j, k) \in [J]^2$, with the first entry denoting an origin firm from which a worker moved and the second entry denoting the destination of the move. To simplify the analysis, we will continue to assume $T = 2$, in which case the set of all edges in the graph can be defined as:

$$E = \left\{ (j, k) \mid (j, k) \in [J]^2, j \neq k, \sum_{i \in [N]} 1 \{ \mathbf{j}(i, 1) = j, \mathbf{j}(i, 2) = k \} > 0 \right\}.$$

Denoting the total number of edges by $|E|$, I will index the edges by $\ell \in \{1, \dots, |E|\} \equiv [E]$, referring to individual edges by $\{e_\ell\}_{\ell \in [E]}$.

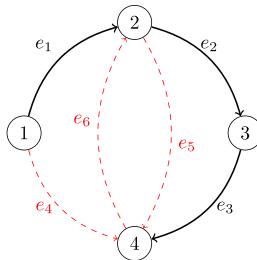


FIG. 1 A mobility network ($J = 4$, $|E| = 6$).

A *walk* is a sequence of edges that join a set of firms. A *trail* is a walk with no repeated edges. A *path* is a trail with no repeated firms. The mobility graph is *connected* if there is a path from any firm to any other firm. A *tree* is a connected graph for which there is a unique path between any pair of firms. A *spanning tree* is any subset of a connected graph that contains all firms and is a tree.

Fig. 1 depicts a connected graph with four firms and six edges. Arrows indicate the directions in which workers move between firms. A spanning tree of this network is given by the solid edges. The dashed edges depart from the tree by generating alternative paths of moving between firms. These alternate paths yield *cycles*: that is, trails that lead us back to where we started. For example, using a minus sign to denote traversal of an edge in reverse, the trail $\{e_1, e_2, e_3, -e_4\}$ is a cycle. A *fundamental cycle* is a cycle formed by departing from the spanning tree using a single edge not in the tree. There are $|E| - J + 1$ distinct fundamental cycles in a connected graph. The other fundamental cycles in this graph are $\{e_2, e_3, e_6\}$ and $\{e_2, e_3, -e_5\}$.

The *incidence matrix* \mathbf{B} provides a mathematical representation of the graph's edges.⁴ Every row of \mathbf{B} represents a firm, while every column represents an edge. A single entry in each column equals 1, denoting that edge's destination firm, and a single entry equals -1 , capturing that edge's origin firm. The remaining entries equal zero. In the graph above \mathbf{B} takes the form:

	e_1	e_2	e_3	e_4	e_5	e_6
1	-1	0	0	-1	0	0
2	1	-1	0	0	-1	1
3	0	1	-1	0	0	0
4	0	0	1	1	1	-1

⁴ Jochmans and Weidner (2019) work with a weighted definition of the incidence matrix. I rely here on an unweighted definition in order to highlight connections to cycles in the graph. Weights are introduced below in Section 3.1.2.

An important property of \mathbf{B} , to which we will return, is that its rows are orthogonal to the graph's cycles. For instance, the cycle $\{e_1, e_2, e_3, -e_4\}$ can be represented by the vector $c_1 = [1, 1, 1, -1, 0, 0]'$. Likewise, the cycles $\{e_2, e_3, e_6\}$ and $\{e_2, e_3, -e_5\}$ are captured by the vectors $c_2 = [0, 1, 1, 0, 0, 1]'$ and $c_3 = [0, 1, 1, 0, -1, 0]'$ respectively. It is easy to verify that $\mathbf{B}c_1 = \mathbf{B}c_2 = \mathbf{B}c_3 = 0$. More generally, $\mathbf{B}c = 0$ for any $E \times 1$ vector c in the linear span (also known as the "cycle space") of the fundamental cycles. For example, the trail $\{e_5, e_6\}$, which can be represented by $c_2 - c_3$, is in this graph's cycle space. In a connected graph, the cycle space is the *nullspace* of \mathbf{B} , meaning it contains the set of all vectors $c \in \mathbb{R}^J$ such that $\mathbf{B}c = 0$.

3.1.1 Firm effects as restricted edge effects

Returning to (2), we can now rewrite the AKM model in a notation directly linked to the structure of the graph. To simplify the analysis, suppose that the vector β is known and define $R = Y_2 - Y_1 - (X_2 - X_1)\beta$ as the $N \times 1$ vector of worker wage changes adjusted for the change in time varying covariates. Let \mathbf{F}_t denote the $N \times J$ matrix of firm assignment indicators in period t , the i 'th row of which can be written $(1\{\mathbf{j}(i, t) = 1\}, 1\{\mathbf{j}(i, t) = 2\}, \dots, 1\{\mathbf{j}(i, t) = J\})$. The AKM model implies

$$R = (\mathbf{F}_2 - \mathbf{F}_1)\psi + \varepsilon,$$

where $\psi = (\psi_1, \dots, \psi_J)'$ is the $J \times 1$ vector of firm fixed effects and $\varepsilon = (\varepsilon_{12} - \varepsilon_{11}, \dots, \varepsilon_{N2} - \varepsilon_{N1})'$ is the $N \times 1$ vector of differences in wage errors obeying $\mathbb{E}[\varepsilon|\mathbf{F}_1, \mathbf{F}_2] = 0$.

We can write the matrix of first differenced firm indicators in terms of the edge dummies via the relation $\mathbf{F}_2 - \mathbf{F}_1 = \mathbf{EB}'$ where \mathbf{E} is an $N \times |E|$ matrix of (directed) edge indicators – i.e., dummies of the form $1\{\mathbf{j}(i, 2) = k\} \cdot 1\{\mathbf{j}(i, 1) = j\}$ for all origin-destination firm pairs (j, k) traversed by movers. Hence, the AKM model is equivalently expressed in terms of the incidence matrix as

$$R = \mathbf{EB}'\psi + \varepsilon.$$

Here, strict exogeneity can be represented as the requirement that $\mathbb{E}[\varepsilon|\mathbf{E}] = 0$.

It is instructive to contrast the AKM model with a model of unrestricted edge fixed effects:

$$R = \mathbf{E}\Delta + u,$$

where Δ is an $|E| \times 1$ vector of edge effects and the $N \times 1$ error vector u obeys $\mathbb{E}[u|\mathbf{E}] = 0$. Section 3.3 introduces assumptions giving these edge effects a causal interpretation. The AKM model imposes $\Delta = \mathbf{B}'\psi$, which entails $|E| - J + 1$ linear restrictions on the edge effects. When these restrictions hold, the two error terms are identical ($u = \varepsilon$). Hence, the AKM model can be thought of as projecting the $|E| \leq J(J - 1)$ edge effects down to only $J - 1$ linearly independent firm effects.

To understand the nature of the edge restrictions entailed by the AKM model, note that for any vector c in the cycle space, $c'\Delta = c'B'\psi = 0$, which follows from the cyclic orthogonality properties of \mathbf{B} discussed earlier. For example, the AKM model imposes that wage changes should be symmetric across origin firm - destination firm pairs, a property that was emphasized by Card et al. (2013) and is reflected in our example of the cycle $c_2 - c_3$. However, the AKM restrictions go far beyond pairwise symmetry, restricting network dependent tuples of edge effects. For example, the fundamental cycle c_1 involves four edges. Though directly visualizing the restrictions pertaining to such 4-cycles is challenging, their logic mirrors the restrictions pertaining to the 2-cycles studied by Card et al. (2013): that “taking a walk” along the graph should have no effect on wages so long as one ends up back at the same firm where the walk began.

It may be helpful here to illustrate this reasoning with a simple thought experiment. Consider two workers of the same age, both of whom are employed at firm j in 2010, where they earn the same wage. In subsequent years, each worker switches employers twice before returning to firm j in 2020. The AKM model stipulates that we should expect these two workers to earn the same wages in 2020, regardless of the identity of their two intermediate employers. Indeed, if our two time periods were 2010 and 2020, these workers would be viewed as “stayers” and equation (2) predicts their wage change depends only on the change in time varying covariates.

Mathematically, these cyclic restrictions exhaust the empirical restrictions of the AKM model on edge effects within a connected set of firms. That is, if $c'\Delta = 0$ for any cycle in the graph space, then there must exist a set of firm effects capable of rationalizing the edges exactly. To understand why, recall that \mathbf{B} 's nullspace coincides with the cycle space of the graph, which implies we can decompose the edge effects as

$$\Delta = \mathbf{B}\dot{\psi} + \mathbf{C}\dot{\eta},$$

where $\dot{\psi}$ is a vector of coefficients from a linear projection of Δ onto \mathbf{B} , \mathbf{C} is an $|E| \times |E| - J + 1$ matrix collecting the graph's fundamental cycles, and $\dot{\eta}$ is an $|E| - J + 1$ vector of “cycle effects” that serve as residuals. Plugging this decomposition into the edge effects model yields,

$$R = \mathbf{EB}'\dot{\psi} + \mathbf{C}\dot{\eta} + u.$$

The AKM model amounts to assuming that $\dot{\eta} = 0$, in which case $\dot{\psi} = \psi$ and $\varepsilon = u$. When there are no cycle effects, then the true dimension of the edge effects is much lower than it appears: the AKM model reduces the $|E|$ edges to $J - 1$ linearly independent firm effects.

While $|E| - J + 1 = 3$ in the graph depicted in Fig. 1, large scale empirical applications can feature hundreds of thousands (or even millions) of restrictions. As with any economic or statistical model, these restrictions are unlikely to be

satisfied exactly. When the restrictions do not hold, the firm effects can be thought of as a linear projection that provides a lower dimensional summary of the edge effects. We will examine the quality of this summary in [Section 3.2](#).

3.1.2 Estimators

Let $\hat{\Delta} = (\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbf{R}$ denote the $|E| \times 1$ vector of estimated edge effects. The normal equations defining the least squares estimator of ψ can be written

$$\mathbf{B}\mathbf{W}\hat{\Delta} = \mathbf{L}\psi,$$

where $\mathbf{W} = \mathbf{E}'\mathbf{E}$ is a diagonal weighting matrix recording the number of workers moving along each edge and $\mathbf{L} = \mathbf{B}\mathbf{W}\mathbf{B}' = (\mathbf{F}_2 - \mathbf{F}_1)'(\mathbf{F}_2 - \mathbf{F}_1)$ is a symmetric $J \times J$ matrix known in graph theory as the *Laplacian*. The Laplacian encodes information about each firm's role in the mobility network. The j th row and k th column of \mathbf{L} equals the negative of the total number of workers moving (in either direction) between firms j and k when $j \neq k$, while the j th diagonal entry of \mathbf{L} gives the total number of workers moving to or from firm j .

\mathbf{L} is singular, which implies there are an infinite number of solutions to the normal equations. [Jochmans and Weidner \(2019\)](#) study the properties of the solution $\mathbf{L}^\dagger \mathbf{B}\mathbf{W}\hat{\Delta}$, where \mathbf{L}^\dagger denotes the Moore-Penrose inverse of \mathbf{L} . I will take a slightly different approach by studying the solution that results when one of the firms is taken as the “reference firm” with zero firm effect. While both solutions yield the same predicted edge effects, the reference firm solution is typically used in practice and happens to also simplify the subsequent theoretical analysis. [Bozzo \(2013\)](#) provides some useful results on connections between the two approaches.

Define $\mathbf{B}_{(1)}$ as the submatrix leaving out the first row of \mathbf{B} and let $\mathbf{L}_{(11)} = \mathbf{B}_{(1)}\mathbf{W}\mathbf{B}'$ denote the submatrix of \mathbf{L} leaving out its first row and column. If we impose the restriction $\psi_1 = 0$, then we obtain the constrained normal equations

$$\mathbf{B}_{(1)}\mathbf{W}\hat{\Delta} = \mathbf{L}_{(11)}\psi_{(1)},$$

where $\psi_{(1)}$ is ψ omitting its first entry. A classic result in graph theory, *Kirchhoff's matrix tree theorem*, states that any cofactor of the unweighted Laplacian matrix gives the number of spanning trees in the graph. When the edges are weighted, generalizations of the theorem (e.g., [Spielman, 2019](#), Theorem 13.4.1) establish that any cofactor of \mathbf{L} gives the total edge weight of the graph's spanning trees, where the weight of each tree is given by the product of the edge weights it contains. A connected graph must have at least one spanning tree. Hence, when the mobility graph is connected, it follows that $\det(\mathbf{L}_{(11)}) > 0$, implying that $\mathbf{L}_{(11)}$ has full rank.

The least squares estimator that results from treating the first firm as the reference can therefore be written

$$\hat{\psi}_{(1)} = \mathbf{L}_{(11)}^{-1}\mathbf{B}_{(1)}\mathbf{W}\hat{\Delta}. \quad (3)$$

Variants of this estimator are heavily used in applied research; however, computation is typically implemented by iterative conjugate gradient (CG) methods rather than direct inversion of $\mathbf{L}_{(11)}$.⁵ CG routines are available in most scientific computing packages including MATLAB and SciPy. The efficiency of these routines is greatly aided by “preconditioning” the problem with an approximate Cholesky factorization of $\mathbf{L}_{(11)}$. In the empirical examples below, I rely on the combinatorial multigrid solver package of Koutis et al. (2011) as a preconditioner.

3.1.3 Combination weights and smoothing

Equation (3) reveals that the estimated firm effects are linear combinations of the average wage changes associated with each edge. In general, the combination weights are such that each firm effect can depend on each element of $\hat{\Delta}$. For example, when the edges in the graph depicted in Fig. 1 each represent a single mover, the firm effect estimates can be written:

$$\hat{\psi}_{(1)} = (\mathbf{B}_{(1)} \mathbf{B}')^{-1} \mathbf{B}_{(1)} \hat{\Delta} = \begin{pmatrix} \frac{7}{12} & -\frac{1}{12} & -\frac{1}{12} & \frac{5}{12} & -\frac{1}{6} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{5}{12} & \frac{1}{12} & \frac{1}{12} & \frac{7}{12} & \frac{1}{6} & -\frac{1}{6} \end{pmatrix} \hat{\Delta}. \quad (4)$$

Recall that $\mathbf{B}_{(1)} c = \mathbf{0}$ for any vector c in the graph’s cycle space. It is easy to verify in the above example that perturbing $\hat{\Delta}$ by adding to it any vector $c \in \{c_1, c_2, c_3\}$ yields no change in the estimated firm effects $\hat{\psi}_{(1)}$. This cyclic invariance property can be thought as offering a form of robustness to certain types of confounding trends in the error ε . For example, a trend shared by the movers traversing edges e_5 and e_6 (i.e., movers between Firms 2 and 4) will “difference out.” Likewise, $\hat{\psi}_{(1)}$ is unaffected by adding a constant to the wage changes of the movers traversing each of the edges e_2 , e_3 , and e_5 .⁶ Adding a constant to the wage changes of movers on edges e_2 and e_3 while subtracting that constant from movers on e_5 also has no effect.

Whether confounding cyclic trends of this nature tend to be present in economic data is an interesting question for future research. Mobility cycles are common among employers in the same industry and region. Suppose the error ε

⁵We have glossed over the issue of how to form R – and consequently $\hat{\Delta}$ – in the first place. Typically, one estimates the coefficient vector β on the time varying covariates X_h in a first step and subtracts them off. This initial adjustment step is also greatly accelerated with CG methods.

⁶When the number of movers differs across edges in a cycle, then the magnitude of a trend shared across the cycle’s edges would need to be inversely proportional to the number of movers along each edge in order to difference out. That is, the firm effect estimates become invariant to perturbing $\hat{\Delta}$ in the direction $\mathbf{W}^{-1}c$ where c is a vector in the cycle space. However, fitting the AKM model directly to the edge effects by unweighted least squares (i.e., setting $\mathbf{W} = \mathbf{I}$ in estimation) restores invariance to cycle specific trends.

takes the form $\varepsilon = \mathbf{C}\eta + u$ where η is a vector of cycle effects driven by demand shocks to those industry-regions. Though this error structure violates the AKM edge restrictions, unweighted firm effect estimates remain unbiased because $(\mathbf{B}_{(1)}\mathbf{B}')^{-1}\mathbf{B}_{(1)}\hat{\Delta} = \psi + (\mathbf{B}_{(1)}\mathbf{B}')^{-1}\mathbf{B}_{(1)}u$.

An important simplification of (3) arises in the case where the mobility graph is a tree. By definition, a tree has J vertices and $J - 1$ edges, which implies the submatrix $\mathbf{B}_{(1)}$ is square. Recall that $\mathbf{L}_{(11)} = \mathbf{B}_{(1)}\mathbf{W}\mathbf{B}'$. Since $\mathbf{L}_{(11)}$ has full rank, $\mathbf{B}_{(1)}$ must also have full rank. Hence, we can write:

$$\hat{\psi}_{(1)} = (\mathbf{B}_{(1)}\mathbf{W}\mathbf{B}')^{-1}\mathbf{B}_{(1)}\mathbf{W}\hat{\Delta} = (\mathbf{B}'_{(1)})^{-1}\mathbf{W}^{-1}\mathbf{B}_{(1)}^{-1}\mathbf{B}_{(1)}\mathbf{W}\hat{\Delta} = (\mathbf{B}'_{(1)})^{-1}\hat{\Delta}.$$

The predicted edge effects implied by the estimated firm effects are given by $\mathbf{B}'_{(1)}\hat{\psi}_{(1)} = \hat{\Delta}$, indicating that the firm effects rationalize the adjusted wage changes $\hat{\Delta}$ with no error. This phenomenon reflects that the firm effects are just-identified by (i.e., “they saturate”) the edge specific wage changes.

Consider the spanning tree depicted in Fig. 1, which is comprised of the graph’s first three edges. The $\mathbf{B}'_{(1)}$ associated with this tree and its inverse are depicted below:

$$\mathbf{B}'_{(1)} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}, \quad (\mathbf{B}'_{(1)})^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

In any spanning tree, one can always ensure that $(\mathbf{B}'_{(1)})^{-1}$ is triangular by ordering the edges of $\mathbf{B}_{(1)}$ according to their distance from Firm 1. However, some of the entries in such a triangle may possess a negative sign if reaching the reference firm requires traversing an edge in reverse.⁷ The triangular structure of $(\mathbf{B}'_{(1)})^{-1}$ ensures that each firm effect is simply the sum of the (oriented) edge effects on the path connecting it to the reference firm. Consequently, the difference in firm effect estimates for any two firms j and k connected by an edge must equal the average wage change of the workers moving directly between them. We will return to this property when discussing causal interpretations of firm effects.

A closely related property can be shown to hold when the graph is a *polytree*, meaning that the undirected graph is a tree but some firm pairs may be connected by edges in both directions. For example, adding an edge from Firm 4 to Firm 3 to the spanning tree depicted in Fig. 1 yields a polytree. Any polytree can be transformed into a simple tree by transferring the weight from one edge to the other in each pair of edges connecting the same firms. This

⁷ For example, choosing Firm 3 as the reference in this spanning tree yields $(\mathbf{B}'_{(3)})^{-1} = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

transformation can be represented by an $|E| \times J - 1$ matrix \mathbf{T} that differences the relevant edge pairs in the incidence matrix. For example,

$$\underbrace{\begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}}_B \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix}}_T = \begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Remarkably, the firm effect estimates are invariant to such transformations. Specifically, when \mathbf{B} represents a polytree, the least squares weights for the transformed graph $(\mathbf{B}_{(1)} \mathbf{W} \mathbf{T} \mathbf{T}' \mathbf{W} \mathbf{B}_{(1)}')^{-1} \mathbf{B}_{(1)} \mathbf{W} \mathbf{T} \mathbf{T}' \mathbf{W}$ equal the weights $\mathbf{L}_{(11)}^{-1} \mathbf{B}_{(1)} \mathbf{W}$ for the untransformed graph.⁸ Since $\mathbf{B}_{(1)} \mathbf{W} \mathbf{T}$ is a square invertible matrix representing a simple tree, the firm effects derived from fitting the AKM model to a polytree can equivalently be written $(\mathbf{T}' \mathbf{W} \mathbf{B}_{(1)}')^{-1} \mathbf{T}' \mathbf{W} \hat{\Delta}$. Consequently, in any polytree, the difference in estimated firm effects between any pair of firms joined by a pair of edges will equal a mover weighted average of the two oriented edge effects connecting them. As in a simple tree, the difference in firm effects for any pair of firms joined by a single edge will depend only on that estimated edge effect.

When the graph is not a tree, the firm effects become over-identified and the vector of predicted wage changes is $\hat{\Delta} \equiv \mathbf{B}_{(1)}' \hat{\psi}_{(1)} = \mathbf{H} \hat{\Delta}$, where $\mathbf{H} = \mathbf{B}_{(1)}' \mathbf{L}_{(11)}^{-1} \mathbf{B}_{(1)} \mathbf{W}$ is an $|E| \times |E|$ weighted projection matrix that is invariant to the choice of reference firm. Like the usual “hat” matrix (Hoaglin and Welsch, 1978), \mathbf{H} ’s diagonal entries $\{h_{\ell\ell}\}_{\ell \in [E]}$ give the leverage of each observation (in this case each edge effect) on the predicted value. One can write the ℓ ’th leverage:

$$h_{\ell\ell} = b_{\ell} \mathbf{L}_{(11)}^{-1} b_{\ell} n_{\ell},$$

where b_{ℓ} is the ℓ ’th column of $\mathbf{B}_{(1)}$ and n_{ℓ} is the ℓ ’th diagonal entry of \mathbf{W} . In large systems, costly inversion of $\mathbf{L}_{(11)}$ can be avoided when by breaking computation into a CG step that solves a linear system and a subsequent matrix multiplication step.⁹ Leverages lie in the interval $[0, 1]$, with larger values indicating that dropping that edge from the data would lead to a greater change in the estimated firm effects. Any edge that is part of a cycle has $h_{\ell\ell} < 1$. An

⁸ The transformation \mathbf{T} maps the edges back into the span of the weighted incidence matrix, implying the weighted orthogonality condition $\mathbf{T}' \mathbf{W} (\mathbf{I} - \mathbf{B}_{(1)}' \mathbf{L}_{(11)}^{-1} \mathbf{B}_{(1)} \mathbf{W}) = 0$. Expanding this condition yields $\mathbf{T}' \mathbf{W} = \mathbf{T}' \mathbf{W} \mathbf{B}_{(1)}' \mathbf{L}_{(11)}^{-1} \mathbf{B}_{(1)} \mathbf{W}$. Premultiplying by $\mathbf{B}_{(1)} \mathbf{W} \mathbf{T}$ gives $\mathbf{B}_{(1)} \mathbf{W} \mathbf{T} \mathbf{T}' \mathbf{W} = \mathbf{B}_{(1)} \mathbf{W} \mathbf{T} \mathbf{T}' \mathbf{W} \mathbf{B}_{(1)}' \mathbf{L}_{(11)}^{-1} \mathbf{B}_{(1)} \mathbf{W}$. Dividing both sides by $\mathbf{B}_{(1)} \mathbf{W} \mathbf{T} \mathbf{T}' \mathbf{W} \mathbf{B}_{(1)}'$ yields the result.

⁹ Note that we can rewrite the ℓ ’th leverage $h_{\ell\ell} = b_{\ell} z_{\ell}$, where $z_{\ell} = \mathbf{L}_{(11)}^{-1} b_{\ell} n_{\ell}$. The first step solves the equation $\mathbf{L}_{(11)} z_{\ell} = b_{\ell} n_{\ell}$ for the vector z_{ℓ} via CG methods. The second step computes $h_{\ell\ell} = b_{\ell} z_{\ell}$ by vector multiplication. This process can be parallelized across edges to recover all of the leverages.

edge with $h_{\ell\ell} = 1$ is known as a *bridge*. Dropping a bridge breaks the graph into two or more connected components, in which case at least one firm effect can no longer be estimated.

When the graph is a tree, all edges are bridges and \mathbf{H} is the identity matrix. However, when the graph exhibits cycles, \mathbf{H} departs from identity and some “smoothing” across edges takes place. The rows of \mathbf{H} give the smoothing weights used to form the prediction for each edge. Each row’s weights sum to one but the entries can be negative. For example, if we assume a single mover traverses each edge of the graph depicted in Fig. 1 then the hat matrix takes the following form:

$$\mathbf{H} = \begin{pmatrix} \frac{7}{12} & -\frac{1}{12} & -\frac{1}{12} & \frac{5}{12} & -\frac{1}{6} & \frac{1}{6} \\ -\frac{1}{12} & \frac{7}{12} & -\frac{5}{12} & \frac{1}{12} & \frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{12} & -\frac{5}{12} & \frac{7}{12} & \frac{1}{12} & \frac{1}{6} & -\frac{1}{6} \\ \frac{5}{12} & \frac{1}{12} & \frac{1}{12} & \frac{7}{12} & \frac{1}{6} & -\frac{1}{6} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

Inheriting the properties of $\mathbf{B}_{(1)}$, these smoothing weights are orthogonal to any vector in the cycle space but are otherwise widely dispersed across the edges. Placing weight on edges throughout the network is efficient when the AKM model restrictions hold. Otherwise, $\tilde{\Delta}$ may provide a poor estimate of Δ .

3.2 Evaluating the AKM restrictions

To evaluate whether the AKM model provides an accurate summary of the wage changes associated with (directed) moves between firm pairs, we study two years of the Veneto Workers History (VHW) data. This dataset has emerged as a popular benchmark in the literature due to the low barriers associated with obtaining access to it.¹⁰ We work with an extract of 1,859,459 person-year observations from the years 1999 and 2001 that was studied previously by Kline et al. (2020). The largest connected set contains 73, 933 firms and 747, 205 workers, 197, 572 of whom switch employers between the two years. These 197, 572 “movers” are spread across 150, 417 edges. Hence, the AKM model implies 76, 485 restrictions on the edge effects.

The AKM model is fit to the log daily wage changes of workers by solving the normal equations using MATLAB’s preconditioned conjugate gradient routine. The only time varying covariate included is an indicator for the year

¹⁰The data can be requested at <https://www.frdb.org/en/dati/dati-inps-carriere-lavorative-in-veneto/>.

being 2001. Job stayers contribute to the firm effect estimates only indirectly via estimation of the year fixed effect $\hat{\beta}$. We use this same year effect estimate to preadjust wage changes before collapsing them to estimated edge effects $\hat{\Delta}$.

3.2.1 Visualizing goodness of fit

[Fig. 2](#) summarizes how the conditional distribution of estimated edge effects varies with the AKM predictions. Each dot depicts the mean edge effect within a bin of predicted edge effects ($\tilde{\Delta}$). The bands around the dots give a sense of dispersion within each bin: the upper limit of each band gives the 75th percentile of estimated edge effects in that bin, while the lower limit gives the 25th percentile.

The AKM model stipulates that, in the absence of noise, the dots should all lie on the dashed 45 degree line. On average, the edge effects do tend to lie remarkably close to the 45 degree line. Moreover, the bands around the dots reveal only modest dispersion around the averages. However, the AKM model was fit to the same data as the edge effects, which induces a mechanical dependence between the two sets of estimates. Indeed, if the graph had been a tree, the edge predictions would all lie exactly on the 45 degree line.

Looping over edges to compute the leverages $\{h_{ee}\}_{e \in [E]}$ reveals that about 22 % of the edges are bridges that must mechanically lie on the 45 degree line. Roughly 44 % of the firm effects are just-identified by one of these bridges. Dropping the bridges leaves 117,657 edges with $h_{ee} < 1$ that connect 41,195 firms. The x's in [Fig. 2](#) depict the mean predictions in this subpopulation, which still track the 45 degree line closely. However, the interquartile range of deviations is amplified. To evaluate whether these deviations are larger than we should expect under the AKM model requires accounting for noise in the estimated edge effects.

3.2.2 Accounting for noise

The noise in the edge effects that concerns us derives from the vector u of wage change errors. One can think of these errors as capturing the idea that if a different worker happened to traverse the same edge, a different wage change would likely result. In what follows, I will use the expectation and variance operators $\mathbb{E}_u[\cdot]$ and $\mathbb{V}_u[\cdot]$ to convey that integration is ultimately being conducted with respect to the edge effects error u introduced in [Section 3.1.1](#). Hence, the expected value of the AKM prediction is $\mathbb{E}_u[\tilde{\Delta}] = \mathbf{H}\Delta$ and the variance matrix of the estimated edge effects is $\mathbb{V}_u[\hat{\Delta}] = (\mathbf{E}'\mathbf{E})^{-1}\mathbf{E}'\mathbb{E}[uu']\mathbf{E}(\mathbf{E}'\mathbf{E})^{-1}$.

Denote the vector of differences between the predicted and estimated edge fixed effects by $\hat{\Delta} - \tilde{\Delta} = \mathbf{M}\hat{\Delta}$, where $\mathbf{M} = (\mathbf{I} - \mathbf{H})$ is the “residual maker” matrix. Let $\hat{\Delta}_\ell$ denote the ℓ 'th entry of $\hat{\Delta}$ and $\tilde{\Delta}_\ell$ the ℓ 'th entry of $\tilde{\Delta}$. A standard goodness of fit statistic is the sum of squared residuals. We will work with a mover weighted version of this statistic: Σ_ℓ

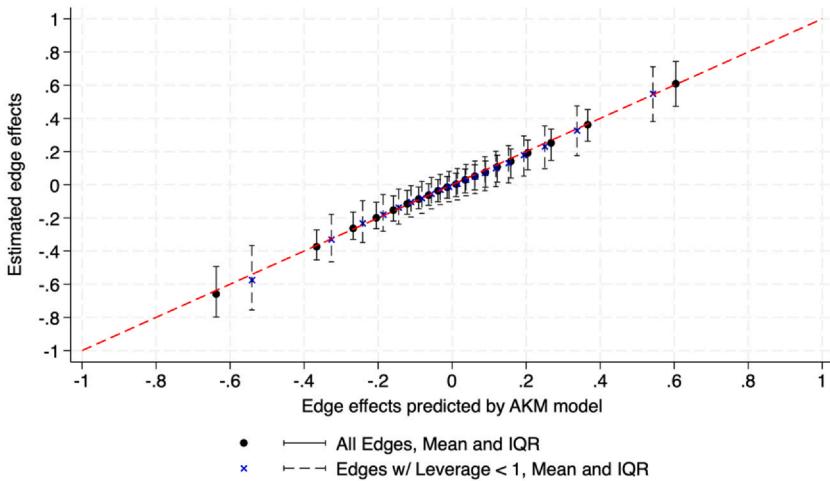


FIG. 2 Log daily wage change of edge ($\hat{\Delta}$) versus AKM prediction ($\tilde{\Delta}$). Notes: The vertical axis depicts binned averages of the elements of $\hat{\Delta}$: the average adjusted log daily wage changes associated with each origin-destination firm edge. The horizontal axis gives bins of $\tilde{\Delta}$: the wage change predicted by the least squares estimates of firm effects. Panel comprised of the 1999 and 2001 waves of the Veneto Work Histories dataset developed by the Economics Department in Universita Ca' Foscari Venezia under the supervision of Giuseppe Tattara.

$n_e(\hat{\Delta}_e - \tilde{\Delta}_e)^2 = (\hat{\Delta} - \tilde{\Delta})'W(\hat{\Delta} - \tilde{\Delta}) = \hat{\Delta}'M'WM\hat{\Delta}$.¹¹ So long as the wage change errors have finite variance, we can write the expectation of this sum as

$$\mathbb{E}_u[\hat{\Delta}'M'WM\hat{\Delta}] = \underbrace{\Delta'M'WM\Delta}_{\text{squared bias}} + \underbrace{\text{trace}(M'WM\mathbb{V}_u[\hat{\Delta}])}_{\text{noise}}.$$

The AKM model stipulates that $M\Delta = 0$, which implies $\Delta'M'WM\Delta = \sum_e n_e(\Delta_e - \mathbb{E}_u[\tilde{\Delta}_e])^2 = 0$. However, the model doesn't restrict the trace term, which captures the expected contribution of noise. If the wage change errors are independent across movers, then $\mathbb{V}_u[\hat{\Delta}]$ is a diagonal matrix and the trace expression simplifies to

$$\text{trace}(M'WM\mathbb{V}_u[\hat{\Delta}]) = \sum_{\ell \in [E]} n_\ell(1 - h_{\ell\ell})\mathbb{V}_u[\hat{\Delta}_\ell],$$

where $\mathbb{V}_u[\hat{\Delta}_\ell]$ is the ℓ th diagonal entry of $\mathbb{V}_u[\hat{\Delta}]$. This formula captures the intuition that high leverage edges are expected to yield smaller residuals

¹¹ The residual maker matrix will not, in general, be symmetric when the number of movers varies across edges. Fortunately, $M'WM = WM$ for any distribution of mover weights, which significantly simplifies the calculations below.

because of overfitting. Conversely, edges with higher noise levels $\mathbb{V}_u[\hat{\Delta}_\ell]$ should yield larger squared residuals.

For edges with more than a single mover, a simple unbiased estimator of $\mathbb{V}_u[\hat{\Delta}_\ell]$ is available: the squared standard error, $\widehat{\mathbb{V}}_u[\hat{\Delta}_\ell] = \frac{1}{n_\ell} s_\ell^2$, where s_ℓ is the standard deviation of adjusted wage changes along edge ℓ . Reflecting the sparsity of the mobility network, only 9459 of the edges that are not bridges have 2 or more movers. Denote this set of edges by \mathcal{E}_{2+} . Combining the leverages with the edge specific standard errors yields an expected sum of squared residuals under the null hypothesis that the AKM model holds of $\sum_{\ell \in \mathcal{E}_{2+}} n_\ell (1 - h_{\ell\ell}) \widehat{\mathbb{V}}_u[\hat{\Delta}_\ell] = 207.58$.

Empirically, the residual sum of squares is $\sum_{\ell \in \mathcal{E}_{2+}} n_\ell (\hat{\Delta}_\ell - \tilde{\Delta}_\ell)^2 = 360.51$. The difference, $360.51 - 207.58 = 152.92$, between the actual and expected sum of squares provides an unbiased estimate of the sum of squared approximation errors: $\sum_{\ell \in \mathcal{E}_{2+}} n_\ell (\Delta_\ell - \mathbb{E}_u[\tilde{\Delta}_\ell])^2$. A natural benchmark for these approximation errors is the (mover-weighted) sum of squared edge effects $\sum_{\ell \in \mathcal{E}_{2+}} n_\ell \Delta_\ell^2$, an unbiased estimate of which is $\sum_{\ell \in \mathcal{E}_{2+}} n_\ell (\hat{\Delta}_\ell^2 - \widehat{\mathbb{V}}_u[\hat{\Delta}_\ell]) = 978.67$. The ratio of these two numbers can be thought of as one minus the (uncentered) R^2 from an infeasible mover-weighted regression of the true edge effects in $\{\Delta_\ell\}_{\ell \in \mathcal{E}_{2+}}$ onto the matrix of first differenced firm dummies.¹² Hence, the data suggest the AKM approximation captures roughly $(1 - 152.92/978.67) \times 100 \approx 84\%$ of the variation in true edge effects. Equivalently, the estimated correlation between the edge effects and the AKM predictions is 0.92.

Of course, this R^2 estimate is itself subject to sampling uncertainty and applies only to a particular population of edges. Table 1 repeats this goodness of fit exercise restricting to edges with more movers. The R^2 estimates are remarkably stable, suggesting that these findings are unlikely to be an artifact of noise. The final column of the table reports the square root of the noise level within edges due to irreducible uncertainty across movers. Depending on the sample of edges considered, the average noise level is four to five times greater than the average squared model error.

The two rows in the second panel of Table 1 impute the noise levels of the edges with a single mover and recompute the relevant quadratic forms over all edges that are not bridges. The first of these rows sets the noise level of the singleton edges equal to twice the average noise level of edges with exactly 2 movers, an imputation that would be valid under homoscedasticity. The second row relaxes the homoscedasticity assumption by allowing an arbitrary linear relationship between the log of the average noise level and the log of the number of movers. This linear relationship, estimates of which are depicted in Appendix Fig. A.1, fits the data well and suggests slightly higher noise levels

¹² The centered R^2 is nearly identical because the mean edge effect in the \mathcal{E}_{2+} sample is .01 and the mean AKM prediction in this sample is .01. In the broader sample of 117,657 edges that are not bridges, the mean edge effect is 0.001 and the mean AKM prediction is 0.005.

TABLE 1 Goodness of fit by edge sample.

Sample	Number of movers	Number of edges	Root mean squared model error	Root mean squared edge effect	R^2	Root mean noise level
At least 2 movers	50,254	9459	0.055	0.140	84.37	0.124
At least 3 movers	39,048	3856	0.041	0.117	87.58	0.092
At least 4 movers	34,941	2487	0.038	0.109	87.95	0.076
At least 1 mover						
Singleton noise level twice edges w/ 2 movers	158,452	117,657	0.119	0.225	71.85	0.195
Singleton noise imputed via log-log regression	158,452	117,657	0.112	0.219	73.86	0.204
Including bridges						
Singleton noise imputed via log-log regression	197,572	150,417	0.100	0.232	81.38	0.200

Notes: All samples but the last are comprised of edges that are not bridges. Letting \mathcal{E} denote the set of edges under consideration and $|\mathcal{E}| = \sum_{e \in \mathcal{E}} n_e$ the number of movers across such edges, the “root mean squared model error” is computed as the square root of $\frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} n_e ((\hat{\Delta}_e - \bar{\Delta}_e)^2 - (1 - h_{ee}) \hat{V}_e [\hat{\Delta}_e])$. “Root mean squared edge effects” is given by the square root of $\frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} n_e (\hat{\Delta}_e^2 - \hat{V}_e [\hat{\Delta}_e])$. The R^2 is the square of the ratio of these two quantities. Root mean noise level gives the square root of $\frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} n_e \hat{V}_e [\hat{\Delta}_e]$. The row labeled “Singleton noise level twice edges w/ 2 movers” imputes $\hat{V}_e [\hat{\Delta}_e]$ for each edge with a single mover as twice the average squared standard error of edges with exactly two movers. The rows labeled “Singleton noise imputed via log-log regression” impute the noise level of edges with a single mover based upon a linear regression among edges with 2-10 movers of the log of the average noise level against an intercept and the log of the number of movers. Imputations conducted separately for bridges and non-bridges.

for the singleton edges. Under both imputations, the R^2 falls modestly to just above 70 %.

Finally, recall that nearly half of the firm effects are just-identified by a bridge, contributing no model error at all to the edge predictions. Applying a corresponding linear imputation of singleton noise levels to the bridges (depicted in Appendix Fig. A.1) yields an estimated sum of squared edge effects across all edges $\sum_{\ell=1}^{150,417} n_\ell (\hat{\Delta}_\ell - \hat{V}_u[\hat{\Delta}_\ell])$ of roughly 10, 671. Hence, the estimated R^2 from an infeasible regression of all edge effects (inclusive of bridges) onto the first differenced firm dummies evaluates to $[1 - (0.112)^2 \times 158, 452/10, 671] \times 100 \approx 81\%$.

In sum, the AKM model provides a highly informative (albeit imperfect) summary of the expected wage effects of worker mobility. If we were using firm effect estimates to predict the wage changes associated with worker moves, these findings suggest that noise would be a greater hindrance than model error. The model errors that are present result from cycles in the mobility network among a highly concentrated subset of firms. One interpretation of these errors is that they reflect heterogeneity in the firm effects faced by different sorts of workers. We now turn to thinking about the conditions under which the estimated AKM firm effects retain a causal interpretation in the presence of such heterogeneous effects.

3.3 Causality

The AKM model bears a strong resemblance to a difference in differences specification with J treatment arms where firm effect differences $\psi_j - \psi_k$ represent average treatment effects and the exogenous mobility assumption ensures “parallel trends.” It is natural then to ask whether least squares estimation of (1) can identify causal effects under non-parametric restrictions on potential outcomes cording to their share of all person-year observations including th of finding conditions under which the edge effects introduced in Section 3.1.1 can be given a causal interpretation. To ease exposition, we will again confine attention to the case where $T = 2$ and ignore time varying covariates, which can be thought of as having been adjusted for in a previous step.

Let $Y_{it}(d_1, d_2)$ denote the potential log wage of worker i in year t who works at firm $d_1 \in [J]$ in period 1 and $d_2 \in [J]$ in period 2. To mimic conventional treatment effects notation, I will use the symbol $D_{it} = \mathbf{j}(i, t)$ to denote the firm employing worker i in period t . We now state three assumptions that endow the average wage changes of workers switching employers between the two periods with a causal interpretation. Our first assumption is an exclusion restriction:

Assumption 1. (Exclusion). $Y_{it}(d_1, d_2) = Y_{it}(d_t)$ for $t \in \{1, 2\}$.

This assumption rules out the possibility that past or future firm assignments affect wages. Assumption 1 is violated in sequential auction models (Postel-Vinay and Robin, 2002b; Cahuc et al., 2006), which posit that hiring

wages are influenced by the firm from which a worker was poached. However, [Di Addario et al. \(2023\)](#) find in Italian data that past employers exhibit a negligible influence on hiring wages outside of the law and banking sectors, suggesting this assumption is likely to provide a reasonable approximation for most workers. When [Assumption 1](#) holds, we can link observed wages to potential wages via the relation $Y_{it} = Y_{it}(D_{it})$.

The next assumption mimics the parallel trends assumption of standard difference in differences models:

Assumption 2 (Parallel trends). $\mathbb{E}[Y_{i2}(j) - Y_{i1}(j)|D_{i1} = j, D_{i2} = k] = 0$
 $\forall k \neq j \in [J]^2$.

[Assumption 2](#) states that, among workers switching between any pair of firms, the average potential wages at their origin firms would not have changed between periods. As noted earlier, we should think of Y_{it} here as pre-adjusted for year and age/experience effects, in which case this amounts to a restriction that potential origin and destination wages exhibit a common time trend. [Card et al. \(2013\)](#) reported event study plots of the average earnings trajectories of workers who transitioned between groups of firms characterized by their leave-out wage quartile. These plots, which are now a standard diagnostic, indicate that workers moving to high wage firms do not experience faster wage growth before moving, nor does their wage trend change upon moving to a new firm, suggesting that [Assumption 2](#) provides a reasonable approximation.

Finally, we make a stationarity assumption on average treatment effects among firm switchers:

Assumption 3 (Stationarity). $\mathbb{E}[Y_{i1}(k) - Y_{i1}(j)|D_{i1} = j, D_{i2} = k] = \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)|D_{i1} = j, D_{i2} = k] \equiv \Delta_{jk}, \quad \forall k \neq j \in [J]^2$.

In a mild abuse of our earlier notation for edge effects, this last condition simply ensures that the average treatment effect Δ_{jk} of moving from firm j to firm k among those who make this transition is not time dependent. The plausibility of this restriction will, of course, depend on nature and length of the sample period under consideration. [Lachowska et al. \(2023\)](#) and [Engbom et al. \(2023\)](#) provide empirical evidence that firm effects are quite stable over the five to seven year horizons typically studied in the literature.

The following proposition establishes that when these conditions are satisfied worker moves between pairs of firms identify average treatment effects on wages.

Proposition 1 (Firm switches identify ATTs). Under assumptions 1–3,

$$\mathbb{E}[Y_{i2} - Y_{i1}|D_{i1} = j, D_{i2} = k] = \Delta_{jk}$$

Proof. The assumptions used in each step of the below proof are listed above the equals sign:

$$\begin{aligned}\mathbb{E}[Y_{i2} - Y_{i1}|D_{i1} = j, D_{i2} = k] &\stackrel{A1}{=} \mathbb{E}[Y_{i2}(k) - Y_{i1}(j)|D_{i1} = j, D_{i2} = k] \\ &= \mathbb{E}[Y_{i2}(k) - Y_{i2}(j) + Y_{i2}(j) - Y_{i1}(j)|D_{i1} = j, D_{i2} = k] \\ &\stackrel{A2}{=} \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)|D_{i1} = j, D_{i2} = k] \\ &\stackrel{A3}{=} \Delta_{jk}.\end{aligned}\quad \square$$

Hence, contrasts of the form in (2) can identify causal estimands under plausible assumptions even if firm effects are heterogeneous. In particular, one does not need the process determining wages to be additively separable in unobserved worker and firm heterogeneity for these assumptions to hold.

While Assumptions 1–3 endow the mean wage changes accompanying firm switches with a causal interpretation, these average causal effects are not sufficient to order firms in terms of their average wage levels. The Δ_{jk} represent average treatment effects for a potentially highly selected group of movers between firm j and firm k . Without further assumptions, this heterogeneity undermines our ability to rank the potential wages offered by firms because wage changes may be intransitive. For example, with three firms, we could have $\Delta_{12} > 0$, $\Delta_{23} > 0$, $\Delta_{13} < 0$ because the workers who move between Firm 1 and Firm 3 are different from those who move between Firm 2 and Firm 3 or Firm 1 and Firm 2.¹³ Assumptions 1–3 do not even rule out the possibility that wage changes between firm pairs are asymmetric – i.e., that $\text{sign}(\Delta_{jk}) = \text{sign}(\Delta_{kj})$ – which is also a form of intransitivity.

3.3.1 Indirect contrasts and spanning trees

The AKM model enforces transitivity by imposing that $\Delta_{jk} = \psi_k - \psi_j$. We discussed in Section 3.1 how this assumption implies restrictions on edges forming a cycle. For example, if workers move from Firm 1 to Firm 2, Firm 2 to Firm 3, and Firm 3 to Firm 1, then the AKM model requires that $\Delta_{12} + \Delta_{23} + \Delta_{31} = 0$. When cyclic restrictions of this nature are violated, least squares estimation of (1) is not guaranteed to provide firm effect estimates that, when contrasted, yield a convex weighted average of treatment effects. This difficulty is familiar from both the difference in differences literature and recent work on least squares estimation in environments with multiple treatment arms (Goldsmith-Pinkham et al., 2022). As in those settings, the problem emerges, in part, from imposing over-identifying restrictions that are violated empirically. Unlike in randomized experiments, however, interpretation problems

¹³ Patterns of this nature are familiar from the social choice literature, where pairwise elections have long been observed to exhibit intransitivities in the form of Condorcet (1785) cycles. Young (1995) provides an accessible introduction to the graph theoretic interpretation of these cycles.

persist even when we saturate the model because the causal contrasts under study (i.e., the “edge effects”) pertain to potentially non-comparable populations.

It was already mentioned in [Section 3.1.3](#) that the firm effect estimates are, in general, a linear combination of all of the edge specific wage changes. The combination weights need not sum to one in each row and can be negative. These negative entries do not immediately undermine a causal interpretation of the firm effect estimates because the edge effects are directed. Returning to the graph depicted in [Fig. 1](#), in the case where a single mover is present along each edge, [equation \(4\)](#) implies that Firm 2’s fixed effect estimate can be written:

$$\begin{aligned}\hat{\psi}_2 &= \underbrace{\frac{7}{12}\hat{\Delta}_{12}}_{\text{direct}} + \underbrace{\frac{3}{12}\hat{\Delta}_{14} + \frac{1}{12}(2\hat{\Delta}_{14} - \hat{\Delta}_{23} - \hat{\Delta}_{34}) + \frac{2}{12}(\hat{\Delta}_{42} - \hat{\Delta}_{24})}_{\text{indirect}} \\ &= \underbrace{\hat{\Delta}_{12}}_{\text{direct}} + \frac{5}{12} \underbrace{(\hat{\Delta}_{14} - \hat{\Delta}_{12} - \hat{\Delta}_{23} - \hat{\Delta}_{34})}_{-c'\hat{\Delta}} + \frac{2}{12} \underbrace{(2\hat{\Delta}_{23} + 2\hat{\Delta}_{34} + \hat{\Delta}_{42} - \hat{\Delta}_{24})}_{(c_2+c_3)'\hat{\Delta}}.\end{aligned}$$

With the first firm effect normalized to zero, it is natural for $\hat{\psi}_2$ to place substantial weight on $\hat{\Delta}_{12}$, which offers a direct contrast of the wages at Firm 1 and Firm 2 for a well defined subpopulation of movers. From the first line, we see a weight of 7/12 is placed on $\hat{\Delta}_{12}$. However, the combination weights are not convex: they sum to 13/12. Moreover, indirect contrasts measuring the effects of moving between other pairs of firms contribute to $\hat{\psi}_2$. The influence of these indirect contrasts is an example of what [Goldsmith-Pinkham et al. \(2022\)](#) term “contamination.”

Under the AKM model, the indirect contrasts contain additional information about Δ_{12} . To see this, note from the second line that by rearranging terms, we can write the firm effect as the direct contrast $\hat{\Delta}_{12}$ plus two terms that have mean zero under the AKM model because they correspond to cycles. However, with unrestricted selection into edges and treatment effect heterogeneity, these indirect contrasts need not be informative about any individual’s causal effect of moving from Firm 1 to Firm 2.

Indirect contrasts can be avoided by pruning the mobility network to a polytree. As discussed in [Section 3.1.3](#), pruning the graph in [Fig. 1](#) to its first three edges yields estimates taking the form

$$\begin{bmatrix} \hat{\psi}_2 \\ \hat{\psi}_3 \\ \hat{\psi}_4 \end{bmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}}_{B_{(1)}^{-1}} \begin{bmatrix} \hat{\Delta}_{12} \\ \hat{\Delta}_{23} \\ \hat{\Delta}_{34} \end{bmatrix}.$$

Importantly, this representation holds no matter how many workers traverse each edge. Here, each firm effect is a simple sum of contrasts $\hat{\Delta}_{jk}$, which provides a causal interpretation to the difference in estimated firm effects

between any two firms that share an edge. For example $\hat{\psi}_3 - \hat{\psi}_2 = \hat{\Delta}_{23}$. However, the interpretation of differences in estimated firm effects between firms that do not share an edge is murky.

For example, the estimator $\hat{\psi}_4 = \hat{\Delta}_{12} + \hat{\Delta}_{23} + \hat{\Delta}_{34}$ is not guaranteed to reveal anything about the relative wage levels of Firm 4 and Firm 1. Fundamentally, without moves from Firm 4 to Firm 1 (which would introduce a cycle into the graph) there is no information in the data directly revealing these firms' relative wage levels for any given individual. For the wage changes of the workers moving between Firms 1 and 2 to even reveal the expected sign of the wage change associated with moving from Firm 1 to Firm 4, we need a transitivity restriction: e.g., that for any three firms $(j, k, m) \in [J]^3$, $\Delta_{jk} > 0, \Delta_{km} > 0 \Rightarrow \Delta_{jm} > 0$.

Ensuring transitivity requires either restricting the treatment effect heterogeneity or restricting selection. We will follow the tradition in the treatment effects literature of avoiding restrictions on the outcome equation and examine a restriction on selection that not only ensures a stable ordering of firms but allows cardinal comparison of firm wage levels.

3.3.2 Restricting selection

The following exogeneity assumption ensures comparability of firm wage levels based upon moves by assuming away selection on treatment effects:

Assumption 4 (No selection on treatment effects). $Y_{i2}(k) - Y_{i2}(j) \perp D_{it}, D_{i2} \forall k \neq j \in [J]^2$.

Importantly, [Assumption 4](#) permits mobility decisions to be related to *average* treatment effects $\mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]$. For example, workers can gravitate towards high wage firms as in the [Burdett and Mortensen \(1998\)](#) model. However, this assumption prohibits selection on “match” components of wages as arises in many models with comparative advantage (e.g., [Gibbons et al., 2005](#); [Eckhout and Kircher, 2011](#); [Haanwinckel, 2023](#); [Gottfries and Jarosch, 2023](#)).

When [Assumption 4](#) does hold, worker mobility identifies unconditional average treatment effects. These average treatment effects necessarily obey transitivity because they pertain to the same population, allowing firm wage levels to be ranked on a common scale. Hence, we can write $\Delta_{jk} = \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)] = \psi_k - \psi_j$, in which case least squares estimation of (1) identifies pairwise average treatment effects within the connected set of firms. We summarize this logic in the following result.

Proposition 2. If Assumptions 1–4 hold then $\Delta_{jk} = \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)] \forall k \neq j \in [J]^2$. Let $\psi = (\psi_2, \dots, \psi_J)'$, where $\psi_j = \mathbb{E}[Y_{i2}(j)] - \mathbb{E}[Y_{i2}(1)]$ for $j \in [2, \dots, J]$, and define the $J - 1 \times 1$ vector $F_{it} = (1\{D_{it} = 2\}, \dots, 1\{D_{it} = J\})'$. If the worker mobility network is connected and $\psi_1 = 0$, then $\mathbb{E}[\hat{\psi}_k - \hat{\psi}_j | \{F_{i2}, F_{i1}\}_{i \in [N]}] = \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)] \forall k \neq j \in [J]^2$, where $\hat{\psi} = [\sum_{i \in [N]} (F_{i2} - F_{i1})(F_{i2} - F_{i1})']^{-1} \sum_{i \in [N]} (F_{i2} - F_{i1})(Y_{i2} - Y_{i1}) = (\hat{\psi}_2, \dots, \hat{\psi}_J)'$.

Proof. $\Delta_{jk} = \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]$ follows directly from [Assumption 4](#). Using [Proposition 1](#), the definition of ψ , and the assumption that $\psi_1 = 0$, we have

$$\begin{aligned}\mathbb{E}[Y_{i2} - Y_{i1}|D_{i2}, D_{i1}] &= \sum_{(j,k) \in \{2, \dots, J\}^2} (\psi_k - \psi_j)(1\{D_{i2} = k\} - 1\{D_{i1} = j\}) \\ &= (F_{i2} - F_{i1})'\psi.\end{aligned}$$

Connectedness of the mobility network ensures the estimator $\hat{\psi}$ is well defined. It follows that $\mathbb{E}[\hat{\psi}|F_{i2}, F_{i1}]_{i \in [N]} = \psi$. The definition of ψ and assumption that $\psi_1 = 0$ imply $\mathbb{E}[\hat{\psi}_k - \hat{\psi}_j|F_{i2}, F_{i1}]_{i \in [N]} = \psi_k - \psi_j \forall k \neq j \in [J]^2$. \square

This proposition implies that, in the absence of selection on treatment effects, the cyclic restrictions discussed in [Section 3.1.1](#) should hold. While these restrictions offer a reasonable approximation to the edge effects, the empirical analysis in [Section 3.2](#) indicated they are unlikely to hold exactly. Of course, the same could be said of most empirical work relying on quasi-experimental variation. Nonetheless, future researchers may find it fruitful to entertain some weakenings of [Assumption 4](#).

One approach is to find richer time varying covariates that plausibly account for selection. Recent work by [Vafa et al. \(2022\)](#) demonstrates that low dimensional embeddings of employment histories can capture significant information about both potential wages and mobility, potentially restoring independence of adjusted wages. Similarly, conditioning on transition patterns less likely to be plagued by selection could improve the credibility of firm effect estimates. For example, [Di Addario et al. \(2023\)](#) show that the sequential auction model of [Bagger et al. \(2014\)](#) predicts that the AKM model restrictions should hold for the subpopulation of workers displaced from their two previous jobs and provide evidence supporting this hypothesis. The scope for selection may also be diminished among subpopulations whose transitions are prompted by plant closures or mass layoffs ([Gibbons and Katz, 1992](#)).

A second approach involves imposing a priori bounds on the maximal selection present in the network. For example, one could constrain $\max_{(j,k) \in [J]^2} |\Delta_{kj} - \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]|$ and seek estimation and inference procedures that perform well subject to this bound, utilizing extensions of the methods discussed in [Armstrong and Kolesár \(2018\)](#), [Armstrong and Kolesár \(2021\)](#), and [Rambachan and Roth \(2023\)](#).¹⁴ In some contexts it might be reasonable to consider asymmetric bounds on selection. For example, a static [Roy \(1951\)](#) selection model would posit that $\Delta_{jk} \geq \mathbb{E}[Y_{i2}(k) - Y_{i2}(j)]$ for voluntary

¹⁴ A major technical hurdle in this setting relative to conventional difference in differences problems is that most edges have very few movers, implying that normality of the estimated edge effects is not assured. However, the ultimate objects of interest in these models are typically low dimensional functionals such as variance components or projection coefficients. Recently proposed estimators of these functionals do provably converge to a normal distribution with consistently estimable variance ([Kline et al., 2020](#)).

moves. While this sort of condition can be violated in sequential auction models and models with compensating differentials, it seems reasonable to expect positive selection on wage gains more often than negative selection.

A third approach is to develop network formation models that deliver dynamic propensity scores for mobility between firms that can be used to make semi-parametric adjustments to wage changes. While some early progress has been made in this direction (Abowd et al., 2019), particularly with the use of stochastic block models (Nimczik, 2017; Jarosch et al., 2024), this literature is still in its infancy. A challenge for future work in this area is evaluating the quality of propensity score models in networks that are extremely sparse.

Finally, a number of authors depart from “design based” assumptions on selection and work with interactive factor models of wage outcomes (Bonhomme et al., 2019; Lei and Ross, 2023). These models rationalize intransitivities in edge effects in terms of latent heterogeneity in the sorts of workers that transition between different edges. To date, however, most research in this vein has worked with lower dimensional representations of the mobility network, typically by clustering firms into a small set of groups, in order to circumvent the incidental parameter biases that emerge from fitting nonlinear models to sparse networks (Chen et al., 2021). An interesting question for future research is whether the clustering step can be skipped and the structure of the underlying (uncoarsened) edge effects more fully rationalized with factor models of this nature.

4 Variance decomposition

Abowd et al. (1999) proposed summarizing the influence of firms on covariate-adjusted wage inequality via the finite sample variance decomposition

$$\mathbb{V}_n[Y_{it} - X'\beta] = \underbrace{\mathbb{V}_n[\alpha_i]}_{\text{person effect variance}} + \underbrace{\mathbb{V}_n[\psi_{j(i,t)}]}_{\text{firm effect variance}} + \underbrace{2C_n[\alpha_i, \psi_{j(i,t)}]}_{\text{sorting}} + \underbrace{\mathbb{V}_n[\varepsilon_{it}]}_{\text{noise}},$$

where n is the number of person-year observations in the sample, $\mathbb{V}_n[x_{it}] = n^{-1} \sum_{i,t} (x_{it} - \mathbb{E}_n[x_{it}])^2$, $\mathbb{E}_n[x_{it}] = n^{-1} \sum_{i,t} x_{it}$, and $C_n[\alpha_i, \psi_{j(i,t)}] = n^{-1} \sum_{i,t} \alpha_i (\psi_{j(i,t)} - \mathbb{E}_n[\psi_{j(i,t)}])$. Attention usually focuses on the firm effect variance $\mathbb{V}_n[\psi_{j(i,t)}]$, which gives a first pass measure of the importance of firms in wage determination. Note that this quantity is person-year weighted, so that the firm effects of larger firms make a greater contribution to wage inequality. The covariance component $C_n[\alpha_i, \psi_{j(i,t)}]$, which is often converted into a correlation coefficient, measures the assortativeness of worker-firm matching.

It has become common to scale the variance and covariance components by $\mathbb{V}_n[Y_{it} - X'\beta]$ in order to give each component a share interpretation. While such exercises allow a complete decomposition of residual wage inequality, the variance shares depend critically on the noise level $\mathbb{V}_n[\varepsilon_{it}]$, which can vary depending on the volatility of earnings in the country being studied, the nature

of the earnings measure (hourly, monthly, quarterly, or annual), and the demographics of the workers under study. As discussed below, cross-fitting and clustering methods both provide approaches to consistently estimating $\mathbb{V}_n[\varepsilon_{it}]$. To maximize comparability across studies then, it is advisable to scale decomposition exercises by the “signal variance” $\mathbb{V}_n[Y_{it} - X'\beta] - \mathbb{V}_n[\varepsilon_{it}]$, which captures the variability of long run expected wages of the worker-firm pairings observed in the data.

Variance shares measure the relative importance of variance components but say nothing about the absolute magnitude of variability present. Variance components are also difficult to interpret because they are measured in squared log points. Standard deviations allow a more direct assessment of the magnitude of worker and firm heterogeneity because they are measured in log points. For example, a finding that $\mathbb{V}_n[\psi_{j(i,t)}]^{1/2} = 0.25$ implies that moving to a standard deviation higher paying firm yields a roughly $[\exp(0.25) - 1] \times 100 \approx 28\%$ higher wage. Moreover, by Chebyshev’s inequality, we know that the (employment-weighted) share of firms with firm effects more than k standard deviations above the mean is at most $1/k^2$. Hence, in this example, no more than 6.25% of person year observations can be at firms with wages 100 log points or more above the mean.

4.1 Limited mobility bias

The exogenous mobility assumption guarantees that least squares will produce unbiased estimates of each fixed effect. It is therefore tempting to plug the least squares estimates $\{\hat{\alpha}_i\}_{i=1}^N$ and $\{\hat{\psi}_j\}_{j=1}^J$ into the \mathbb{V}_n and \mathbb{C}_n operators to form estimates of the relevant variance components. Unfortunately, doing so will produce biased estimates because these operators are quadratic functions. To understand the problem, observe that for any unbiased estimator $\hat{\psi}_j$ of ψ_j , we can write

$$\begin{aligned}\mathbb{E}_\varepsilon[\hat{\psi}_j^2] &= \mathbb{E}_\varepsilon[(\hat{\psi}_j - \psi_j + \psi_j)^2] = \mathbb{E}_\varepsilon[(\hat{\psi}_j - \psi_j)^2 + 2\psi_j(\hat{\psi}_j - \psi_j) + \psi_j^2] \\ &= \mathbb{V}_\varepsilon[\hat{\psi}_j] + \psi_j^2 > \psi_j^2,\end{aligned}$$

where $\mathbb{E}_\varepsilon[\cdot]$ denotes expectation with respect to the mean zero noise terms $\{\varepsilon_{it}\}_{i \in [N], t \in [T]}$ in (1) and $\mathbb{V}_\varepsilon[\cdot]$ gives the corresponding variance. Hence, estimation noise leads the square of the estimator to provide an upwardly biased estimate of the square of the parameter. A similar argument reveals that for any unbiased person effect estimator $\hat{\alpha}_i$ of α_i and any unbiased firm effect estimator $\hat{\psi}_j$ of ψ_j that $\mathbb{E}_\varepsilon[\hat{\alpha}_i \hat{\psi}_j] = \alpha_i \psi_j + \mathbb{C}_\varepsilon[\hat{\alpha}_i, \hat{\psi}_j]$. Abowd et al. (2002) termed these biases in the context of fixed effects estimation (1) “limited mobility bias” on account of the observation that if the number of movers between each firm were to grow infinitely large, the noise would disappear and the bias along with it. Andrews et al. (2008) derived the nature of the bias in plugin estimates of

the variance components in the AKM decomposition more formally and established that the covariance $\mathbb{C}_n[\hat{\alpha}_i, \hat{\psi}_{j(i,t)}]$ between person and firm effects must be biased down.

When a consistent estimator $\hat{\mathbb{V}}_\varepsilon[\hat{\psi}_j]^{1/2}$ of the standard error $\mathbb{V}_\varepsilon[\hat{\psi}_j]^{1/2}$ is available, one can form a bias corrected estimate of each ψ_j^2 with $\hat{\psi}_j^2 - \hat{\mathbb{V}}_\varepsilon[\hat{\psi}_j]$; that is, by subtracting off the squared standard error from the plugin estimate. Likewise, an unbiased estimate of the variance of firm wage effects $\theta_\psi = \mathbb{V}_n[\psi_{j(i,t)}] = \mathbb{E}_n[\psi_{j(i,t)}^2] - \mathbb{E}_n[\psi_{j(i,t)}]^2$ can be obtained from its debiased analogue

$$\hat{\theta}_\psi = \mathbb{E}_n[\hat{\psi}_{j(i,t)}^2 - \hat{\mathbb{V}}_\varepsilon[\hat{\psi}_{j(i,t)}]] - \{\mathbb{E}_n[\hat{\psi}_{j(i,t)}]^2 - \hat{\mathbb{V}}_\varepsilon[\mathbb{E}_n[\hat{\psi}_{j(i,t)}]]\}. \quad (5)$$

[Krueger and Summers \(1988\)](#) implemented a bias correction of this form when computing the variance of industry wage fixed effects. Replacing $\hat{\psi}_{j(i,t)}$ with $\hat{\alpha}_i$ in the above formula yields a bias corrected variance of person effects $\hat{\theta}_\alpha$. The bias corrected covariance between person and firm effects can be obtained from the formula $\hat{\theta}_{\alpha,\psi} = \mathbb{E}_n[\hat{\alpha}_i \hat{\psi}_{j(i,t)} - \hat{\mathbb{C}}_\varepsilon[\hat{\alpha}_i, \hat{\psi}_{j(i,t)}]]$.

[Andrews et al. \(2008\)](#) proposed a correction for the AKM variance and covariance components under the assumption that the ε_{it} are *iid*. However, these corrections yielded small changes in the variance components, corrections that appeared to be too small given the magnitude of the biases found in the subsampling exercises reported in [Andrews et al. \(2012\)](#). [Card et al. \(2013, Online Appendix 3\)](#) conjectured that this under-correction was likely a result of unmodeled heteroscedasticity and serial correlation in wage innovations, properties that had been well documented in the literature on earnings dynamics (e.g., [MaCurdy, 1982](#); [Abowd and Card, 1989](#); [Meghir and Pistaferri, 2004](#)).

It is tempting to use conventional heteroscedasticity-consistent standard errors to estimate and remove the bias. However, these “standard errors” and their bootstrap analogues are known to exhibit bias when the number of parameters being estimated is proportional to the number of observations ([Bickel and Freedman, 1981](#); [MacKinnon and White, 1985](#); [Mammen, 1993](#); [Cattaneo et al., 2018](#); [El Karoui and Purdom, 2018](#)). [Kline et al. \(2020\)](#) proposed replacing the usual heteroscedasticity consistent standard errors (e.g., [White, 1980](#); [MacKinnon and White, 1985](#)) with heteroscedasticity *unbiased* variance estimates derived from cross-fitting that are robust to arbitrary heteroscedasticity.

4.2 Cross-fitting and bias correction

Cross-fitting can be thought of as a version of sample splitting designed to remove overfitting biases while making maximally efficient use of the data

(Newey and Robins, 2018).¹⁵ To understand the logic behind this approach, it is useful to rewrite (1) in the notation

$$Y_m = \mathbf{D}_m \alpha + \mathbf{F}_m \psi + \varepsilon_m, \quad (6)$$

where Y_m is a vector of all of the wages in worker-firm match $m \in \{1, \dots, M\} \equiv [M]$. For example, if a worker spends three years at a job, then that match yields a 3×1 vector of wages. The matrix \mathbf{D}_m is comprised of worker dummies; it has as many rows as match m has time periods and it has N columns corresponding to each worker in the sample. The vector α collects the person effects. The matrix \mathbf{F}_m is comprised of firm dummies. We assume one firm effect has been normalized to zero so that \mathbf{F}_m has $J - 1$ columns and the vector ψ collects $J - 1$ firm effects. I have again abstracted from the time varying covariates, which can be partialled out in a first stage. Throughout this discussion, we will treat the $\{\mathbf{D}_m, \mathbf{F}_m\}_{m \in [M]}$, along with (α, ψ) as fixed, leaving ε_m as the only source of randomness in the model. The $\{\varepsilon_m\}_{m \in [M]}$ are assumed to be mutually independent and to exhibit mean zero.

We will write $\mathbb{V}_\varepsilon[\varepsilon_m] = \boldsymbol{\Omega}_m$ which conveys both that the noise level may vary from match to match and that arbitrary within match correlation of the errors ε_m is permitted. The variance of least squares estimates of firm effects takes the usual “sandwich” form

$$\mathbb{V}_\varepsilon[\hat{\psi}] = \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1} \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \boldsymbol{\Omega}_m \tilde{\mathbf{F}}_m \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1},$$

where $\tilde{\mathbf{F}}_m$ is the matrix of firm dummies that results after partialling out the worker dummies – i.e., after deviating the firm indicators from their worker specific means. Hence, if we knew the $\{\boldsymbol{\Omega}_m\}_{m \in [M]}$, we could compute “match clustered” standard errors that allow us to bias correct the square of each firm effect by subtracting off its squared standard error.

Let $\hat{\psi}_m$ denote the vector of firm effects derived from fitting (1) by least squares when leaving out the observations for match m and $\hat{\alpha}_{-m}$ the corresponding vector of person effects. For $\hat{\alpha}_{-m}$ to exist, we need that every worker has at least two worker-firm matches. Assume for the moment then that the sample has been restricted to job switchers so that $\hat{\alpha}_{-m}$ exists. This assumption is without loss of generality since job stayers do not contribute to estimation of the firm effects but only to the firm weights used to define the variance of interest. Another option is to consider long differences – i.e., to omit all but the first and last periods – and to treat the first and last wage error of job stayers as independent, which may be plausible in longer panels.

Define the cross-fit residual as

¹⁵ Sorkin (2018), Drenik et al. (2023), and Card et al. (2024) estimate vectors of firm effects using two independent half samples of workers. The covariance between the two samples provided an unbiased (and transparent) estimate of the variance of the latent firm effects. Unfortunately, the connected set can grow much smaller when the sample is split and randomness in how the split was chosen contributes to the variability of the estimator.

$$\hat{\varepsilon}_m = Y_m - \mathbf{D}_m \hat{\alpha}_{-m} - \mathbf{F}_m \hat{\psi}_{-m} = \varepsilon_m + \xi_{-m},$$

where $\xi_{-m} \equiv \mathbf{D}_m(\alpha - \hat{\alpha}_{-m}) + \mathbf{F}_m(\psi - \hat{\psi}_{-m})$ is a mean zero vector of noise arising from estimation error in the coefficients $(\hat{\alpha}_{-m}, \hat{\psi}_{-m})$. Note that $\mathbb{E}_\varepsilon[\varepsilon_m \xi'_{-m}] = 0$ because the noise is independent across matches. Hence, unlike traditional regression residuals, which tend to be too small due to overfitting, the cross-fit residuals are generally too large, as $\mathbb{E}_\varepsilon[\hat{\varepsilon}_m \hat{\varepsilon}'_m] = \Omega_m + \mathbb{E}_\varepsilon[\xi_{-m} \xi'_{-m}]$.

[Kline et al. \(2020\)](#) propose multiplying the cross-fit residual by the outcome, which yields the unbiased estimator

$$\hat{\Omega}_m = Y_m \hat{\varepsilon}'_m = (\mathbf{D}_m \alpha + \mathbf{F}_m \psi + \varepsilon_m)(\varepsilon_m + \xi_{-m})'.$$

Unbiasedness of $\hat{\Omega}_m$ for Ω_m follows from the observation that $\mathbf{D}_m \alpha + \mathbf{F}_m \psi$ is a matrix of constants and the presumed independence of ε_m from ξ_{-m} .¹⁶ Hence, an unbiased estimator of $\mathbb{V}_\varepsilon[\hat{\psi}]$ is

$$\hat{\mathbb{V}}_\varepsilon[\hat{\psi}] = \left(\sum_{m \in [M]} \tilde{\mathbf{F}}'_m \tilde{\mathbf{F}}_m \right)^{-1} \left(\sum_{m \in [M]} \tilde{\mathbf{F}}'_m \hat{\Omega}_m \tilde{\mathbf{F}}_m \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}'_m \tilde{\mathbf{F}}_m \right)^{-1}. \quad (7)$$

Note that unlike the classic HC2 and HC3 estimators of [MacKinnon and White \(1985\)](#), the cross-fit variance estimator is unbiased for any sample size. A corresponding formula for the covariance between person and firm effects is provided in the appendix.

Unbiasedness does not guarantee that the variance estimate for any particular firm effect will be accurate. In fact, a necessary consequence of unbiasedness is that there must be some probability that the realized variance estimate for each of the diagonal terms $\hat{\mathbb{V}}_\varepsilon[\hat{\psi}_j]$ is negative. However, [Kline et al. \(2020\)](#) show that weighted averages of the estimated variances, such as the average estimated noise level $\mathbb{E}_n[\hat{\mathbb{V}}_\varepsilon[\hat{\psi}_{j(i,t)}]]$ are guaranteed to converge to $\mathbb{E}_n[\mathbb{V}_\varepsilon[\hat{\psi}_{j(i,t)}]]$ as the sample size grows large. If we have restricted estimation to the firm movers, we can also compute the weighted average noise level, which reweights the firms according to their share of all person-year observations including the firm stayers. Consequently, unbiased estimation of the variance of firm effects does not require taking a stand on the serial correlation of the stayer wage errors. Bias corrected estimates of firm variance components can often be measured quite precisely. For instance, [Kline et al. \(2020\)](#) obtain a bias corrected point estimate of the person-year weighted variance of firm effects of 0.024 (which we will replicate shortly) with a corresponding standard error of only 0.0006.

¹⁶ As mentioned earlier, one typically preadjusts log wages for time varying covariates in a first step, which introduces a small higher order bias due to estimation error $\hat{\beta} - \beta$ influencing both y_m and $\hat{\varepsilon}_m$. Even so, it is often wise to ensure y_m has mean zero before applying cross-fitting in order to reduce the variability of $\hat{\Omega}_m$ and we will do so in our empirical example below.

4.2.1 Leave-out connectedness

An important requirement of cross-fitting methods is that the model must be estimable after leaving out any particular observation. Within a given connected set, many firms may be connected by only a single move, implying their ψ_j would not be estimable if that worker's wage observations were dropped. Kline et al. (2020) find when using only two periods of data that 43 % of the firms in the largest connected set are “just-connected” in this manner. Workers who move to or from such firms have no residual associated with their wage change, prohibiting an assessment of the level of noise in their wages and consequently a bias correction. Fundamentally then, the variance of firm effects is only identified within the leave-out connected set that prunes the just-connected firms.

To assess how this pruning might change estimands, Kline et al. (2020, Table IV) report the results of further restricting the set of firms under study to be connected when any *two* matches are left out. Requiring that each firm effect be estimable when any two matches are left out further reduces the number of estimable firm effects by 43 %. Surprisingly, the effect of this restriction turns out to be negligible, nudging the point estimate of the variance of firm effects from 0.240 to 0.238. Similar insensitivity to these leave out requirements is found for subsamples of older and younger workers. One reason for this insensitivity is that weakly connected firms typically employ few workers and hence make a small contribution to the overall (person-year weighted) variance of firm effects. Another is that in finite samples, there is a large degree of randomness in which firms happen to be connected, a phenomenon consistent with standard random search models exhibiting Poisson arrival of mobility events.

4.2.2 Bounding and imputation

Existing applications of the cross-fitting correction report variance components describing heterogeneity within the leave-out connected set of workers and firms. Moving the goalposts to estimate whatever target parameter is identified by a research design is standard fare in empirical economics (Crump et al., 2009; Imbens, 2010). It is nonetheless prudent to examine the extent to which the leave-out connected set might differ from the broader population of workers and firms. Fortunately, it is relatively straightforward to compute bounds on variance components describing the broader connected set of firms.

The key insight that allows the construction of bounds is to note that the noise level Ω_m in any match must obey the bound: $0 \leq \Omega_m \leq \mathbb{E}_\varepsilon[Y_m Y']$. The upper bound follows from observing that

$$\mathbb{E}_\varepsilon[Y_m Y'] = (\mathbf{D}_m \alpha + \mathbf{F}_m \psi)(\mathbf{D}_m \alpha + \mathbf{F}_m \psi)' + \Omega_m.$$

The first term in this sum is the outer product of a vector and therefore must be positive semi-definite. Consequently, the upper bound is sharp, arising when $D_m\alpha + F_m\psi$ equals a vector of zeros.

Intuitively, the wages associated with a just-connected match could be pure noise, in which case $\Omega_m = \mathbb{E}_\varepsilon[Y_m Y']$, or they could be entirely noiseless, in which case $\Omega_m = 0$. This observation suggests estimating bounds on $\mathbb{V}_\varepsilon[\hat{\psi}]$ using $\hat{\Omega}_m = Y_m Y'$ as an upper bound and $\hat{\Omega}_m = 0$ as a lower bound on the noise contribution of just-connected matches. As before, we use the leave-out estimator $\hat{\Omega}_m = Y_m \hat{\varepsilon}'$ for leave-out connected matches. Denote the resulting estimated upper and lower bounds by $\hat{\mathbb{V}}_\varepsilon^+[\hat{\psi}]$ and $\hat{\mathbb{V}}_\varepsilon^-[\hat{\psi}]$ respectively. Plugging these bounds on the noise level into (5) yields a corresponding lower bound estimate $\hat{\theta}_\psi^-$ and upper bound estimate $\hat{\theta}_\psi^+$ on the variance of firm effects $\theta_{\psi} = \mathbb{V}_n[\psi_{j(i,t)}]$:

$$\hat{\theta}_\psi^- = \mathbb{E}_n \left[\hat{\psi}_{j(i,t)}^2 - \hat{\mathbb{V}}_\varepsilon^+[\hat{\psi}_{j(i,t)}] \right] - \left\{ \mathbb{E}_n[\hat{\psi}_{j(i,t)}]^2 - \hat{\mathbb{V}}_\varepsilon^+[\mathbb{E}_n[\hat{\psi}_{j(i,t)}]] \right\},$$

$$\hat{\theta}_\psi^+ = \mathbb{E}_n \left[\hat{\psi}_{j(i,t)}^2 - \hat{\mathbb{V}}_\varepsilon^-[\hat{\psi}_{j(i,t)}] \right] - \left\{ \mathbb{E}_n[\hat{\psi}_{j(i,t)}]^2 - \hat{\mathbb{V}}_\varepsilon^-[\mathbb{E}_n[\hat{\psi}_{j(i,t)}]] \right\}.$$

These bounds, which are consistent for the corresponding population bounds under conditions that parallel the case where all observations are leave-out connected, may be especially useful in thin samples or environments characterized by low mobility. A corresponding approach to bounding the covariance $\theta_{\alpha,\psi} = \mathbb{C}_n(\alpha_i, \psi_{j(i,t)})$ is detailed in the appendix.

4.2.3 An empirical example

To illustrate these ideas, we now return to the benchmark VHW sample introduced in Section 3.2. With two years of data, estimating the firm effects in levels and first differences is numerically equivalent, which implies that we can think of Y_m as a scalar measuring wage changes without loss of generality. As shown in the top panel of Table 2, roughly 83 % of movers in the largest connected set are also in the leave-out connected set. These leave-out connected workers exhibit mildly higher average wages that are slightly less dispersed. The lower panels of the table report estimates of the AKM variance decomposition in each sample.

Squaring the plug-in and bias corrected standard deviations of firm effects reported in the second column of Table 2 reproduces the firm effect variances reported in Table II of Kline et al. (2020). The cross-fitting bias correction has substantial bite in the leave-out sample, cutting the estimated standard deviation of firm effects from 18.9 to 15.5 log points. It is natural to worry, however, that this bias reduction constitutes a pyrrhic victory, as the roughly 40 % of connected firms that are not leave-out connected may differ from those that are connected. How large of a bias correction would we have obtained if we knew the error variances in the original connected set?

TABLE 2 Sample composition and variance components in Veneto, Italy.

	Connected set	Leave-out connected set
Number of Person Year Observations	1,859,459	1,319,972
Number of Movers	197,572	164,203
Number of Firms	73,933	42,489
Mean Log Wage	4.7507	4.8066
Standard Deviation of Log Wage	0.4455	0.4293
Standard Deviation of Firm Effects		
Plug-in	0.2161	0.1892
Bias Corrected	[0.1421, 0.1847]	0.1549
Connected at Random	0.1643	
Standard Deviation of Person Effects		
Plug-in	0.3712	0.3634
Bias Corrected	[0.3216, 0.3423]	0.3345
Connected at Random	0.3320	
Impute Stayer Noise Level	[0.3138, 0.3353]	0.3267
CAR (Impute Stayer Noise)	0.3245	
Covariance of Firm and Worker Effects		
Plug-in	-0.0053	0.0039
Bias Corrected	[0.0063, 0.0188]	0.0146
Connected at Random	0.0128	

Notes: This table reports properties of the connected and leave-out connected sets in a panel comprised of the 1999 and 2001 waves of the Veneto Work Histories dataset developed by the Economics Department in Università Ca' Foscari Venezia under the supervision of Giuseppe Tattara. The standard deviation of firm effects refers to the person-year weighted standard deviation of firm effects in a regression of log daily wages on worker fixed effects, firm fixed effects, and a year fixed effect. "Plug-in" refers to the OLS estimates. "Bias corrected" estimate uses the cross-fit bias correction. Intervals correspond to bounds on the variance component in question resulting from the assumption that the noise levels of just-connected movers equal either zero or that mover's squared wage change. "Connected at random" bias corrects by imputing the average error variance of leave-out connected movers to just-connected movers. "Impute Stayer Noise Level" bias corrects assuming that workers who don't switch jobs have the average noise level of leave-out connected movers. Intervals again correspond to bounds on the variance component in question resulting from the assumption that the noise levels of just-connected movers equal either zero or that mover's squared wage change. "CAR (Impute Stayer Noise)" bias corrects assuming that both just-connected movers and job stayers exhibit the average noise level of leave-out connected movers.

The first column of the bottom panel of [Table 2](#) sheds light on this question. The upper bound on the variance of firm effects assumes that the matches at just-connected firms have error variance zero, yielding an upper bound on the standard deviation of firm effects of 18.5 log points. Coincidentally, this upper bound is very near the plug-in estimate of the standard deviation of firm effects in the leave-out connected sample. Conversely, if we assume matches at each just-connected firm have error variance equal to the squared wage change involving that firm, then we attain a lower bound standard deviation of 14.2 log points. Finally, if the just-connected matches exhibit a noise level equal to the average of the leave-out connected matches – an assumption that I will term “connected at random” (CAR) – then we can form a bias correction by imputing for every just-connected match the average cross-fit noise level of wage changes of movers in the leave-out connected set.¹⁷ The CAR imputation yields an estimated standard deviation of firm effects of 16.4 log points.

Evidently, the bias corrected standard deviation estimate in the leave-out connected sample is almost exactly halfway between the lower bound and CAR estimates in the broader connected sample. Moreover, the range of estimates is relatively narrow. Little seems to have been lost here by restricting to the leave-out connected set. If the CAR estimate had been very different from the bias corrected estimate, however, we might have come away more concerned about selection bias. Hence, the CAR estimate seems like a useful diagnostic to report in addition to the standard bias-corrected estimates describing the leave-out connected set.

An equivalent exercise can be conducted with the variance of person effects and the covariance between person and firm effects. Bias correcting the standard deviation of person effects in the leave-out connected set reduces its magnitude by about 3 log points, which is comparable to the effects of bias correction on the standard deviation of firm effects. In the broader sample of connected workers, the bounds on the standard deviation are quite narrow, ranging from 32.2 to 34.2 log points. Like the bias corrected estimate in the leave-out connected set, the CAR estimate of the standard deviation of person effects in the broader connected set is 33 log points.

As was noted earlier, it is possible that the noise level of job stayers has been underestimated by neglecting serial correlation. While job stayers do not contribute to estimation of firm effects, they are essential for the estimation of person effects. Underestimation of job stayer noise levels could therefore lead to overestimation of the variance of person effects. To assess this possibility, we also report the person effect standard deviation that would result if job stayers had the same average noise level as job movers. Doing so in the

¹⁷ One could argue that this assumption should be termed “connected completely at random” (CCAR) as the imputation is not conditioned on any covariates.

leave-out connected set yields a marginally smaller person effect standard deviation of 32.7 log points. In the broader connected sample, this imputation lowers both the upper and lower bounds on the person effect standard deviation by slightly less than a log point. Likewise, the CAR estimate in the connected sample falls by nearly a log point and is essentially indistinguishable from the estimate in the leave-out connected sample.

Finally, bias correcting the covariance between worker and firm effects in the leave-out connected set yields small increases. Fortunately, bias correcting the covariance does not require recovering the noise level of stayers because the covariance must reflect estimation error in firm effects, which depend entirely on movers. In the broader connected sample, the bounds are again fairly narrow. Moreover, the CAR estimate of covariance is close to the bias corrected covariance in the leave out connected set.

Using the bias corrected estimates in the leave out connected sample yields a correlation coefficient of 0.28. If we ascribe to the stayers the noise level of the movers, the correlation rises negligibly to 0.29 because the person effect variance falls. In the broader connected set the correlation is an increasing function of the unknown noise level of the just-connected movers. Consequently, we can obtain lower bound under the assumption that the noise level is zero and an upper bound under the assumption that the noise level is given by the squared wage change. It turns out that this yields a non-trivial range of possible correlation coefficients [0.09, 0.40]. However, these bounds entertain the implausible possibility that the wage changes of just-connected movers are either all noise or all signal. The CAR estimate of correlation is 0.23 and imputing the mover noise level to the stayers raises this correlation negligibly to 0.24. These estimates are quite close to our bias corrected estimate in the leave out connected sample, suggesting that selection is probably not a major concern here.

In sum, we can be relatively confident that trimming has little effect on the person-year weighted variance of worker or firm effects. More ambiguity is present regarding the correlation between worker and firm effects but the agreement between CAR estimates and estimates in the leave-out connected set suggest selection bias is also likely to be mild in this dimension. Future research in this area could consider more sophisticated imputation schemes that allow noise levels of just-connected workers to be estimated based upon features of the worker-firm mobility network. Finally, our experimentation with imputation schemes for the noise levels of job stayers suggests that person effect variances are unlikely to be dramatically overstated by cross-fitting approaches neglecting the serial correlation of job stayers. In settings where serial correlation is a known concern, the proposed imputation strategy based on the average estimated noise level of movers offers a potentially attractive way of circumventing the problem.

4.3 Clustering approaches

[Bonhomme et al. \(2019\)](#) analyze a version of the AKM model in which firm heterogeneity is restricted to be discrete. They assume the firm effects can be represented in a lower dimensional space via the relation

$$\psi_j = \sum_{k=1}^K T_{jk} \bar{\psi}_k, \quad (8)$$

where the $\{T_{jk}\}_{k=1}^K$ are indicators for the latent type of the j 'th firm effect obeying $\sum_{k=1}^K T_{jk} = 1$ and the $\{\bar{\psi}_k\}_{k=1}^K$ are the wage effects of those firm types. In their baseline specification, they work with $K = 10$ types, a choice that has been focal in the subsequent literature.

Directly imposing (8) and optimizing jointly over the indicators T_{jk} and the locations $\bar{\psi}_k$ via nonlinear least squares is a non-convex and often intractable computational problem. To circumvent this obstacle, [Bonhomme et al. \(2019\)](#) propose a two step approach. First, they apply a variant of K -means clustering ([Forgy, 1965; Lloyd, 1982](#)) to firm wage distributions to obtain firm type assignments \hat{T}_{jk} . These type assignments are then treated as regressors in second step estimation of the model

$$Y_{it} = \alpha_i + \sum_{k=1}^K \hat{T}_{j(i,t)k} \bar{\psi}_k + X'\beta + \varepsilon_{it}.$$

Rather than estimate this equation by OLS, they treat the α_i as normal mixtures with means that depend on \hat{T}_{jk} , which further reduces the number of parameters to be estimated, and maximize the likelihood via the EM algorithm ([Dempster et al., 1977](#)). Once the type specific parameters have been estimated, the type estimates can (in principle) be updated, yielding reclassified firm and worker type assignments that provide approximations to one step maximum likelihood estimates of the full model.

In some respects, the clustering approach mirrors the earlier literature on industry wage differentials (e.g., [Krueger and Summers, 1988](#)). Rather than using as regressors indicators for 20 or so 2-digit industries, the “industries” are treated as latent random variables to be reconstructed via clustering of firm wage distributions. By reducing the high dimensional AKM specification down to a low dimensional model, the clustering approach sidesteps the usual incidental parameters problem, substantially reducing the biases associated with squaring estimated parameters. Clustering also circumvents the requirement to limit the analysis to the largest connected set of firms, as one only needs the estimated firm types \hat{T}_{jk} , rather than each individual firm, to be connected by worker mobility for the second step model to be estimable. Interactions between estimated worker and firm types can also be treated as regressors, allowing estimation of non-separable models. These interactions turn out to be negligible in Swedish data, however, raising the estimated R^2 of the model

from 74.8 % to 75.8 %, leading Bonhomme et al. (2019) to conclude that “complementarities explain only a small part of the variance of log-earnings.”¹⁸

The advantages of the clustering approach come at the cost of strong assumptions on the data generating process. For one thing, it seems implausible that there exist large groups of firms that offer exactly the same wage premiums. The restriction in (8) is at best an approximation and one that inevitably leads to understatement of firm effect variances by neglecting within-type variability. Neglected covariances between any within firm type employer heterogeneity and worker heterogeneity can also lead to bias in the estimates of worker-firm sorting. Card et al. (2023) note both of these problems when revisiting the industry wage differential literature, where they find substantial variation in employer wage premiums within industry along with significant worker-firm sorting. It seems unlikely that any partition of firms into 10 or even 10,000 groups would entirely resolve these problems.

Even if (8) were to hold exactly, the type assignments \hat{T}_{jk} will be noisy for small firms, which can generate bias in the estimated locations parameters $\{\hat{\psi}_k\}_{k=1}^K$. Indeed, the formal assumptions used by Bonhomme et al. (2019) to establish consistency of the two step clustering approach require that the number of wage observations at the smallest firm grow with the number of firms. This potential for bias that arises with finite sized firms is a cost of having to estimate a regressor instead of relying on a predetermined grouping such as industry, firm size, or geography. Another cost concerns interpretability. While some judgement calls are involved in choosing industry and geographic categories, variation across them is substantially easier to interpret than variation across firm groups determined via K -means clustering of wage distributions.

A related conceptual difficulty is that the type assignments are determined based on cross-sectional wage distributions rather than worker mobility. However, any cross-sectional distribution of wages could be driven by worker sorting rather than firm heterogeneity. The ability to separate the two comes only from the assumption that the economy possesses a finite number of well separated firm types. Parametric identification of this nature is contrary to the ethos of the AKM approach, which relies entirely on worker mobility to separate worker and firm heterogeneity.

The robustness exercises reported in Bonhomme et al. (2019, Table III) reveal that the estimated variance of firm effects can, in fact, be quite sensitive to the details of the procedure used to form the type assignments. They find

¹⁸ In fact, their estimated interactions are smaller than those reported by Card et al. (2013, Table III) for German data, who find that allowing for unrestricted worker-firm match effects raises the adjusted R^2 by roughly 2 % points. In a recent analysis of US earnings data, Lamadon et al. (2022, Table A6) report that adding worker-firm interactions to an additive group fixed effects model fit to raises the R^2 by less than one percentage point.

that clustering firms into ten groups based on their cross-sectional wage distributions yields a variance of firm effects that accounts for 2.6 % of the overall variance of earnings in Swedish administrative data. Splitting those groups by firm value added raises the share of wage variance explained by firms to 3.4 %. Reclassifying the firm types – which can be thought of as choosing the firm groups to directly approximate the firm effects of the movers – raises the estimated contribution of firm effects to 4.1 % of the variance.

A recent paper by [Bonhomme et al. \(2023\)](#) relaxes (8) by assuming

$$\psi_j = \sum_{k=1}^K T_{jk} (\bar{\psi}_k + v_k), \quad (9)$$

where each $\{v_k\}_{k=1}^K$ is a mean zero normally distributed random effect with a different variance. By allowing for within firm type dispersion, this correlated random effects (CRE) approach generally picks up a greater degree of firm dispersion. For instance, [Lamadon et al. \(2022, Table A6\)](#) find that firm effects explain only 3.2 % of annual earnings variance in US tax data when using the two-step estimator imposing (8), whereas [Bonhomme et al. \(2023, Table F2\)](#) estimate that share at 6.2 % in a six year panel of the same data using the CRE estimator predicated on (9). However, the CRE estimator still relies on functional form assumptions to separate worker and firm types. In particular, the estimator is predicated on moment conditions imposing that $\alpha_i - \mathbb{E}[\alpha_i | T_{jk}]$ is independent of v_k , which implies there is no worker firm sorting within firm types, while the type assignments \hat{T}_{jk} are still based on a first step clustering routine applied to the cross-sectional wages of job stayers.

[Bonhomme et al. \(2023\)](#) find in both Monte Carlo exercises and real datasets that both their CRE estimator and the cross-fitting estimator of [Kline et al. \(2020\)](#) successfully address limited mobility bias. On average the parametric CRE estimator yields modestly smaller firm effect estimates than the bias corrected estimator based on cross-fitting. It is difficult to assess the extent to which these differences arise from violations of the functional form assumptions baked into the CRE model. A traditional justification for CRE methods is that, by exploiting additional restrictions, they can offer more efficient (albeit less robust) estimates ([Chamberlain, 1982; Angrist and Newey, 1991](#)). Monte Carlo evidence suggests that the CRE estimates of variance components are indeed likely to be more efficient than the cross-fitting estimator when the CRE assumptions hold. Hence, the CRE approach may be useful in small samples where precision is a practical concern. Another potentially important use case for the CRE estimator is settings with extremely limited mobility, where restricting to the leave out connected set would drop an unacceptably large share of the units under study (e.g., [Fenizia, 2022](#)). When using such approaches, it may be worthwhile to pursue iteratively updated versions of the estimator, which have been found to yield improved performance in some settings ([Bonhomme et al., 2019; Lentz et al., 2022](#)).

4.4 How variable are worker and firm effects?

Bias corrected estimates of worker and firm contributions to wage inequality have now been reported in many countries. The figure below depicts bias-corrected estimates of worker and firm effect variability drawn from nine recent studies utilizing the cross-fitting correction of Kline et al. (2020). Rather than focus on variances or variance shares, I compare the standard deviation of person effects to the standard deviation of firm effects, the units of which are directly interpretable in log points. When reported, multiple specifications from the same study are included to illustrate the sensitivity of estimates to the sample period and population. The list of studies depicted is provided in Appendix Table A.1. Some studies that used bias corrections could not be included because they failed to report the magnitude of the variance components, relying on variance shares without reporting the marginal variance.

The 45 degree line through the origin of Fig. 3 gives what one should expect if worker and firm components are equally important and scale with the overall level of inequality in an economy. Perhaps surprisingly, many of the estimates lie very near this line. As expected, the scale of inequality appears most pronounced in middle income countries such as Mexico, South Africa, and Brazil, while Italy, the US, and Sweden are relatively more equal in both dimensions. The estimates falling below the 45 degree line come predominantly from high income countries and from Brazil. Interestingly, these studies all find comparable standard deviations of firm effects near 0.25. However, the standard deviations of worker effects vary widely from sample to sample.

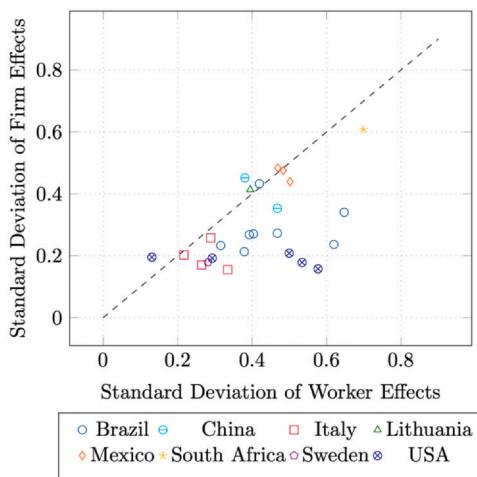


FIG. 3 Bias corrected standard deviations of firm and worker fixed effects by country.

To some extent, this variability of person effect variances is to be expected given that many of the estimates partition by race or sex, groups within which we expect person effects to be less dispersed. For example, four of the Brazilian estimates are from [Gerard et al. \(2021\)](#), who report estimates separately by race and sex using a 12 year panel, which accounts for some of the lowest worker effect standard deviations in that country. However, person effect variances also seem to vary with other features of the data including the time horizon studied.

[Abowd and McKinney \(2023\)](#), for example, find a nearly identical standard deviation of firm effects in 3 year and 24 year extracts of annualized earnings records from the LEHD. However, in the 3 year panel, the bias corrected standard deviation of person effects is roughly 50 % larger than the standard deviation of firm effects, while in the 24 year panel, the person effects exhibit a standard deviation roughly 33 % below that of the firm effects. Likewise, [Lachowska et al. \(2023\)](#) find using hourly wage data from Washington state that person effects are substantially more dispersed in a 2 year panel than a 12 year panel. While it is tempting to conclude that this sensitivity to time scale reflects drift in the person effects, [Lachowska et al. \(2023\)](#) demonstrate that person effect estimates remain strongly correlated across decades.

Recall that the variance of person effects among firm stayers cannot be estimated by cross-fitting at the match level. A majority of the studies considered include firm stayers and it is reasonable to assume that these studies treat the errors of firm stayers as serially independent, as this is the default option provided in the most widely used software package used to implement the cross-fitting correction.¹⁹ The estimated person effect variances may therefore be subject to an upward bias stemming from neglected serial correlation, albeit a smaller one than if no correction were implemented. It seems likely then that the tendency for shorter panels to yield larger person effect variances reflects this tendency to under-correct, as adjacent observations are more strongly correlated. If the person effect variances are in fact upwardly biased due to serial correlation, then it is even more surprising that so many studies yield estimates near the 45 degree line.

Though the estimated person effect variances appear sensitive to sample composition, the firm effect standard deviations are remarkably resilient. Among the estimates depicted here, the firm effect standard deviations all exceed 0.15 and for high income countries cluster around 0.20. A potentially useful comparison comes from [Bonhomme et al. \(2023\)](#), who estimated decompositions using the CRE clustering method in harmonized datasets of annualized earnings from five high income countries (Austria, Italy, Norway, Sweden, and the U.S.) but declined to report a variance of person effects. Averaging their estimates of the standard deviation firm effects across

¹⁹ See <https://github.com/rsaggio87/LeaveOutTwoWay> for details.

countries and samples yields a mean value of roughly 0.14, which is a bit below the estimates reported for rich countries in the figure above.²⁰ Two features of Bonhomme et al. (2023)'s harmonization procedure contribute to their finding of smaller average firm effect variances: i) they study annual earnings rather than daily or hourly wages and ii) they impose a minimum earnings threshold of approximately the annualized minimum wage. Bonhomme et al. (2023) note that relaxing either of these restrictions boosts the estimated variance of firm effects (see Figures F2 and F10 of their online appendix). While one can debate which earnings records should be discarded when activity measures are unavailable, it is uncontroversial that when days or hours worked are available, this information should be used to provide a more accurate wage measure that circumvents the need to select on the dependent variable. A reasonably informed guess then, is that across a wide range of high income countries, the standard deviation of firm effects in daily or hourly wages typically ranges between 15 and 20 log points.

For middle income countries, the standard deviation of firm effects appears to be higher, perhaps as high as 0.4 in some cases. It is plausible that firm effects are more important in developing countries, where search frictions and misallocation have been argued to be more prevalent (Hsieh and Klenow, 2009). However, many of these studies are very recent and have yet to clear peer review. It will be important to see estimates from more countries and research teams before drawing strong conclusions about the relationship between economic development and the dispersion in firm pay components.

The median firm effect standard deviation estimate among all those pictured in Fig. 3 is 0.26 and an unweighted average of them is 0.30. If, in high income countries, the standard deviation of firm wage effects is somewhere between 15 and 20 log points, then switching to a standard deviation higher firm yields wages 16–22 % higher – a very substantial effect size. For comparison, Chetty et al. (2011) estimate that a standard deviation increase in kindergarten classroom quality in the Project STAR experiment raises adult earnings by 13 % points. These findings suggest workplace heterogeneity is an important contributor to wage inequality.

5 Regressing firm effects on observables

Examining how fixed effects covary with observables can help to demystify the nature of these fundamentally unobservable objects. Many of the empirical findings summarized in Section 2 were derived from regressing estimated firm fixed effects on observed features of workers and firms. Besides greater robustness to modeling assumptions, an important advantage of fixed effects methods over more structured random effects approaches (e.g. Hanushek,

²⁰ Bonhomme et al. (2023) report the match weighted variance of firm effects rather than the person-year weighted variance.

1974; Amemiya, 1978) is that fixed effect estimates can be shared with different research teams, who can subsequently use them to examine different downstream hypotheses via “second step” regressions. I will now review the logic of these downstream regressions and discuss the subtleties of inference on second step projection coefficients. These ideas will be illustrated with an example to the firm size wage premium in the VHW data.

5.1 One step vs two

Suppose we are interested in the relationship between the vector of population firm effects ψ and a set of firm covariates such as firm size and the average education level of the firm’s employees. Descriptive relationships of this nature are often summarized with linear projections of the form

$$\psi = \mathbf{Z}\theta + v,$$

where \mathbf{Z} is a matrix of firm covariates and the parameter of interest is $\theta = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\psi$. By construction, the projection error v obeys $\mathbf{Z}'v = 0$. Plugging this relationship into (6) yields:

$$Y_m = \mathbf{D}_m\alpha + \mathbf{F}_m\mathbf{Z}\theta + \mathbf{F}_m v + \varepsilon_m.$$

Since the projection error v is orthogonal to \mathbf{Z} , one might be tempted by this representation to estimate θ from a least squares regression of Y_m on $(\mathbf{D}_m, \mathbf{F}_m\mathbf{Z})$ – i.e., on person dummies plus the firm characteristics. There are two difficulties with this logic. The first objection, which is largely pedantic, has to do with weighting. The cross product $\sum_{m \in [M]} (\mathbf{F}_m\mathbf{Z})'\mathbf{F}_m v = \sum_{m \in [M]} \mathbf{Z}'\mathbf{F}'\mathbf{F}_m v$ will not, in general, equal zero unless all firms are the same size. Hence, orthogonality need not hold in the microdata even if it holds across firms. Of course, if we had initially defined the estimand θ as the firm size weighted projection, then the relevant v would satisfy orthogonality in the microdata.

A more significant objection is that even if $\mathbf{F}_m v$ is uncorrelated with $\mathbf{F}_m\mathbf{Z}$, it is still likely to be correlated with \mathbf{D}_m . The fact that higher wage workers tend work at higher wage firms suggests $\sum_{m \in [M]} \mathbf{D}'\mathbf{F}_m v > 0$, which violates the exogeneity requirements of least squares. This violation will not only tend to generate bias in the estimated person effects but also in estimates of θ because $\mathbf{F}_m\mathbf{Z}$ is correlated with \mathbf{D}_m . Hence, unless one has a strong reason to suspect that the elements of \mathbf{Z} account for all of the correlation between worker and firm wage effects, dropping the firm dummies as controls (i.e., treating v as an uncorrelated random effect) will tend to generate bias.

The two step approach is to first compute the fixed effects $\hat{\psi}$ and then regress them on \mathbf{Z} to obtain the projection coefficient $\hat{\theta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{\psi}$. Under strict exogeneity, the firm effects are unbiased. The projection coefficient, which is just a linear combination of the estimated firm effects, inherits this

property, obeying $E_\varepsilon[\hat{\theta}] = \theta$. Hence, the two step estimator provides robust estimates of the projection regardless of the dependence between worker and firm effects.

Another advantage of the two step estimator is that it can foster scientific cooperation: the research team that produces $\hat{\psi}$ need not be the team that has access to \mathbf{Z} . Fixed effects estimates are often computed once on population microdata by expert researchers and then made available to outside teams who do may not have access to the same microdata files (e.g., [Bellmann et al., 2020](#)). These sorts of data sharing arrangements enable a broader range of hypotheses and external data sources to be brought to bear on questions of scientific interest.

5.2 Variance estimation

The variance of the estimated projection coefficient is

$$\mathbb{V}_\varepsilon[\hat{\theta}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbb{V}_\varepsilon[\hat{\psi}]\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}.$$

While second step regressions will yield unbiased estimates of linear projection coefficients, the standard errors produced by conventional software packages will mistakenly assume that the noise $\hat{\psi} - \psi$ in the second step regressand is independent across firms – i.e., that $\mathbb{V}_\varepsilon[\hat{\psi}]$ is diagonal. Neglecting correlation between the estimated firm effects can lead to severe understatement (or overstatement) of the uncertainty in second step regression coefficients.

The sign of this bias in the estimated standard errors is theoretically ambiguous because the residuals from the second step regression will tend to overstate the intrinsic noise level of each estimated fixed effect. To take an extreme example, suppose that the wage disturbances ε_{it} in (1) are exactly zero. In such a case, the first step regression will fit perfectly, yielding $\hat{\psi} = \psi$. However, a second step regression of firm fixed effects on observed firm characteristics will nonetheless yield residuals capturing unexplained variation in the vector ψ of true firm effects. Consequently, conventional software packages will produce a positive standard error estimate despite the fact that the true firm effects are fixed and exhibit no uncertainty.

Standard errors reflecting only the uncertainty associated with the ε_{it} are easily computed by using the cross-fit variance estimates introduced in (7). For example, [Kline et al. \(2020\)](#) considered a second step regression wherein \mathbf{Z} included a constant, the share of workers over age 35, firm size, and their interaction. Note that, as in our earlier discussion of cross-fitting, interest centers on the finite population of J firms actually measured in our dataset rather than an abstract “super-population” from which those firms were drawn. Replacing the unknown $\mathbb{V}_\varepsilon[\hat{\psi}]$ with $\hat{\mathbb{V}}_\varepsilon[\hat{\psi}]$ yields an unbiased estimate of the variance of the second step regression coefficient that can be used for

inference. Fortunately, computation does not require that the entire $\hat{V}_\epsilon[\hat{\psi}]$ matrix be computed or stored.²¹

While it is straightforward for research agencies to release fixed effect estimates to the public and their (squared) standard errors, it is not feasible to release entire variance matrices. In principle, one could conduct inference relying only on the fixed effect standard errors by considering worst case correlation patterns. However, doing so could lead to extremely conservative inferences. An interesting area for future work is understanding what low dimensional features of $\hat{V}_\epsilon[\hat{\psi}]$ can be reported that would enable accurate inference on projection coefficients without knowledge of the Z under consideration by the research team.²²

5.3 Revisiting the firm size wage premium

Fig. 4 illustrates the use of these methods by studying how the relationship between firm effects and firm size varies by province in Veneto. Returning to the firm effect estimates studied in **Table 2**, the matrix Z is chosen to include indicators for the firm size categories utilized by Bloom et al. (2018) interacted with indicators for which of Veneto's seven provinces contains the firm in question. As a normalization, the smallest firm size category of 1–10 employees has been set to zero in each province, so that each of the included estimates represents a within province firm size "premium." To reduce clutter, we have dropped the province of Rovigo which is so small that it lacks any firms in the largest two size categories. By contrast, more than one thousand firms are present in each size category of the pictured provinces.

Confidence intervals based on naive heteroscedasticity-robust standard errors computed via a second step regression are shown alongside those based on the cross-fit variance matrix $\hat{V}_\epsilon[\hat{\psi}]$. The cross-fit standard errors reflecting uncertainty attributable to ϵ turn out to be about 75 % larger than the naive standard errors on average. As a result the 95 % confidence intervals based on cross-fitting turn out to bracket those based on naive standard errors in all cases. Evidently, the downward bias in naive standard errors attributable to neglecting correlation among the estimated firm effects outweighs the upward bias attributable to treating the firm effects as random draws from a broader population.

In all six pictured provinces, firm effects tend to increase with firm size. However, the size profiles differ substantially across provinces and in some cases appear non-monotone. In Verona and Padova the largest firms exhibit fixed effects averaging approximately 40 log points more than the smallest firms, while in Venice the corresponding gap in firm wage effects is only about 9 log points. These orderings reverse, however, in the next largest firm size

²¹ A computationally efficient approach to estimation of $\hat{V}_\epsilon[\hat{\psi}]$ is automated and detailed in the LeaveOutTwoWay package available at <https://github.com/rsaggio87/LeaveOutTwoWay>.

²² For example, in a lower dimensional context, Firth and De Menezes (2004) propose reporting "quasi-variances" that can be used for inference on unknown contrasts.

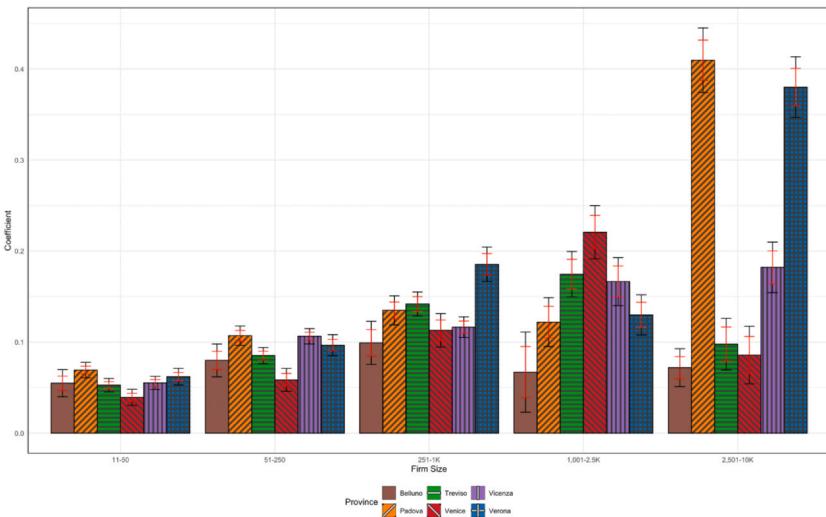


FIG. 4 Mean firm effects by firm size and province. Notes: Sample comprised of firms in leave-out connected set described in Table 2. Bar height gives coefficient from second step regression of firm effects onto province indicators plus interactions with indicators for firm size category. Omitted firm size category in each province is 1–10. Both confidence intervals derived by adding ± 1.96 standard errors to the point estimate. Outer confidence interval (depicted in black) relies on $(Z'Z)^{-1}Z'\hat{\psi}Z(Z'Z)^{-1}$ as estimator of the asymptotic variance. Inner confidence interval (depicted in red) relies on $\frac{J}{J-k}(Z'Z)^{-1}Z'(\text{diag}(M_Z\hat{\psi}))^2Z(Z'Z)^{-1}$ as estimator of the asymptotic variance, where $M_Z = I - Z(Z'Z)^{-1}Z'$.

category. In Venice, for example, firms with 1000–2500 employees are estimated to pay roughly 22 log points more than firms with 1–10 employees, while in Verona the premium is only 13 log points.

While the premiums relative to the base firm size category are precisely estimated in each province, it is not completely obvious which of these premiums differ from one another given that the estimates are all correlated. A useful rule of thumb is that we can conclude that the estimands are different from one another if their confidence intervals do not overlap.²³ Based on this heuristic, we can safely infer that in both Venice and Treviso the firm size premium in the 1000–2500 employee category exceeds the premium in the 2500–10,000 employee category, indicating that the firm size premiums are not monotone in these regions.

²³ For any two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, we have $\mathbb{V}(\hat{\theta}_1 - \hat{\theta}_2) = \mathbb{V}(\hat{\theta}_1) + \mathbb{V}(\hat{\theta}_2) - 2C(\hat{\theta}_1, \hat{\theta}_2) \leq \{\mathbb{V}(\hat{\theta}_1)^{1/2} + \mathbb{V}(\hat{\theta}_2)^{1/2}\}^2$, where the upper bound binds with equality when the two estimators are perfectly negatively correlated. Consequently, $\mathbb{V}(\hat{\theta}_1)^{1/2} + \mathbb{V}(\hat{\theta}_2)^{1/2}$ provides a conservative standard error on the difference between the estimators. A test that evaluates whether $|\hat{\theta}_1 - \hat{\theta}_2| > c \cdot [\mathbb{V}(\hat{\theta}_1)^{1/2} + \mathbb{V}(\hat{\theta}_2)^{1/2}]$ for some critical value c (e.g., 1.96 as in Fig. 4) amounts to evaluating whether $[\hat{\theta}_1 - c\mathbb{V}(\hat{\theta}_1)^{1/2}, \hat{\theta}_1 + c\mathbb{V}(\hat{\theta}_1)^{1/2}] \cap [\hat{\theta}_2 - c\mathbb{V}(\hat{\theta}_2)^{1/2}, \hat{\theta}_2 + c\mathbb{V}(\hat{\theta}_2)^{1/2}] = \emptyset$.

We can also infer that size premiums tend to differ by region, though our visual rule of thumb becomes less decisive in smaller firm size categories. To obtain a more accurate assessment of regional differences in the average premium for firms with 10–50 employees, I reparameterize \mathbf{Z} to include interactions between province and firm size categories. The resulting standard error estimates reveal that it is possible to reject at the 5 % level the null hypothesis that the premiums in the 10–50 employee category are equal in Venice and Vicenza. By contrast, the premiums in Vicenza and Treviso cannot be distinguished from each other even at the 10 % level.

6 Hiring origins and state dependence

The basic AKM specification views wage determination as fundamentally static: the expected wage arising from a match between a worker and firm depends only on their underlying (time-invariant) types. Search theoretic models, by contrast, often predict that wages should be influenced by the circumstances surrounding how the match was formed – e.g., whether the worker was “poached” from another firm or hired from unemployment, as unemployed workers typically have worse outside options than their employed counterparts. Consistent with this view, [Faberman et al. \(2022\)](#) provide survey based evidence that job offers received by currently employed workers pay higher wages than those received by unemployed workers with similar characteristics.

An influential framework for modeling such state dependence comes from the class of sequential auction models pioneered by [Postel-Vinay and Robin \(2002a,b\)](#), where on the job search gives rise to a series of bilateral competitions between firms for workers. These competitions mirror first price auctions, with firms tailoring their wage bids based upon the willingness to pay of the rival they face. Consequently, the wages offered to new hires differ based on where a worker is hired from and which firm is hiring them. Tailoring of this nature can, in principle, contribute greatly to cross-sectional inequality by amplifying the role of luck: an early job displacement can lower wages throughout a worker’s career by persistently degrading their outside options.

[Di Addario et al. \(2023\)](#) study the empirical predictions of sequential auction models for hiring wages using an extension of the AKM model, in which a separate fixed effect is allowed for each possible hiring origin. These hiring origin fixed effects are meant to proxy for the worker’s outside option. Letting y_{im} denote the log hiring wage of the i th worker at their m ’th job, they consider a linear model taking the form:

$$Y_{im} = \alpha_i + \underbrace{\psi_{j(i,m)}}_{\text{destination effect}} + \underbrace{\lambda_{h(i,m)}}_{\text{origin effect}} + X'\delta + \varepsilon_{im}, \quad \text{for } i \in [n], m \in [M_i]. \quad (10)$$

Here, the function $\mathbf{j}: [n] \times [M_i] \rightarrow [J]$ returns the identity of the firm hiring the worker at their m 'th job. The function $\mathbf{h}: [n] \times [M_i] \rightarrow [J] \cup \{U\}$ returns the origin of the new hire, which can either be the identity of a prior employer from which the worker was “poached” or unemployment (denoted as “ U ”). Thus, each firm j has a pair (ψ_j, λ_j) of fixed effects.

[Di Addario et al. \(2023\)](#) term the specification in (10) a “dual wage ladder” (DWL) model because hiring wages depend on two dimensions of firm heterogeneity. As in the AKM model, α_i is a person fixed effect that can be ported from employer to employer, while the vector X_{im} includes time varying covariates, including work experience and indicators for the year that the match was formed. The term $\psi_{j(i,m)}$ is a destination firm fixed effect that, like the traditional AKM firm effect, must be forfeited upon separating from the employer. The distinctive feature of the DWL specification is the origin firm fixed effect, $\lambda_{h(i,m)}$, which captures a form of state dependence in wage setting. According to the DWL model, two workers with the same α_i , hired by the same firm from two different origins – e.g., non-employment and the most productive firm in the economy – will be paid different wages.

The error ε_{im} measures omitted factors that vary across matches at the time of hiring. Each of these errors is assumed to have mean zero, which is a version of the traditional exogenous mobility assumption used to justify least squares estimation. An important feature of standard sequential auction models is that bilateral competitions are presumed to be efficient: i.e., the more productive firm always wins the auction. If firm productivity is time invariant, then conditioning on $\mathbf{j}(i, m)$ and $\mathbf{h}(i, m)$ is equivalent to conditioning on the productivity of the origin and destination firm, which given log-linear wage contracts implies the errors $\{\varepsilon_{im}\}_{i \in [n], m \in [J]}$ are strictly exogenous.

6.1 Structural interpretation

[Di Addario et al. \(2023\)](#) show formally that the model of [Bagger et al. \(2014\)](#), which nests the seminal model of [Postel-Vinay and Robin \(2002b\)](#) when consumption utility is assumed to be logarithmic, yields (10) as the reduced form for hiring wages. It is useful to review this argument both to understand the structural interpretation of the origin and destination fixed effects and the justification for the exogenous mobility assumption on the reduced form errors. The [Bagger et al. \(2014\)](#) model implies that the log hiring wage offered by a firm of productivity level p , to a worker of productivity type ϵ , with labor market experience X , who is currently employed at a firm of productivity q can be written as the generalized linear function

$$\alpha(\epsilon) + g(X) + \psi(p) + \lambda(q) + \varepsilon.$$

Hires from unemployment follow the same equation with the productivity of the incumbent firm q set equal to the flow value of leisure b , which is assumed to be common for all workers.

The term $\alpha(\epsilon)$ is a worker fixed effect capturing general human capital, which is rewarded equally by all employers. Likewise, $g(X)$ captures the returns to experience, while the error term \mathcal{E} captures idiosyncratic innovations to the worker's general human capital. By assumption, neither of these terms influence worker mobility, which depends solely on the firm productivities p and q . The destination firm effect, $\psi(p)$, equals $\beta \ln p + I(p, \beta)$, while the hiring origin effect, $\lambda(q)$, is given by $(1 - \beta) \ln q - I(q, \beta)$, where $\beta \in [0, 1]$ indexes worker bargaining strength. The function $I(p, \beta)$, which is decreasing in both its arguments and obeys $I(p, 1) = 0$, captures the expected utility of the wage growth associated with moving from a firm with productivity p to the most productive firm in the economy. Hence, the difference $I(p, \beta) - I(q, \beta)$ captures the expected utility of the wage growth associated with moving from an incumbent firm with productivity q to a poaching firm with productivity p .

Inspection of these equations reveals that when β is small, the destination effect $\psi(p)$ will be decreasing in p , which can be interpreted as a compensating differential for the anticipated wage growth associated with moving. By contrast, for any value of $\beta < 1$, $\lambda(q)$ will be increasing in q , which reflects that it is more difficult to poach workers from firms that can afford to pay them more. When $\beta = 1$, the term $\lambda(q)$ becomes zero and the model reduces to a version of the AKM model with only destination firm effects. Remarkably, $\psi(p) + \lambda(p) = \ln p$ for any value of β , implying that a firm's productivity can be recovered by summing its origin and destination effects. Since workers view more productive firms as fundamentally more desirable than less productive firms, this sum recovers the ordering of the underlying "job ladder" in expected utility governing worker flows.

6.2 Testable restrictions

In the [Bagger et al. \(2014\)](#) model, firms are differentiated only by productivity. Consequently, the origin and destination effects are deterministic functions of one another. [Di Addario et al. \(2023\)](#) show that it is possible to exploit this feature of the model to bound the bargaining power of workers using the excess variance of the destination effects over the origin effects. Letting \mathbb{V}_p denote the variance across firms, the following bound on β is obtained by exploiting the fact that $\frac{\partial}{\partial \ln p} I(p, \beta) \in [-(1 - \beta)^2/\beta, 0]$:

$$\beta \geq 1/2 + \frac{\mathbb{V}_p[\psi(p)] - \mathbb{V}_p[\lambda(p)]}{2\mathbb{V}_p[\psi(p) + \lambda(p)]}. \quad (11)$$

As discussed earlier, if β were very close to 1, we should expect the origin effects to be negligible and for destination effects to be large as workers extract from firms the greatest wage they can afford: $\ln p$. This bound formalizes the converse idea that when destination effects are large relative to origin effects, worker bargaining power must be strong. When $\beta > 1/2$, the following lower

bound can be shown to hold on the correlation between the two dimensions of firm heterogeneity:

$$\text{corr}(\psi(p), \lambda(p)) \geq \sqrt{\frac{\mathbb{V}_p[\psi(p)]}{\mathbb{V}_p[\psi(p) + \lambda(p)]}} \left(1 - \frac{3}{10} \sqrt{\frac{\mathbb{V}_p[\lambda(p)]}{\mathbb{V}_p[\psi(p) + \lambda(p)]}}\right).$$

Intuitively, when β is large, both the origin and destination effects must be strongly increasing in productivity, yielding a high correlation. However, a large β also yields relatively larger destination effects than origin effects. The correlation bound formalizes this link, effectively providing a test of the presence of a unidimensional firm hierarchy.²⁴

6.3 It ain't where you're from, it's where you're at

Di Addario et al. (2023) fit (10) to Italian social security data using the average daily wage of each worker in their first year of employment with a firm as a proxy for their hiring wage. A poaching event is presumed to have taken place whenever a worker resigns from their job as opposed to being laid off or fired for cause. If the worker did not resign from their previous job, they are assumed to have been hired from unemployment. While there are reasons to suspect that stated resignations provide an imperfect perfect proxy of when bilateral competition between firm pairs is taking place (McLaughlin, 1991; Postel-Vinay and Turon, 2014), Italian workers poached according to this criterion turn out to have much shorter durations of non-employment between jobs than workers involved in other sorts of separations.

Di Addario et al. (2023) find a roughly 3.5 log point gap between the estimated value of λ_U (the origin effect associated with unemployment) and the average origin effect of poached workers, $\mathbb{E}_n[\lambda_{h(i,m)} | h(i, m) \neq U]$, implying a modest penalty for being hired from unemployment. The bias corrected variance of origin effects, $\mathbb{V}_n[\lambda_{h(i,m)}]$, turns out to be extremely small, accounting for less than 1 % of the variance of hiring wages across job movers. By contrast, the variance of destination effects, $\mathbb{V}_n[\psi_{j(i,m)}]$, explains 24 % of the variance of hiring wages.

As mentioned in Section 4, variance shares can be somewhat difficult to interpret given that noise levels vary across samples. The estimated standard deviation of destination effects in their sample of job movers is roughly 0.26, which is only slightly above the typical bias corrected standard deviation of AKM firm effects reported for the US and Italy in Fig. 3. By contrast, the origin effects have a standard deviation among all job movers of 0.04 and a standard deviation of 0.08 among the roughly 1/3 of job transitions that involve poaching a worker from another firm.

While an 8 % wage change is not negligible, this standard deviation of origin effects turns out to be far less than would be predicted by the Bagger et al. (2014)

²⁴ Roussille and Scuderi (2023) reject a unidimensional model of firm valuations in favor of a mixture model with three distinct hierarchies using data from an online job board for software engineers.

model. The standard deviation across firms of the destination effects, $\sqrt{V_p}[\psi(p)]^{1/2}$, is 0.26 (the same as was found across workers), while the corresponding standard deviation of origin effects, $\sqrt{V_p}[\lambda(p)]^{1/2}$, is only 0.07. Applying the formula in (11) implies that $\beta \geq 0.88$. In addition to being intuitively implausible, this value of β would require an extremely high correlation between the origin and destination effects of 0.84. In practice, the bias corrected correlation is only 0.25, indicating that the model cannot rationalize the covariance structure of the origin and destination effects under *any* distribution of firm productivities.

6.4 Information and conduct

The order of magnitude difference in scale between firm origin and destination effects suggests either that the identity of one's current employer doesn't convey much information about outside options at the time of a poaching attempt or that firms are unable (or unwilling) to tailor offers to those outside options. To assess the former hypothesis, one could collect more granular proxies of outside options. Perhaps interacting the identity of the incumbent firm with detailed job titles or tenure would be more predictive of hiring wages? The second possibility, that firms are not able or willing to tailor wage offers, is more difficult to evaluate. Firms often report having some latitude to tailor wages to worker circumstances and [Caldwell et al. \(2024a\)](#) provide evidence that wages are strongly related to previous firm pay among those firms that engage in bargaining. On the other hand, survey evidence suggests offer matching is rare empirically ([Faberman et al., 2022](#); [Caldwell et al., 2024a](#)). Moreover, receiving an outside offer does not seem to be associated with large wage gains on average ([Guo, 2023](#)).

Even if firms typically do have the ability to tailor wages, the informational requirements of tying wage offers to best predictors of outside options are formidable. Sequential auction models are predicated on a perfect information benchmark where each firm knows the willingness to pay of the rival firm for the worker in question, leading them to offer a rival dependent wage.²⁵ By contrast, the famous [Burdett and Mortensen \(1998\)](#) model effectively assumes that firms know nothing about workers' outside options, which is why they are willing to commit to offering the same wages to unemployed workers and workers searching on the job. As [Postel-Vinay and Robin \(2002a\)](#) acknowledge "reality lies somewhere in between our complete information story and Burdett's and Mortensen's incomplete information assumption."

How to think about this middle ground between wage posting and sequential auction models remains a frontier area of research. One approach is to view the economy as comprised of a mixture of wage posting firms ala [Burdett and Mortensen \(1998\)](#) and tailoring firms ala [Postel-Vinay and Robin](#)

²⁵ Workers are also assumed to be fully informed about the match surplus available at the two rival firms. [Jäger et al. \(2024\)](#) provide evidence suggesting that workers at low wage firms tend to underestimate their outside options.

(2002b). While coherent models of this nature have been proposed (Postel-Vinay and Robin, 2004; Flinn and Mullins, 2017), empirical evidence on how wage setting conduct varies across employers remains in its infancy. A recurrent finding from estimation of these models is that counter offers and negotiation are more common among higher skilled workers (Caldwell and Harmon, 2019; Flinn and Mullins, 2021). This finding likely resonates among academic economists, many of whom have experienced the majority of their salary growth by receiving outside offers. Indeed, the sequential auction paradigm of bilateral competition appears to be a good one for academia, which is a hierarchical industry where employers have good information about the ability of rival institutions to compete for talent. It is unclear how many other labor markets are characterized by this sort of competition.

Breaking their variance decompositions down by industry, Di Addario et al. (2023) find that destination effects are orders of magnitude more variable than origin effects in most sectors of the Italian economy. The key exceptions are finance/banking and the legal sector, where origin and destination effects exhibit comparable variability. Both of these sectors are hierarchical and plausibly exhibit more information regarding the ability of firms to pay to retain workers than other sectors. The finance/banking industry is the only sector where the correlation bound is satisfied, suggesting perhaps that it too exhibits the sort of unidimensional competition described in sequential auction models. In less skilled sectors, by contrast, employers are likely more difficult to rank. As a result, less information may be conveyed by the identity of one's previous employer. In these settings, worker outside options seem more likely to be private information, an idea that is central to the idea of monopsonistic models of wage setting.

7 Conclusion

While much has been learned about which firms pay high wages and their contribution to wage inequality, plenty of work remains. Some questions this review has touched upon that appear particularly ripe for exploration include:

- 1. Dispersion and Development:** Why are firm wage effects more dispersed in less developed countries? One possibility is that labor market frictions are more pronounced in these economies, leading to greater misallocation. Another is that measurement differences, especially the prevalence of informal work, play a confounding role.
- 2. Accounting for Cycles:** What accounts for the cyclic component of edge effects? Cycles could reflect either economic shocks shared by closely connected firms or differences in the sorts of workers moving along different parts of the mobility network. The former view has difficulty explaining the documented stability of firm effects. The latter interpretation suggests something important may have been missed by existing models of non-separable wages, estimates of which typically exhibit small departures from linearity.

3. **Intransitive Firms:** To what extent do firm rankings, in both wages and desirability, vary with worker and job characteristics? Does accounting for this heterogeneity amplify or mute the total contribution of firms to inequality?
4. **Hiring Origins and Conduct:** When and where do hiring origins matter for wage determination? Do markets where the dispersion of origin effects is larger exhibit greater wage effects of receiving outside offers? How does the reason for separation (e.g., ostensible quits vs layoffs) influence the degree of state dependence in wages?
5. **Worker Mobility Post-Layoff:** Why do mass layoffs sometimes lead workers to move to higher-wage firms? Does the prevalence of this behavior vary with labor market institutions?
6. **Understanding Network Structure:** What network formation models produce realistic mobility patterns? How effective are these models at predicting the next firm that will employ a worker? How do network-based definitions of labor markets align with workers' perceptions as measured in surveys?
7. **Reproducibility:** How can fixed effect estimates be shared most effectively? Transparency and replicability are crucial components of the data science revolution ([Donoho, 2024](#)). Future work could enable wider access not only to point estimates but also to measures of uncertainty, lowering the barriers to downstream inference and prediction.

Appendix: Covariance between person and firm effects

Here, I detail how to construct an unbiased estimator of $\mathbb{C}_\varepsilon[\hat{\alpha}_i, \hat{\psi}_j]$ for each (i, j) pair in $[N] \times [J - 1]$ that can be used to bias correct the covariance. I then discuss how bounds can be formed on the covariance.

From (6), we can write the OLS estimators

$$\hat{\alpha} = \alpha + \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \tilde{\mathbf{D}}_m \right)^{-1} \sum_{m \in [M]} \tilde{\mathbf{D}}_m' \varepsilon_m,$$

$$\hat{\psi} = \psi + \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1} \sum_{m \in [M]} \tilde{\mathbf{F}}_m' \varepsilon_m,$$

where $\tilde{\mathbf{D}}_m$ is the matrix of worker indicators after having partialled out the matrix of firm indicators. Hence,

$$\begin{aligned} (\hat{\alpha} - \alpha)(\hat{\psi} - \psi)' &= \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \tilde{\mathbf{D}}_m \right)^{-1} \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \varepsilon_m \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \varepsilon_m \right)' \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1} \\ &= \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \tilde{\mathbf{D}}_m \right)^{-1} \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \varepsilon_m \tilde{\varepsilon}_m' \tilde{\mathbf{F}}_m \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1} \\ &\quad + \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \tilde{\mathbf{D}}_m \right)^{-1} \left(\sum_{m \in [M]} \sum_{l \neq m} \tilde{\mathbf{D}}_m' \varepsilon_m \varepsilon_l' \tilde{\mathbf{F}}_l \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1}. \end{aligned}$$

Independence across matches implies that the final line has expectation zero, allowing us to write

$$\mathbb{E}_\varepsilon[(\hat{\alpha} - \alpha)(\hat{\psi} - \psi)'] = \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \tilde{\mathbf{D}}_m \right)^{-1} \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \Omega_m \tilde{\mathbf{F}}_m \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1}.$$

We can estimate this covariance matrix with

$$\hat{\mathbb{E}}_\varepsilon[(\hat{\alpha} - \alpha)(\hat{\psi} - \psi)'] = \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \tilde{\mathbf{D}}_m \right)^{-1} \left(\sum_{m \in [M]} \tilde{\mathbf{D}}_m' \hat{\Omega}_m \tilde{\mathbf{F}}_m \right) \left(\sum_{m \in [M]} \tilde{\mathbf{F}}_m' \tilde{\mathbf{F}}_m \right)^{-1}.$$

The lower triangle of this estimated matrix gives the relevant unbiased estimators $\hat{C}_\varepsilon[\hat{\alpha}_i, \hat{\psi}_j]$ of $C_\varepsilon[\hat{\alpha}_i, \hat{\psi}_j]$. The debiased estimator of covariance between person and firm effects is:

$$\hat{\theta}_{\alpha, \psi} = \mathbb{E}_n[\hat{\alpha}_i \hat{\psi}_{j(i,t)}] - \hat{C}_\varepsilon[\hat{\alpha}_i, \hat{\psi}_{j(i,t)}].$$

To bound this covariance in the broader connected sample that is not leave out connected, we can again apply the bound $0 \leq \Omega_m \leq \mathbb{E}[Y_m Y_m']$. Upwardly and downwardly biased estimators of the relevant covariances $\mathbb{E}_\varepsilon[(\hat{\alpha} - \alpha)(\hat{\psi} - \psi)']$ are obtained by replacing $\hat{\Omega}_m$ with $Y_m Y_m'$ or zero, respectively, in just-connected matches contributing to $\hat{\mathbb{E}}_\varepsilon[(\hat{\alpha} - \alpha)(\hat{\psi} - \psi)']$. One then applies the bias correction formula above replacing $\hat{C}_\varepsilon[\hat{\alpha}_i, \hat{\psi}_{j(i,t)}]$ with either its upwardly or downwardly biased estimate (Fig. A.1) and (Table A.1).

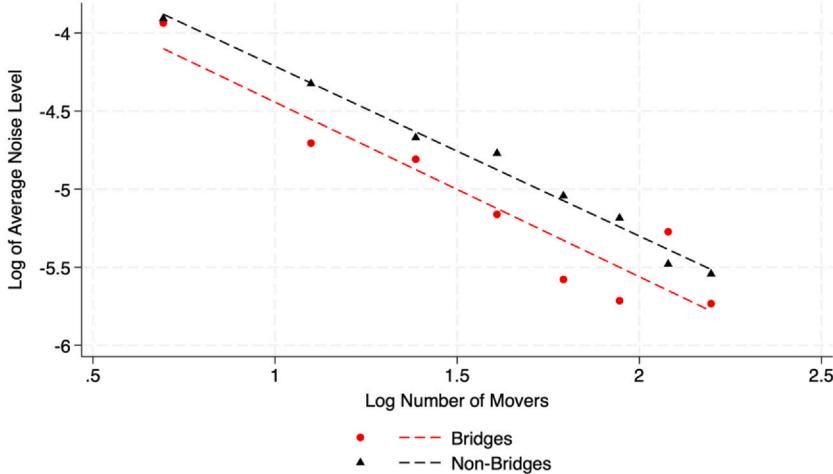


FIG. A.1 Noise level by number of movers per edge. Notes: The vertical axis depicts the natural logarithm of the average of $\hat{\Psi}[\hat{\Delta}_e]$ among edges with a given number of movers. Number of movers considered ranges from 2 to 10. The line of best fit (depicted above) has intercept -3.325 and slope -1.118 when fit on edges that are bridges and an intercept of -3.126 and slope -1.087 when fit on edges that are not bridges. Using these lines of best fit to impute the variance of singleton bridges, the imputed noise level of singleton bridges is $\exp(-3.325) \approx 0.036$ and the imputed noise level for singleton edges that are not bridges is $\exp(-3.12) \approx 0.044$.

TABLE A.1 Studies included in Fig. 3.

Study	Source	Country / Region	Years	Total Var	Worker Var	Firm Var	2 * Cov
Kline et al. (2020)	Table 2	Veneto, Italy	1999, 2001	0.184	0.112	0.024	0.029
Lachowska et al. (2023)	Table 3	Washington State	2002–2014	0.407	0.250	0.043	0.075
	Table 2	Washington State	2002–2003	0.360	0.285	0.032	0.024
Engbom and Moser (2022)	Table 2	Washington State	2013–2014	0.426	0.333	0.025	0.052
	Table 2	Brazil	1994–1998	0.709	0.176	0.187	0.120
		Brazil	2014–2018	0.444	0.154	0.072	0.070
Haanwinckel (2022)	Table 1	Brazil	1998–2001	0.688	0.419	0.116	0.098
		Brazil	2011–2013	0.577	0.384	0.056	0.097
Engbom et al. (2023)	Table 3	Sweden	1985–2016	0.268	0.079	0.032	0.005
Bassier	Table 2	South Africa	2011–2016	1.320	0.488	0.370	0.106
Casarico and Lattanzio (2023)	Table D.2	Italy, Men	1995–2015	0.176	0.070	0.029	0.043
		Italy, Women	1995–2015	0.160	0.047	0.041	0.024

Gerard et al. (2022)	Table 2	Brazil, White men Brazil, non-White men	2002–2014 2002–2014	0.449 0.332	0.163 0.100	0.073 0.054	0.102 0.058
		Brazil, White women	2002–2014	0.498	0.219	0.075	0.123
		Brazil, non-White women	2002–2014	0.324	0.144	0.045	0.060
Di Addario et al. (2023)	Table 3 / A.3	Italy	2005–2015	0.279	0.083	0.066	0.047
Garcia-Louzao and Ruggieri	Table 2	Lithuania	2000–2020	0.595	0.156	0.171	0.053
Perez Nuno-Ledesma (2022)	Table A.4	Mexico	2004–2008	0.596	0.252	0.193	0.099
		Mexico	2009–2013	0.627	0.234	0.226	0.111
		Mexico	2014–2018	0.628	0.220	0.234	0.121
Abowd and McKinney (2023)	Table 4	LEHD	2012–2014	3.330	0.086	0.037	0.023
		LEHD	1994–2017	3.423	0.017	0.038	0.018
Guo et al. (2024)	Table 2	China, Native China, Migrant	2006–2014 2006–2014	0.530 0.572	0.219 0.145	0.125 0.204	0.054 0.081

References

- Abowd, J.M., Card, D., 1989. On the covariance structure of earnings and hours changes. *Econometrica* 57 (2), 411–445.
- Abowd, J.M., Creecy, R.H., Kramarz, F., et al., 2002. Computing person and firm effects using linked longitudinal employer-employee data. Technical Reports, Center for Economic Studies (US Census Bureau).
- Abowd, J.M., Kramarz, F., Lengermann, P., McKinney, K.L., Roux, S., 2012. Persistent inter-industry wage differences: rent sharing and opportunity costs. *IZA Journal of Labor Economics* 1, 1–25.
- Abowd, J.M., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. *Econometrica* 67 (2), 251–333.
- Abowd, J.M., Lengermann, P., McKinney, K.L., 2003. The measurement of human capital in the US economy. Tech. rep., Citeseer.
- Abowd, J.M., McKinney, K.L., 2023. Mixed-effects methods for search and matching research. arXiv preprint arXiv:2308.15445.
- Abowd, J.M., McKinney, K.L., Schmutte, I.M., 2019. Modeling endogenous mobility in earnings determination. *Journal of Business & Economic Statistics* 37 (3), 405–418.
- Akerlof, G.A., Yellen, J.L., 1990. The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics* 105 (2), 255–283.
- Amemiya, T., 1978. A note on a random coefficients model. *International Economic Review* 793–796.
- Andrews, M.J., Gill, L., Schank, T., Upward, R., 2008. High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (3), 673–697.
- Andrews, M.J., Gill, L., Schank, T., Upward, R., 2012. High wage workers match with high wage firms: clear evidence of the effects of limited mobility bias. *Economics Letters* 117 (3), 824–827.
- Angrist, J.D., Newey, W.K., 1991. Over-identification tests in earnings functions with fixed effects. *Journal of Business & Economic Statistics* 9 (3), 317–323.
- Armstrong, T.B., Kolesár, M., 2018. Optimal inference in a class of regression models. *Econometrica* 86 (2), 655–683.
- Armstrong, T.B., Kolesár, M., 2021. Finite-sample optimal estimation and inference on average treatment effects under unconfoundedness. *Econometrica* 89 (3), 1141–1177.
- Bagger, J., Fontaine, F., Postel-Vinay, F., Robin, J.-M., 2014. Tenure, experience, human capital, and wages: a tractable equilibrium search model of wage dynamics. *American Economic Review* 104 (6), 1551–1596.
- Bagger, J., Lentz, R., 2019. An empirical model of wage dispersion with sorting. *The Review of Economic Studies* 86 (1), 153–190.
- Bassier, I., Dube, A., Naidu, S., 2022. Monopsony in movers: the elasticity of labor supply to firm wage policies. *Journal of Human Resources* 57 (S), S50–s86.
- Bellmann, L., Lochner, B., Seth, S., Wolter, S., 2020. Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for....].
- Bertheau, A., Acabbi, E.M., Barceló, C., Gulyas, A., Lombardi, S., Saggio, R., 2022. The unequal cost of job loss across countries. Tech. rep., IZA DP No. 15033.
- Bertheau, A., Acabbi, E.M., Barceló, C., Gulyas, A., Lombardi, S., Saggio, R., 2023. The unequal consequences of job loss across countries. *American Economic Review: Insights* 5 (3), 393–408.

- Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. *The Annals of Statistics* 1196–1217.
- Bloom, N., Guvenen, F., Smith, B.S., Song, J., von Wachter, T., 2018. The disappearing large-firm wage premium. In: AEA Papers and Proceedings, volume 108, 317–322.
- Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., Setzler, B., 2023. How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics* 41 (2), 291–322.
- Bonhomme, S., Lamadon, T., Manresa, E., 2019. A distributional framework for matched employer employee data. *Econometrica* 87 (3), 699–739.
- Boza, I., Reizer, B., 2024. The role of flexible wage components in gender wage difference.
- Bozzo, E., 2013. The moore–penrose inverse of the normalized graph laplacian. *Linear Algebra and Its Applications* 439 (10), 3038–3043.
- Brown, C., Hamilton, J., Medoff, J.L., 1990. Employers Large and Small. Harvard University Press.
- Brown, C., Medoff, J., 1989. The employer size-wage effect. *Journal of Political Economy* 97 (5), 1027–1059.
- Bruns, B., 2019. Changes in workplace heterogeneity and how they widen the gender wage gap. *American Economic Journal: Applied Economics* 11 (2), 74–113.
- Burdett, K., Mortensen, D.T., 1998. Wage differentials, employer size, and unemployment. *International Economic Review* 257–273.
- Cahuc, P., Postel-Vinay, F., Robin, J.-M., 2006. Wage bargaining with on-the-job search: Theory and evidence. *Econometrica* 74 (2), 323–364.
- Caldwell, S., Haeggele, I., Heining, J., 2024a. Bargaining in the labor market.
- Caldwell, S., Haeggele, I., Heining, J., 2024b. Firm pay and worker search.
- Caldwell, S., Harmon, N., 2019. Outside options, bargaining, and wages: Evidence from coworker networks. Unpublished manuscript, Univ. Copenhagen.
- Cappelli, P., Chauvin, K., 1991. An interplant test of the efficiency wage hypothesis. *The Quarterly Journal of Economics* 106 (3), 769–787.
- Card, D., Cardoso, A.R., Heining, J., Kline, P., 2018. Firms and labor market inequality: evidence and some theory. *Journal of Labor Economics* 36 (S1), S13–S70.
- Card, D., Cardoso, A.R., Kline, P., 2016. Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics* 131 (2), 633–686.
- Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of west German wage inequality. *The Quarterly Journal of Economics* 128 (3), 967–1015.
- Card, D., Rothstein, J., Yi, M., 2023. Industry wage differentials: A firm-based approach. Tech. rep., National Bureau of Economic Research.
- Card, D., Rothstein, J., Yi, M., 2024. Industry wage differentials: a firm-based approach. *Journal of Labor Economics* 42 (S1), S11–S59.
- Cattaneo, M.D., Jansson, M., Newey, W.K., 2018. Inference in linear regression models with many covariates and heteroscedasticity. *Journal of the American Statistical Association* 113 (523), 1350–1361.
- Chamberlain, G., 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18 (1), 5–46.
- Chamberlain, G., 1984. Panel data. *Handbook of Econometrics* 2, 1247–1318.
- Chen, M., Fernández-Val, I., Weidner, M., 2021. Nonlinear factor models for network and panel data. *Journal of Econometrics* 220 (2), 296–324.

- Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., Yagan, D., 2011. How does your kindergarten classroom affect your earnings? Evidence from project star. *The Quarterly Journal of Economics* 126 (4), 1593–1660.
- Chetty, R., Hendren, N., 2018. The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics* 133 (3), 1107–1162.
- Condorcet, M.D., 1785. *Essay on the Application of Analysis to the Probability of Majority Decisions*. Imprimerie Royale, Paris.
- Coudin, E., Maillard, S., Tô, M., 2018. Family, firms and the gender wage gap in france. Tech. rep., IFS Working Papers.
- Crane, L.D., Hyatt, H.R., Murray, S.M., 2023. Cyclical labor market sorting. *Journal of Econometrics* 233 (2), 524–543.
- Crump, R.K., Hotz, V.J., Imbens, G.W., Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. *Biometrika* 96 (1), 187–199.
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (methodological)* 39 (1), 1–22.
- Di Addario, S., Kline, P., Saggio, R., Sølvsten, M., 2023. It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages. *Journal of Econometrics* 233 (2), 340–374.
- Donoho, D., 2024. Data science at the singularity. *Harvard Data Science Review* 6, 1.
- Dostie, B., Li, J., Card, D., Parent, D., 2023. Employer policies and the immigrant-native earnings gap. *Journal of Econometrics* 233 (2), 544–567.
- Drenik, A., Jäger, S., Plotkin, P., Schoefer, B., 2023. Paying outsourced labor: direct evidence from linked temp agency-worker-client data. *Review of Economics and Statistics* 105 (1), 206–216.
- Dustmann, C., Fitzenberger, B., Schönberg, U., Spitz-Oener, A., 2014. From sick man of europe to economic superstar: Germany's resurgent economy. *Journal of Economic Perspectives* 28 (1), 167–188.
- Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., VomBerge, P., 2022. Reallocation effects of the minimum wage. *The Quarterly Journal of Economics* 137 (1), 267–328.
- Eeckhout, J., Kircher, P., 2011. Identifying sorting—in theory. *The Review of Economic Studies* 78 (3), 872–906.
- ElKaroui, N., Purdom, E., 2018. Can we trust the bootstrap in high-dimensions? The case of linear models. *The Journal of Machine Learning Research* 19 (1), 170–235.
- Engbom, N., Moser, C., Sauermann, J., 2023. Firm pay dynamics. *Journal of Econometrics* 233 (2), 396–423.
- Faberman, R.J., Mueller, A.I., Şahin, A., Topa, G., 2022. Job search behavior among the employed and non-employed. *Econometrica* 90 (4), 1743–1779.
- Fenizia, A., 2022. Managers and productivity in the public sector. *Econometrica* 90 (3), 1063–1084.
- Finkelstein, A., Gentzkow, M., Williams, H., 2016. Sources of geographic variation in health care: evidence from patient migration. *The Quarterly Journal of Economics* 131 (4), 1681–1726.
- Firth, D., De Menezes, R.X., 2004. Quasi-variances. *Biometrika* 91 (1), 65–80.
- Flinn, C., Mullins, J., 2017. Firms' choices of wage-setting protocols in the presence of minimum wages. Technical Reports, Discussion paper. New York University.
- Flinn, C., Mullins, J., 2021. Firms' choices of wage-setting protocols. Technical Reports, Discussion paper. New York University.
- Forgy, E.W., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics* 21, 768–769.

- Gerard, F., Lagos, L., Severnini, E., Card, D., 2021. Assortative matching or exclusionary hiring? The impact of employment and pay policies on racial wage differences in Brazil. *American Economic Review* 111 (10), 3418–3457.
- Gibbons, R., Katz, L., 1992. Does unmeasured ability explain inter-industry wage differentials? *The Review of Economic Studies* 59 (3), 515–535.
- Gibbons, R., Katz, L.F., Lemieux, T., Parent, D., 2005. Comparative advantage, learning, and sectoral wage determination. *Journal of Labor Economics* 23 (4), 681–724.
- Goldschmidt, D., Schmieder, J.F., 2017. The rise of domestic outsourcing and the evolution of the German wage structure. *The Quarterly Journal of Economics* qjx008.
- Goldsmith-Pinkham, P., Hull, P., Kolesár, M., 2022. Contamination bias in linear regressions. Tech. rep., National Bureau of Economic Research.
- Gottfries, A., Jarosch, G., 2023. Dynamic monopsony with large firms and an application to non-competes.
- Groshen, E.L., 1991. Sources of intra-industry wage dispersion: how much do employers matter? *The Quarterly Journal of Economics* 106 (3), 869–884.
- Guo, J., 2023. The response of wages to rejected offers.
- Haanwinckel, D., 2023. Supply, demand, institutions, and firms: a theory of labor market sorting and the wage distribution. Technical Reports, National Bureau of Economic Research.
- Haltiwanger, J., Hyatt, H.R., Spletzer, J.R., 2024. Rising top, falling bottom: industries and rising wage inequality. *American Economic Review* 114 (10), 3250–3283.
- Hanushek, E.A., 1974. Efficient estimators for regressing regression coefficients. *The American Statistician* 28 (2), 66–67.
- Hoaglin, D.C., Welsch, R.E., 1978. The hat matrix in regression and anova. *The American Statistician* 32 (1), 17–22.
- Holzer, H.J., Katz, L.F., Krueger, A.B., 1991. Job queues and wages. *Quarterly Journal of Economics* 106 (3), 739–768.
- Hsieh, C.-T., Klenow, P.J., 2009. Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics* 124 (4), 1403–1448.
- Huitfeldt, I., Kostøl, A.R., Nimezik, J., Weber, A., 2023. Internal labor markets: a worker flow approach. *Journal of Econometrics* 233 (2), 661–688.
- Imbens, G.W., 2010. Better late than nothing: some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic literature* 48 (2), 399–423.
- Jäger, S., Roth, C., Roussille, N., Schoefer, B., 2024. Worker beliefs about outside options. *The Quarterly Journal of Economics* qjae001.
- Jarosch, G., Nimezik, J.S., Sorkin, I., 2024. Granular search, market structure, and wages. *Review of Economic Studies* rdae004.
- Jochmans, K., Weidner, M., 2019. Fixed-effect regressions on network data. *Econometrica* 87 (5), 1543–1560.
- Katz, L.F., Summers, L.H., Hall, R.E., Schultze, C.L., Topel, R.H., 1989. Industry rents: evidence and implications. *Brookings Papers on Economic Activity. Microeconomics* 1989, 209–290.
- Kline, P., Saggio, R., Sølvsten, M., 2020. Leave-out estimation of variance components. *Econometrica* 88 (5), 1859–1898.
- Koutis, I., Miller, G.L., Tolliver, D., 2011. Combinatorial preconditioners and multilevel solvers for problems in computer vision and image processing. *Computer Vision and Image Understanding* 115 (12), 1638–1646.
- Krueger, A.B., Summers, L.H., 1988. Efficiency wages and the inter-industry wage structure. *Econometrica: Journal of the Econometric Society* 259–293.

- Lachowska, M., Mas, A., Saggio, R., Woodbury, S.A., 2023. Do firm effects drift? Evidence from Washington administrative data. *Journal of Econometrics* 233 (2), 375–395.
- Lachowska, M., Mas, A., Woodbury, S.A., 2020. Sources of displaced workers' long-term earnings losses. *American Economic Review* 110 (10), 3231–3266.
- Lagos, L., 2019. Labor market institutions and the composition of firm compensation: evidence from Brazilian collective bargaining.
- Lamadon, T., Mogstad, M., Setzler, B., 2022. Imperfect competition, compensating differentials, and rent sharing in the US labor market. *American Economic Review* 112 (1), 169–212.
- Lehmann, T., 2023. Non-wage job values and implications for inequality. Available at SSRN 4373816.
- Lei, L., Ross, B., 2023. Estimating counterfactual matrix means with short panel data. arXiv preprint arXiv:2312.07520.
- Lentz, R., Piyapromdee, S., Robin, J.-M., 2022. The anatomy of sorting—evidence from danish data.
- Li, J., Dostie, B., Simard-Duplain, G., 2023. Firm pay policies and the gender earnings gap: the mediating role of marital and family status. *ILR Review* 76 (1), 160–188.
- Lloyd, S., 1982. Least squares quantization in pcm. *IEEE Transactions on Information Theory* 28 (2), 129–137.
- MacKinnon, J.G., White, H., 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29 (3), 305–325.
- MacCurdy, T.E., 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18 (1), 83–114.
- Maestas, N., Mullen, K.J., Powell, D., Von Wachter, T., Wenger, J.B., 2023. The value of working conditions in the United States and implications for the structure of wages. *American Economic Review* 113 (7), 2007–2047.
- Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics* 255–285.
- Mas, A., 2024. Compensating differentials. *Handbook of Labor Economics*.
- McLaughlin, K.J., 1991. A theory of quits and layoffs with efficient turnover. *Journal of Political Economy* 99 (1), 1–29.
- Meghir, C., Pistaferri, L., 2004. Income variance dynamics and heterogeneity. *Econometrica* 72 (1), 1–32.
- Murphy, K.M., Topel, R.H., 1990. Efficiency wages reconsidered: theory and evidence. *Advances in the Theory and Measurement of Unemployment*. Springer, pp. 204–240.
- Newey, W.K., Robins, J.R., 2018. Cross-fitting and fast remainder rates for semiparametric estimation. arXiv preprint arXiv:1801.09138.
- Nimczik, J.S., 2017. Job mobility networks and endogenous labor markets.
- Oi, W.Y., Idson, T.L., 1999. Firm size and wages. *Handbook of Labor Economics* 3, 2165–2214.
- Page, L., Brin, S., Motwani, R., Winograd, T., et al., 1999. The pagerank citation ranking: bringing order to the web.
- Postel-Vinay, F., Robin, J.-M., 2002a. The distribution of earnings in an equilibrium search model with state-dependent offers and counteroffers. *International Economic Review* 43 (4), 989–1016.
- Postel-Vinay, F., Robin, J.-M., 2002b. Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica* 70 (6), 2295–2350.
- Postel-Vinay, F., Robin, J.-M., 2004. To match or not to match?: Optimal wage policy with endogenous worker search intensity. *Review of Economic Dynamics* 7 (2), 297–330.
- Postel-Vinay, F., Turon, H., 2014. The impact of firing restrictions on labour market equilibrium in the presence of on-the-job search. *The Economic Journal* 124 (575), 31–61.

- Rambachan, A., Roth, J., 2023. A more credible approach to parallel trends. *Review of Economic Studies* 90 (5), 2555–2591.
- Roussille, N., Scuderi, B., 2023. Bidding for talent: a test of conduct in a high-wage labor market.
- Roy, A.D., 1951. Some thoughts on the distribution of earnings. *Oxford Economic Papers* 3 (2), 135–146.
- Schmieder, J.F., VonWachter, T., Heining, J., 2023. The costs of job displacement over the business cycle and its sources: evidence from Germany. *American Economic Review* 113 (5), 1208–1254.
- Shapiro, C., Stiglitz, J.E., 1984. Equilibrium unemployment as a worker discipline device. *The American Economic Review* 74 (3), 433–444.
- Slichter, S.H., 1950. Notes on the structure of wages. *The Review of Economics and Statistics* 32 (1), 80–91.
- Sockin, J., 2022. Show me the amenity: are higher-paying firms better all around?
- Song, J., Price, D.J., Guvenen, F., Bloom, N., Von Wachter, T., 2019. Firming up inequality. *The Quarterly Journal of Economics* 134 (1), 1–50.
- Sorkin, I., 2018. Ranking firms using revealed preference. *The Quarterly Journal of Economics* 133 (3), 1331–1393.
- Sorkin, I., Wallskog, M., 2023. The slow diffusion of earnings inequality. *Journal of Labor Economics* 41 (S1), S95–S127.
- Spielman, D., 2019. Spectral and algebraic graph theory.
- Stigler, G.J., 1962. Information in the labor market. *Journal of Political Economy* 70 (5, Part2), 94–105.
- Vafa, K., Palikot, E., Du, T., Kanodia, A., Athey, S., Blei, D.M., 2022. Career: a foundation model for labor sequence data. arXiv preprint arXiv:2202.08370.
- Walters, C.R., 2024. Empirical bayes methods in labor economics. *Handbook of Labor Economics*.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society* 817–838.
- Yeh, C., Macaluso, C., Hershbein, B., 2022. Monopsony in the US labor market. *American Economic Review* 112 (7), 2099–2138.
- Young, P., 1995. Optimal voting rules. *Journal of Economic Perspectives* 9 (1), 51–64.

This page intentionally left blank

Chapter 3

Empirical Bayes methods in labor economics[☆]

Christopher Walters*

UC Berkeley, NBER, and Amazon

*Corresponding author. e-mail address: crwalters@econ.berkeley.edu

Chapter Outline

1 Introduction	184	3.2 Non-parametric priors and posteriors	221
2 Empirical Bayes basics	186	3.3 Partial identification	224
2.1 An empirical Bayes recipe	186	3.4 EB for multiple testing: large-scale inference	227
2.2 Gains from shrinkage	192	3.5 Ranking problems	230
2.3 Practical shrinkage issues	196	3.6 Compound decisions and shrinkage strategies	233
2.4 Generalizations of linear shrinkage	200	3.7 Non-parametric EB application: firm-level labor market discrimination	235
2.5 EB decision rules	204		
2.6 Precision-dependence	206		
2.7 Connections to machine learning	211		
2.8 Linear shrinkage application: school value-added in Boston	213		
3 Non-parametric empirical Bayes	216	4 Conclusion	253
3.1 Bias-corrected variance estimation	216	References	255

☆ The material builds on the 2022 NBER Methods Lecture “Empirical Bayes Methods: Theory and Application.” I thank Jiaying Gu for her collaboration on that lecture as well as course participants for engaging comments and questions. I also thank participants in several related workshops including the AEA Continuing Education Program, Bonn/Mannheim CRC Summer School, Northwestern Causal Inference Workshop, Empirical Bayes Mixtape Session, and UCLA CCPR seminar. I am grateful to Thomas Lemieux and Christian Dustmann for their work organizing the HOLE Volume, to participants at the 2023 HOLE Volume Conference, and to the Rockwool Foundation Berlin for providing funding for the conference. The Massachusetts Department of Elementary and Secondary Education generously provided the data used for school value-added analysis. Finally, I thank Joshua Angrist, Peter Hull, Patrick Kline, Parag Pathak, and Evan Rose for collaborations and discussions that were essential to the development of this chapter.

1 Introduction

Labor economists in a variety of research areas increasingly drill down to study large numbers of finely-grained, unit-specific parameters. Large-scale administrative data sets and new research designs have allowed researchers to sharpen the focus of scientific inquiry into economic impacts of specific institutions and individuals. In the economics of education, for example, a classic topic is the return to education, traditionally defined as the causal effect of an extra year of schooling on earnings (Card, 1999). This question has evolved into studies of the effects of attending more versus less selective colleges (Dale and Krueger, 2002, 2014) and impacts of specific colleges on labor market success (Mountjoy and Hickman, 2021; Chetty et al., 2023). Building on Krueger and Summers' (1988) seminal study of industry-specific wage premia, a large recent literature looks at the role of individual firms in wage determination (Abowd et al., 1999; Card et al., 2013, 2018; Bonhomme et al., 2023). More broadly, many areas of applied microeconomics have seen the rise of *value-added* studies considering causal effects of individual units such as neighborhoods, teachers, schools, managers, doctors, health insurance plans, hospitals, nursing homes, police officers, and judges (Chetty and Hendren, 2018; Chetty et al., 2018, 2014a; Angrist et al., 2017, 2024a; Fenizia, 2022; Chan et al., 2022; Abaluck et al., 2021; Einav et al., 2022; Goncalves and Mello, 2021; Frandsen et al., 2023).

Empirical Bayes (EB) methods provide a powerful suite of econometric tools for settings with large numbers of unit-specific parameters. The EB framework, pioneered by Robbins (1951, 1956, 1964), is designed for contexts involving multiple parallel estimation problems for many similar units. EB leverages this common structure by pooling information on all units to estimate a distribution of parameters in the population being studied. This estimated distribution is then used as an empirical prior to construct posterior predictions for each individual unit. On average, the resulting EB estimators and decision rules often perform better than approaches that consider each unit separately.

An EB approach can be used to accomplish several key objectives in value-added analyses. First, the estimated EB prior characterizes the *distribution* of value-added across units. This condenses large sets of value-added estimates into interpretable summaries of heterogeneity, such as the variance of value-added. Second, by “borrowing strength” from other similar units via the estimated prior (Morris, 1983), EB improves estimates of individual value-added parameters. Third, EB methods are useful for making decisions. EB is intimately linked to a *compound decision problem* in which an analyst faces repeated decisions across many units and seeks to minimize an aggregate loss function. A clear view of the loss function allows a researcher to convey the information about each unit’s value-added that is most relevant for the economic objective at hand.

This chapter offers an overview of empirical Bayes methods in labor economics. Many excellent surveys of EB and related methods are already available (see, e.g., [Efron, 2012](#); [Bonhomme and Denis, 2024](#); [Koenker and Gu, 2024](#)). I do not aim to improve upon the technical elements of these treatments or break new theoretical ground on the properties of EB. My goal is to provide an accessible toolkit for labor economists using EB methods in value-added studies. Along the way I will touch on practical issues that arise in EB analyses of microeconomic data, offer concrete examples and guidance on EB implementation, and make connections between EB and other methods that may be more familiar to labor economists. Applications to school value-added in Boston and employer-level discrimination in the US labor market will illustrate the EB toolkit in action.

The remainder of the chapter is organized in two broad sections. [Section 2](#) provides a basic introduction to the empirical Bayes approach. I start with a simple three-step EB recipe: (1) estimate a parameter and associated standard error for each unit; (2) pool the estimates and standard errors to estimate the prior distribution (a procedure known as *deconvolution*); and (3) use the estimated prior to generate posterior predictions for each unit (also known as *empirical Bayes shrinkage*). Applying this recipe with a normal prior distribution gives rise to *linear shrinkage* estimators of the sort considered by [James and Stein \(1961\)](#). While linear shrinkage is nominally based on a normality assumption, the James/Stein Theorem establishes that this method improves aggregate mean squared error (MSE) even if the normal model is wrong. Thus, we can view parametric empirical Bayes as a device for motivating procedures that perform well for a broader class of data generating processes.

After covering the basics of linear shrinkage I consider several variations on the theme. These extensions include the use of shrinkage in a regression context, posteriors that incorporate observed unit characteristics, and scenarios with multiple parameters per unit or multiple research designs for estimating the same parameter. I also discuss EB decision rules for objectives other than mean squared error, outline approaches to incorporating dependence between effect sizes and standard errors, and make connections between EB shrinkage and commonly-used machine learning algorithms. [Section 2](#) ends with an EB analysis of middle school value-added in Boston based on data from [Angrist et al. \(2017\)](#). [Section 3](#) introduces non-parametric empirical Bayes methods that relax the normality and independence assumptions maintained earlier in the chapter. Here I link the EB framework with recent approaches to bias-corrected variance component estimation (e.g., [Kline et al., 2020](#)), which may be used to estimate the variance of value-added under weak assumptions. I then consider non-parametric methods for estimating the full prior distribution, including non-parametric maximum likelihood (NPML; [Robbins, 1950](#); [Kiefer and Wolfowitz, 1956](#)) and flexible smooth deconvolution estimators ([Efron, 2016](#)). These non-parametric prior estimation procedures give rise to

corresponding non-parametric EB posteriors, which generalize James/Stein-style shrinkage to account for features of the value-added distribution beyond the mean and variance.

The second half of Section 3 covers variants of non-parametric EB that are relevant for empirical practice. These include cases with partial identification of priors and posteriors, EB approaches to multiple testing, and ranking problems. I connect ideas from throughout the chapter in a non-parametric version of the compound decision framework, and contrast EB decision rules motivated by different economic objectives. Revisiting a labor market correspondence experiment studied by Kline, Rose and Walters (2022, 2024), I apply non-parametric EB methods to flexibly estimate distributions of race and gender discrimination across large US employers.¹ The chapter concludes in Section 4 with thoughts on directions for future applications of empirical Bayes methods in labor economics.

2 Empirical Bayes basics

2.1 An empirical Bayes recipe

Consider a hierarchical data structure with individuals nested within groups. We observe data Y_i on N individuals indexed by i , each associated with one of J groups (I refer to the groups interchangeably as *units*). Let $D_i \in \{1, \dots, J\}$ denote the group for individual i . A group-specific parameter θ_j determines the distribution of Y_i for individuals with $D_i = j$. For example, Y_i might represent test scores for students nested within schools, with θ_j the causal effect of school j on student achievement; or Y_i might measure log earnings for workers nested within firms, with θ_j a pay premium associated with working at firm j . I next outline a three-step empirical Bayes recipe for estimating the unit-specific θ_j parameters, characterizing the distribution of these parameters across units, and using this distributional information to refine the estimate for each unit.

Step 1: Estimation

The first step of an EB analysis uses the data for group j to form an estimate $\hat{\theta}_j$ of the parameter θ_j along with a corresponding standard error s_j . In much of what follows I will assume these estimates are unbiased, normally distributed, and mutually independent, with sampling variances equal to their squared standard errors:

$$\hat{\theta}_j | \theta_j, s_j \sim N(\theta_j, s_j^2). \quad (1)$$

The conditioning on (θ_j, s_j) in this expression is to emphasize that at this stage these quantities are treated as fixed parameters rather than random variables. Normality of $\hat{\theta}_j$ may be justified either by a parametric model for the

¹ Software for implementing these methods is available in this chapter's replication package.

microdata Y_i or by an asymptotic approximation with a growing number of individuals in each group. In the latter case a central limit theorem is typically invoked to argue that $\hat{\theta}_j$ is asymptotically normal with limiting variance s_j^2 . I later discuss scenarios where such asymptotic approximations break down.

Two examples of the estimation step will help fix ideas as I develop the EB recipe.

Example 1: Normal means

Suppose the data Y_i are normally distributed with a group-specific mean and variance:

$$Y_i | D_i = j, \theta_j, \sigma_j \sim N(\theta_j, \sigma_j^2). \quad (2)$$

The natural estimator of θ_j in this case is the sample mean $\hat{\theta}_j = n_j^{-1} \sum_{i=1}^N D_{ij} Y_i$, where $D_{ij} = 1\{D_i = j\}$ indicates that observation i is in group j and $n_j = \sum_{i=1}^N D_{ij}$ is the number of observations in this group. The parametric model in (2) implies $\hat{\theta}_j$ is normally distributed with mean θ_j and variance $s_j^2 = \sigma_j^2/n_j$.

Example 2: School value-added

Consider a causal model of school effectiveness in which potential academic achievement for student i if she attends school j is given by

$$Y_i(j) = \theta_j + a_i. \quad (3)$$

Here θ_j represents the causal contribution of school j to student achievement (school j 's *value-added*), while a_i captures all student-specific factors that influence achievement including family background, earlier educational investments, and innate ability. I normalize $E[a_i] = 0$ so that $\theta_j = E[Y_i(j)]$ equals the population mean potential outcome for school j . The additive structure in equation (3) implies school value-added is constant across students – for any student i , $\theta_j - \theta_k$ is the causal effect of attending school j rather than k . The observed outcome Y_i is the potential outcome associated with the school that i attends: $Y_i = \sum_{j=1}^J D_{ij} Y_i(j)$.

School enrollment is not randomly assigned, so uncontrolled comparisons of test scores across schools are likely to be contaminated by differences in student ability. This motivates an empirical strategy that adjusts for observed characteristics such as demographics and lagged achievement. Let X_i represent a vector of these control variables, de-meaned so that $E[X_i] = 0$. Write the linear projection of a_i on X_i as:

$$a_i = X_i' \gamma + \epsilon_i, \quad (4)$$

where $\gamma = E[X_i X_i']^{-1} E[X_i a_i]$, and $E[\epsilon_i] = E[X_i \epsilon_i] = 0$ by definition of γ . I can then write the observed outcome for student i as

$$Y_i = \sum_{j=1}^J \theta_j D_{ij} + X_i' \gamma + \epsilon_i. \quad (5)$$

The residual ϵ_i is uncorrelated with X_i by construction but need not be uncorrelated with the school indicators D_{ij} . Studies of such *value-added models* (VAMs) typically proceed under the assumption that $E[D_{ij}\epsilon_i] = 0 \forall j$. This is a selection-on-observables restriction requiring additive controls for X_i to eliminate any relationship between potential outcomes and school attendance. Under this assumption, Ordinary Least Squares (OLS) estimation of (5) using a random sample of students recovers unbiased estimates $\hat{\theta}_j$ of the causal value-added parameters θ_j . The squared standard errors s_j^2 are the first J diagonal elements of the asymptotic covariance matrix of the OLS coefficient vector $(\hat{\theta}_1, \dots, \hat{\theta}_J, \gamma')'$, which may be estimated using the standard [White \(1980\)](#) heteroskedasticity-robust variance matrix or another appropriate variance estimator. With a sufficient number of students per school we can treat $\hat{\theta}_j$ as normally distributed with variance approximately equal to s_j^2 .

2.1.1 Step 2: Deconvolution

The second step of the EB recipe is predicated on a model of the group-level parameters as random draws from a probability distribution. Suppose the θ_j 's are generated by independent draws from a common cumulative distribution function G :

$$\theta_j \sim G, j \in \{1, \dots, J\}. \quad (6)$$

The *mixing distribution* G is central to the empirical Bayes approach. Equations (1) and (6) imply the estimates $\hat{\theta}_j$ are a mixture of draws from G and normally-distributed sampling error. In other words, the marginal distribution of $\hat{\theta}_j$ is a *convolution* of G and normal noise with variance s_j^2 .

Though G will play the role of a Bayesian prior in much of what follows, it is worth highlighting that this distribution represents an objective description of the variation in parameters θ_j rather than subjective beliefs about their likely values. In the school value-added context (Example 2), we might wonder whether there are substantial differences between schools' contributions to student learning. The variance of the mixing distribution, given by $\sigma_\theta^2 = \int (\theta - \mu_\theta)^2 dG(\theta)$ with $\mu_\theta = \int \theta dG(\theta)$, provides a quantitative answer to this question. Likewise, the gap in effectiveness between the most- and least-effective schools is a feature of G ; the difference in value-added between 90th and 10th-percentile schools is $G^{-1}(0.9) - G^{-1}(0.1)$. The second step of an EB analysis answers such questions empirically via *deconvolution*: extract an estimate \hat{G} of the mixing distribution G from the noisy group-specific estimates $\hat{\theta}_j$ and standard errors s_j , and use \hat{G} to quantify heterogeneity in parameters across groups.

While there is nothing subjective about G , a practical empiricist might nonetheless be uncomfortable with the random effects model in (6). On a conventional fixed effects view, the parameters θ_j are just a list of J unknown numbers, so it may seem unnatural to treat them as random draws from a probability distribution. In some cases model (6) is justified by a hierarchical

sampling process – perhaps the groups were drawn at random from a larger superpopulation, as in some multi-site randomized trials (see, e.g., [Walters, 2015](#)). In other cases, however, we observe all the relevant groups, so it is unattractive to appeal to such a superpopulation view. In a school value-added analysis for a particular district like Boston, where are the missing schools that might have otherwise been sampled?

An important theme of this chapter is that EB methods perform well even judged on purely frequentist terms, independent of such philosophical considerations. As discussed further below, estimators and decision rules incorporating distributional information based on an estimate of G can improve on approaches treating each $\hat{\theta}_j$ in isolation. These improvements are achieved even if the stylized model in (6) is wrong. We can therefore think of G as a device for motivating procedures with desirable frequentist properties, irrespective of whether the parameters are fixed or random. In what follows I will often make use of continuous models for G , which must be misspecified on a fixed effects view; for a committed frequentist, these models may be seen as useful approximations rather than literal descriptions of the data-generating process.

To illustrate the basic mechanics of EB, the remainder of this section will mostly focus on a parametric normal model for G :

$$\theta_j | s_j \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2). \quad (7)$$

Model (7) is written conditional on s_j , which implies the effect sizes θ_j are independent of the sampling variances s_j^2 across groups, a restriction I relax later. With this parametric model deconvolution amounts to estimating the two *hyperparameters* μ_θ and σ_θ . The “hyperparameter” label refers to the fact that μ_θ and σ_θ govern the distribution of lower-level parameters (the θ_j ’s). Simple estimators for these hyperparameters are given by:

$$\hat{\mu}_\theta = \frac{1}{J} \sum_{j=1}^J \hat{\theta}_j, \quad (8)$$

$$\hat{\sigma}_\theta^2 = \frac{1}{J} \sum_{j=1}^J [(\hat{\theta}_j - \hat{\mu}_\theta)^2 - s_j^2]. \quad (9)$$

The subtraction of s_j^2 in (9) is a *bias-correction* that accounts for excess variability of the $\hat{\theta}_j$ ’s due to statistical noise. We should expect some chance variation in the estimated $\hat{\theta}_j$ ’s even if all the latent θ_j ’s are equal, so the naive sample variance of $\hat{\theta}_j$ ’s is too large relative to the variance of the mixing distribution. Subtracting the average squared standard error removes this excess noise. A finding that $\hat{\sigma}_\theta^2 > 0$ indicates *overdispersion* in the $\hat{\theta}_j$ ’s: the observed estimates differ more than should be expected from sampling error, implying variation in the underlying θ_j ’s. With a growing number of groups J , the bias-corrected variance estimate $\hat{\sigma}_\theta^2$ is consistent for the

hyperparameter σ_θ^2 .² Section 3 considers unbiased estimation of the mixing variance in finite samples and non-parametric deconvolution methods for flexibly estimating the hyperparameters of a non-normal G .

2.1.2 Step 3: Shrinkage

The third and final step of an EB analysis uses the noisy $\hat{\theta}_j$'s together with the estimated mixing distribution \hat{G} to form posteriors for each θ_j . Specifically, we treat the deconvolved \hat{G} as a prior, then use Bayes' rule to perform an update based on $(\hat{\theta}_j, s_j)$ and generate a posterior distribution for θ_j . We are often interested in a particular feature of this posterior distribution such as the posterior mean. The use of an estimated \hat{G} as a prior when forming posterior predictions is commonly known as *empirical Bayes shrinkage*.

Suppose first that the mixing distribution G is known. Equations (1) and (6) along with Bayes' rule imply the cumulative distribution function for θ_j conditional on $(\hat{\theta}_j, s_j)$ evaluated at a point t is given by:

$$\Pr[\theta_j \leq t | \hat{\theta}_j, s_j] = \frac{\int_{-\infty}^t \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \theta}{s_j}\right) dG(\theta)}{\int_{-\infty}^{\infty} \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \theta}{s_j}\right) dG(\theta)} \equiv \mathcal{P}(t | \hat{\theta}_j, s_j; G). \quad (10)$$

I will sometimes refer to $\mathcal{P}(t | \hat{\theta}_j, s_j; G)$ as an *oracle* posterior distribution because it coincides with the posterior beliefs of an oracle who knows the mixing distribution G a priori. With the normal mixing distribution in model (7), this oracle posterior distribution is also normal:

$$\theta_j | \hat{\theta}_j, s_j \sim \mathcal{N}(\theta_j^*, V_j^*), \quad (11)$$

$$\theta_j^* = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_j^2} \right) \hat{\theta}_j + \left(\frac{s_j^2}{\sigma_\theta^2 + s_j^2} \right) \mu_\theta, \quad (12)$$

$$V_j^* = \frac{\sigma_\theta^2 s_j^2}{\sigma_\theta^2 + s_j^2}. \quad (13)$$

The posterior mean θ_j^* shrinks the noisy estimate $\hat{\theta}_j$ towards the prior mean μ_θ in proportion to its signal-to-noise ratio. When $\hat{\theta}_j$ is completely uninformative or the prior distribution is degenerate ($s_j^2 \rightarrow \infty$ or $\sigma_\theta^2 \rightarrow 0$) the posterior mean equals the prior mean μ_θ . As the precision of $\hat{\theta}_j$ or the variability of the prior grow large ($s_j^2 \rightarrow 0$ or $\sigma_\theta^2 \rightarrow \infty$) the posterior mean approaches the estimate $\hat{\theta}_j$. In between these extremes, θ_j^* is a convex weighted average of the unbiased estimate $\hat{\theta}_j$ and the prior mean μ_θ .

² Standard errors for the hyperparameter estimates in the normal/normal model can be calculated as $SE(\hat{\mu}_\theta) = J^{-1} \sqrt{\sum_j (\hat{\theta}_j - \hat{\mu}_\theta)^2}$ and $SE(\hat{\sigma}_\theta^2) = J^{-1} \sqrt{2 \sum_j (\hat{\sigma}_\theta^2 + s_j^2)^2}$.

An analyst that does not know G cannot calculate the oracle posterior formulas in [equations \(10\)–\(13\)](#). The empirical Bayes approach is to plug in an estimate \hat{G} of the mixing distribution, resulting in an empirically feasible posterior:

$$\mathcal{P}(t|\hat{\theta}_j, s_j; \hat{G}) = \frac{\int_{-\infty}^t \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \theta}{s_j}\right) d\hat{G}(\theta)}{\int_{-\infty}^{\infty} \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \theta}{s_j}\right) d\hat{G}(\theta)}, \quad (14)$$

where \hat{G} is the deconvolution estimate from step 2. With the normal model in [\(7\)](#), this means substituting the estimated hyperparameters $\hat{\mu}_{\theta}$ and $\hat{\sigma}_{\theta}^2$ from [equations \(8\)–\(9\)](#) into the posterior formulas in [equations \(12\)–\(13\)](#). This substitution yields an empirical Bayes posterior mean:

$$\hat{\theta}_j^* = \left(\frac{\hat{\sigma}_{\theta}^2}{\hat{\sigma}_{\theta}^2 + s_j^2} \right) \hat{\theta}_j + \left(\frac{s_j^2}{\hat{\sigma}_{\theta}^2 + s_j^2} \right) \hat{\mu}_{\theta}. \quad (15)$$

The EB posterior mean $\hat{\theta}_j^*$ shrinks the unbiased estimate $\hat{\theta}_j$ for group j using distributional information based on the ensemble of estimates for all J groups, which enters through the estimated hyperparameters $\hat{\mu}_{\theta}$ and $\hat{\sigma}_{\theta}^2$. [Morris \(1983\)](#) refers to this adjustment as “borrowing strength from the ensemble.” EB shrinkage refines the estimate for each individual unit by interpreting it in the context of results for a larger pool units that are similar in some relevant sense. [Efron \(2012\)](#) labels such refinements “learning from the experience of others.”

In the remainder of this chapter I will often refer to the estimator in [equation \(15\)](#) as a *linear shrinkage estimator* to distinguish it from EB posterior means derived from more elaborate non-normal models for G . While linear shrinkage is motivated by normality, the resulting predictions are likely to have good properties even when G is not normal. In particular, $\hat{\theta}_j^*$ coincides with the fitted value from an infeasible ordinary least squares regression of the unobserved θ_j on $\hat{\theta}_j$.³ By standard properties of OLS regression, we can then think of the linear shrinkage estimate as a best linear approximation to the true (possibly nonlinear) conditional mean of θ_j given $\hat{\theta}_j$. As I discuss next, linear approximations of this form turn out to improve upon the unbiased estimates $\hat{\theta}_j$ (in a particular sense) regardless of the form of G .

2.1.3 Recap: A three-step EB recipe

The standard progression of an EB analysis is summarized in the following three-step recipe:

- 1. Estimation:** Compute an estimate $\hat{\theta}_j$ and corresponding standard error s_j for each unit j .

³ When θ_j is independent of s_j we have $\hat{\theta}_j^* = \left(\frac{\text{Cov}(\theta_j, \hat{\theta}_j | s_j)}{\text{Var}(\hat{\theta}_j | s_j)} \right) \hat{\theta}_j + \left(1 - \frac{\text{Cov}(\theta_j, \hat{\theta}_j | s_j)}{\text{Var}(\hat{\theta}_j | s_j)} \right) \hat{\mu}_{\theta}$.

2. **Deconvolution:** Use the estimates and standard errors $\{\hat{\theta}_j, s_j\}_{j=1}^J$ to compute an estimate \hat{G} of the mixing distribution.
3. **Shrinkage:** Treating \hat{G} as a prior, update with $(\hat{\theta}_j, s_j)$ to form posterior predictions $\hat{\theta}_j^*$ for each unit.

The key outputs of the analysis are typically a summary of parameter heterogeneity based on \hat{G} from the deconvolution step along with posterior predictions for each unit generated in the shrinkage step.

2.2 Gains from shrinkage

2.2.1 MSE improvements in the normal/normal model

In what sense do the EB posterior means $\hat{\theta}_j^*$ improve upon the unbiased estimates $\hat{\theta}_j$? I first explore this question in the context of the normal/normal model defined by [equations \(1\)](#) and [\(7\)](#), focusing on a mean squared error (MSE) criterion.⁴ Assume the hyperparameters μ_θ and σ_θ^2 are known, and consider the MSE of the oracle posterior mean θ_j^* and the unbiased estimate $\hat{\theta}_j$ conditional on the unknown parameter θ_j and standard error s_j :

$$E[(\hat{\theta}_j - \theta_j)^2 | \theta_j, s_j] = s_j^2, \quad (16)$$

$$E[(\theta_j^* - \theta_j)^2 | \theta_j, s_j] = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_j^2} \right)^2 s_j^2 + \left(\frac{s_j^2}{\sigma_\theta^2 + s_j^2} \right)^2 (\theta_j - \mu_\theta)^2. \quad (17)$$

Since $\hat{\theta}_j$ is unbiased, its MSE is given by its sampling variance s_j^2 . The posterior mean θ_j^* shrinks $\hat{\theta}_j$ toward a constant μ_θ , thereby reducing variance in exchange for an increase in bias. This results in a squared bias term in the conditional MSE formula, reflected in the second term in [\(17\)](#).

These expressions show that if an analyst is only interested in one unit (say θ_1), it is not clear which of the two estimators is better. The linear shrinkage posterior is less variable than the unbiased estimate, but for any particular unit this variance reduction may be outweighed by increased bias. Moreover, [equation \(17\)](#) shows that the conditional MSE of θ_j^* is not uniform in the true parameter θ_j . The bias introduced by shrinkage is worse for units that are more atypical in the sense of having θ_j 's farther from the mean μ_θ . This bias can be arbitrarily large if the mixing distribution has unbounded support. An analyst interested in a specific unit and concerned about worst-case MSE might reasonably prefer the unbiased estimate $\hat{\theta}_j$.

Next, consider an analyst who cares about reporting estimates with low MSE for many units simultaneously. This analyst expects to study many units drawn from G and wants to do well on average across all of them. We can

⁴ See [Angrist et al. \(2023\)](#) for related discussion in the context of school value-added models.

evaluate the performance of $\hat{\theta}_j$ and θ_j^* for this purpose by integrating conditional MSE over the mixing distribution:

$$E[(\hat{\theta}_j - \theta_j)^2 | s_j] = \int E[(\hat{\theta}_j - \theta_j)^2 | \theta_j = \theta, s_j] dG(\theta) = s_j^2, \quad (18)$$

$$E[(\theta_j^* - \theta_j)^2 | s_j] = \int E[(\theta_j^* - \theta_j)^2 | \theta_j = \theta, s_j] dG(\theta) = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_j^2} \right) s_j^2. \quad (19)$$

This integration averages the squared bias of θ_j^* over values of θ_j while leaving MSE for $\hat{\theta}_j$ unchanged.

The signal-to-noise ratio $\sigma_\theta^2 / (\sigma_\theta^2 + s_j^2)$ is less than one, so equations (18) and (19) establish that unconditional MSE for the shrinkage estimator is below that of the unbiased estimator. While shrinkage increases conditional MSE for atypical values of θ_j far from the mean, it improves MSE for values close to the mean, and on average over repeated draws from G the θ_j 's cannot all be atypical. The reduction in variance therefore outweighs the increased conditional bias on average, reducing overall MSE. In fact, since θ_j^* is the conditional mean of θ_j given $(\hat{\theta}_j, s_j)$, it must have lowest average MSE of all functions of $(\hat{\theta}_j, s_j)$ under the normal/normal model. As long as the hyperparameter estimates $\hat{\mu}_\theta$ and $\hat{\sigma}_\theta^2$ are sufficiently precise we should expect the EB posterior mean $\hat{\theta}_j^*$ to inherit the properties of the oracle posterior mean θ_j^* and improve average MSE relative to $\hat{\theta}_j$.

2.2.2 The James/Stein theorem

I next show that the gains from EB shrinkage apply more generally outside the normal random effects setup. This is a classic result due to James and Stein (Stein, 1956; James and Stein, 1961), referred to here as the *James/Stein Theorem*. Suppose each $\hat{\theta}_j$ is unbiased for the corresponding θ_j and normally distributed with common sampling variance of $s_j^2 = s^2 \forall j$:

$$\hat{\theta}_j | \theta_j \sim N(\theta_j, s^2). \quad (20)$$

We are interested in finding an estimator with low total MSE summed over all J units, treating the θ_j 's as fixed but unknown parameters. For any estimator δ_j , this objective is given by:

$$MSE_\delta = \sum_{j=1}^J E[(\delta_j - \theta_j)^2 | \theta_j]. \quad (21)$$

By equation (16), total MSE of the unbiased estimates $\hat{\theta}_j$ is equal to Js^2 :

$$MSE_{\hat{\theta}} = \sum_{j=1}^J E[(\hat{\theta}_j - \theta_j)^2 | \theta_j] = Js^2. \quad (22)$$

Now consider an alternative estimator that shrinks each $\hat{\theta}_j$ toward a constant μ as follows:

$$\hat{\theta}_j^{JS} = \left(1 - \frac{(J-2)s^2}{\sum_{k=1}^J (\hat{\theta}_k - \mu)^2}\right)\hat{\theta}_j + \left(\frac{(J-2)s^2}{\sum_{k=1}^J (\hat{\theta}_k - \mu)^2}\right)\mu. \quad (23)$$

This estimator turns out to improve MSE relative to the unbiased $\hat{\theta}_j$'s whenever the number of units J is at least 3. Assuming $J \geq 3$, we have

$$\begin{aligned} MSE_{\hat{\theta}}^{JS} &= \sum_{j=1}^J E[(\hat{\theta}_j^{JS} - \theta_j)^2 | \theta_j] \\ &\leq Js^2 - \frac{(J-2)^2 s^4}{(J-2)s^2 + \sum_{j=1}^J (\theta_j - \mu)^2} \\ &< Js^2 \\ &= MSE_{\hat{\theta}}. \end{aligned} \quad (24)$$

See Chapter 1 of [Efron \(2012\)](#) for a simple proof.

This result shows that the unbiased estimator $\hat{\theta}_j$ is inadmissible under squared error loss since it is dominated by $\hat{\theta}_j^{JS}$. This may seem counterintuitive since $\hat{\theta}_j$ is the maximum likelihood estimator of θ_j under model (20). Indeed, as discussed in [Efron and Morris \(1975\)](#), the James/Stein result was initially met with surprise and encountered resistance even among statisticians.

Where does the linear shrinkage rule in [equation \(23\)](#) come from? It should be clear based on the preceding discussion that this estimator is the output of an empirical Bayes procedure based on a normal mixing distribution. Suppose we adopt the prior that $\theta_j \sim N(\mu, \sigma^2)$. [Equation \(12\)](#) implies that if the hyperparameters μ and σ^2 are known, the posterior mean for θ_j is given by $\theta_j^* = (\sigma^2/[s^2 + \sigma^2])\hat{\theta}_j + (s^2/[s^2 + \sigma^2])\mu$. If μ is known but σ^2 is not, the shrinkage factor $s^2/(s^2 + \sigma^2)$ must be estimated. The quantity $(J-2)s^2/\sum_{j=1}^J (\hat{\theta}_j - \mu)^2$ provides an unbiased estimate of this shrinkage factor as long as $J \geq 3$, resulting in an EB posterior mean equal to $\hat{\theta}_j^{JS}$ ⁵.

While the James/Stein estimator can be derived from an EB approach based on a normal mixing distribution, linear shrinkage reduces MSE whether or not this model is true. In particular, the result in [\(24\)](#) holds for any configuration of θ_j 's and any μ . For the purpose of reducing aggregate MSE, the θ_j 's need not be drawn from a normal distribution, or even viewed as random effects at all. The unknown parameters also need not be centered at μ , though shrinkage will result in larger MSE improvements if the prior mean is near the average of

⁵ With normal noise and a normal mixing distribution, the marginal distribution of each estimate is $\hat{\theta}_j \sim N(\mu, s^2 + \sigma^2)$. This implies the scaled sum of squared deviations $(s^2 + \sigma^2)^{-1} \sum_{j=1}^J (\hat{\theta}_j - \mu)^2$ follows a χ^2 distribution with J degrees of freedom. Its reciprocal then follows an inverse χ^2 distribution, which has mean $1/(J-2)$ for $J \geq 3$. It follows that $E[(J-2)s^2/\sum_{j=1}^J (\hat{\theta}_j - \mu)^2] = s^2/(s^2 + \sigma^2)$ when $J \geq 3$.

the θ_j 's. The sum of squares in the denominator of the shrinkage coefficients in [equation \(23\)](#) provides an empirical measure of how far the parameters tend to fall from μ , letting the data dictate the appropriate amount of shrinkage.

Versions of the James/Stein result generalize to more complex settings than the simple homoskedastic noise model in [equation \(20\)](#). These include scenarios where the noise variance s^2 is unknown so must be estimated from the microdata, different noise variances or correlation in noise across units, or an estimated prior mean ([Lindley, 1962](#); [Efron and Morris, 1973b](#); [Bock, 1975](#)). The case where the standard error s_j varies across groups is especially relevant since this is likely to be true in any realistic value-added analysis. With at least three units for each value of s_j the linear shrinkage formula in [\(23\)](#) can in principle be implemented separately for every s_j , in which case the basic James/Stein Theorem immediately applies. This amounts to using an unrestricted conditional prior distribution of the form

$$\theta_j | s_j \sim \mathcal{N}(\mu(s_j), \sigma^2(s_j)) \quad (25)$$

for EB shrinkage. In practice shrinking separately for each value of s_j may not be feasible, which motivates less flexible priors imposing restrictions on how the distribution of θ_j varies with s_j . An important caution is that such restrictions may break the James/Stein guarantee of a reduction in MSE.⁶ The standard EB posterior mean in [\(15\)](#) assumes independence of θ_j and s_j , and may perform poorly if effect sizes and standard errors are correlated ([Chen, 2023](#)). I return to such precision-dependence issues in [Section 2.6](#).

2.2.3 Compound decision problems

The James/Stein Theorem shows that EB shrinkage reduces MSE even when the normal random effects model motivating this method is wrong. Then why does shrinkage yield improvements? The requirement that $J \geq 3$ in the James/Stein Theorem provides a clue. Rather than an assumption on the data generating process, the value of EB shrinkage depends on the objective of the econometric analyst. The gains highlighted in the James/Stein Theorem arise because of the form of the loss function in [equation \(21\)](#), which cares about total mean squared error summed over all J units. This idea is generalized with a loss function of the form

$$\mathcal{L}(\delta_1, \dots, \delta_J; \theta_1, \dots, \theta_J) = \sum_{j=1}^J \ell(\delta_j, \theta_j), \quad (26)$$

⁶Since the James/Stein Theorem implies a reduction in MSE for shrinkage toward any constant, it is not necessary to allow the prior mean to depend on s_j to guarantee an improvement. Implementing the shrinkage formula in [\(23\)](#) separately for each s_j – but shrinking all units towards a common prior mean μ – reduces aggregate MSE as long as there are at least 3 units with each value of s_j . In contrast, imposing restrictions on how the prior variance varies with s_j may break the James/Stein guarantee.

where δ_j represents a decision for unit j and $\ell(\theta_j, \delta_j)$ gives the loss associated with making decision δ_j for a unit with parameter θ_j .

Such a loss function gives rise to a *compound decision problem* (Robbins, 1951; Gu and Koenker, 2016) in which a sequence of similar independent decisions are considered for each of several units and the decisionmaker seeks to minimize aggregate loss. A decisionmaker with this loss function views units as *exchangeable* – the labels on the units are irrelevant, so the decisionmaker does not distinguish between different configurations of outcomes across units that lead to the same total loss. This notion of exchangeability motivates the abstraction in model (6) treating the θ_j 's as draws from a common distribution G . The James/Stein Theorem reveals that using this conceptual device to incorporate distributional information via EB shrinkage improves expected outcomes in a compound decision problem.

Applying shrinkage outside the compound decision context can quickly lead to *reductio ad absurdum* arguments against EB methods. Should a labor economist shrink estimates of the elasticity of labor supply, the return to schooling, and the intergenerational income elasticity toward one another?⁷ Shrinkage seems inappropriate in this scenario because it is hard to imagine a decision for which total error across these three parameters is the relevant criterion. On the other hand, a framework that seeks to minimize aggregate error seems more natural when studying distributions of parameters across large sets of similar units like firms, schools, or neighborhoods. This makes the EB framework appealing for value-added analysis. Sections 2.5 and 3.6 present further analysis of EB methods in a compound decision setup.

2.3 Practical shrinkage issues

This subsection considers some issues of interpretation and implementation that often arise in value-added studies using empirical Bayes methods.

2.3.1 Distributions of true parameters, unbiased estimates, and posterior means

In any EB value-added analysis it is worth keeping in mind the distinction between true unobserved effects, unbiased estimates, and shrunk posterior means. For instance, suppose we are interested in summarizing heterogeneity in effectiveness across schools by looking at the variance of school value-added. This is a question about the variability of the true parameters θ_j , which

⁷ This facetious example intentionally mimics Robbins' (1951) discussion of jointly shrinking observations on “a butterfly in Ecuador...an oyster in Maryland...[and] the temperature of a star.” As noted by Lai and Siegmund (2018), examples of this type led to early skepticism regarding the practical value of the compound decision framework (see also Efron and Morris, 1973a).

are unobserved. Due to sampling error, the variance of the observed unbiased estimates $\hat{\theta}_j$ is too big relative to the variance of true effects:

$$\text{Var}(\hat{\theta}_j|s_j) = \sigma_\theta^2 + s_j^2 > \sigma_\theta^2. \quad (27)$$

In contrast, the variance of the oracle posterior means θ_j^* is too small relative to the variance of true effects:

$$\text{Var}(\theta_j^*|s_j) = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_j^2} \right) \sigma_\theta^2 < \sigma_\theta^2. \quad (28)$$

With precisely-estimated hyperparameters the EB posterior means $\hat{\theta}_j^*$ should be expected to exhibit variance less than σ_θ^2 as well. As a result, the mixing variance σ_θ^2 will typically lie in between the variances of $\hat{\theta}_j$ and $\hat{\theta}_j^*$ across groups.

The bias-corrected variance estimator in equation (9) splits this difference to recover the variance of unobserved effects. This estimator removes the expected contribution of noise from the variance of $\hat{\theta}_j$'s, yielding a consistent estimate of the true mixing variance σ_θ^2 . The estimate $\hat{\sigma}_\theta^2$ is therefore the right metric for summarizing variation in effects across units. More generally, the deconvolution estimate \hat{G} from step 2 of the EB recipe is the appropriate object to use for studying questions about the distribution of latent parameters. A plot of the shrunk posterior means $\hat{\theta}_j^*$ should not be expected to reproduce features of the mixing distribution G .

2.3.2 Shrinkage and regression

Another common application of EB shrinkage arises in the context of group-level OLS regression. Suppose we are interested in an OLS regression of an observed variable for unit j , Z_j , on the unknown parameter θ_j :

$$Z_j = \alpha_0 + \alpha_1 \theta_j + e_j. \quad (29)$$

The slope coefficient $\alpha_1 = \text{Cov}(Z_j, \theta_j)/\text{Var}(\theta_j)$ measures the change in Z_j associated with a one-unit increase in θ_j . Examples of analyses in this mold include Chetty et al.'s (2014b) study linking labor market outcomes to teacher test score effects and Abdulkadiroglu et al.'s (2020) study relating measures of school popularity to test score value-added.

Regression (29) is not feasible since the regressor θ_j is not observed, and substituting the noisy estimate $\hat{\theta}_j$ in its place biases the resulting slope coefficient. To see this, suppose the standard deviation of noise in $\hat{\theta}_j$ is common across units ($s_j = s \forall j$) and the estimation error $\hat{\theta}_j - \theta_j$ is uncorrelated with e_j . Then the slope coefficient from a regression of Z_j on $\hat{\theta}_j$ is

$$\begin{aligned}
 \frac{\text{Cov}(Z_j, \hat{\theta}_j)}{\text{Var}(\hat{\theta}_j)} &= \frac{\text{Cov}(\alpha_0 + \alpha_1 \theta_j + e_j, \theta_j + (\hat{\theta}_j - \theta_j))}{\text{Var}(\theta_j + (\hat{\theta}_j - \theta_j))} \\
 &= \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s^2} \right) \alpha_1.
 \end{aligned} \tag{30}$$

Putting the noisy estimate $\hat{\theta}_j$ on the right-hand side of a regression therefore yields bias toward zero. This is an instance of the standard result that classical measurement error in a regressor causes attenuation bias.

Empirical Bayes shrinkage corrects this bias. The regression of Z_j on the oracle posterior mean θ_j^* yields

$$\begin{aligned}
 \frac{\text{Cov}(Z_j, \theta_j^*)}{\text{Var}(\theta_j^*)} &= \frac{\text{Cov}\left(Z_j, \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s^2}\right)\hat{\theta}_j + \left(\frac{s^2}{\sigma_\theta^2 + s^2}\right)\mu_\theta\right)}{\text{Var}\left(\left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s^2}\right)\hat{\theta}_j + \left(\frac{s^2}{\sigma_\theta^2 + s^2}\right)\mu_\theta\right)} \\
 &= \left(\frac{\sigma_\theta^2 + s^2}{\sigma_\theta^2}\right) \frac{\text{Cov}(Z_j, \hat{\theta}_j)}{\text{Var}(\hat{\theta}_j)} \\
 &= \alpha_1.
 \end{aligned} \tag{31}$$

[Equation \(31\)](#) shows that using the shrunk posterior mean θ_j^* as a regressor recovers the same coefficient as using the true θ_j . This is a version of errors-in-variables regression, which uses the signal-to-noise ratio of the independent variable to correct attenuation bias due to measurement error. An empirically-feasible version of this correction substitutes the EB posterior mean $\hat{\theta}_j^*$ in place of θ_j^* .⁸ Note that while shrinkage fixes attenuation bias due to noise in $\hat{\theta}_j$, this noise still reduces the precision of the resulting estimate since the shrunk regressor θ_j^* has lower variance than the ideal regressor θ_j by [equation \(28\)](#).

The situation is reversed when the unknown parameter θ_j appears on the left-hand side of a regression rather than the right. Consider the OLS regression

$$\theta_j = \beta_0 + \beta_1 Z_j + u_j. \tag{32}$$

The slope coefficient $\beta_1 = \text{Cov}(\theta_j, Z_j)/\text{Var}(Z_j)$ measures the change in θ_j associated with a one unit increase in Z_j . For example, [Chetty and Hendren \(2018\)](#) investigate the correlates of neighborhood quality by regressing

⁸ When the standard error s_j varies across units a regression of Z_j on θ_j^* recovers α_1 if θ_j is independent of s_j , but may not recover α_1 when θ_j and s_j are correlated (see [Section 2.6](#)). One solution in this heteroskedastic case is to shrink all of the $\hat{\theta}_j$'s using a pooled signal-to-noise ratio equal to $\kappa = \sigma_\theta^2/\text{Var}(\hat{\theta}_j)$, which can be estimated with $\hat{\kappa} = \hat{\sigma}_\theta^2/[J^{-1}\sum_j (\hat{\theta}_j - \hat{\mu}_\theta)^2]$. A regression of Z_j on $\kappa\hat{\theta}_j$ recovers α_1 regardless of correlation between θ_j and s_j , though this estimator will be less efficient than regressing Z_j on θ_j^* when θ_j and s_j are independent.

neighborhood effects on location characteristics, and [Kline et al. \(2020\)](#) study the attributes of high-paying employers by regressing firm earnings effects on firm covariates. Assuming the estimation error $\hat{\theta}_j - \theta_j$ is independent of Z_j across units, replacing θ_j with the unbiased estimate $\hat{\theta}_j$ on the left-hand side of (32) recovers the target regression coefficient:

$$\frac{Cov(\hat{\theta}_j, Z_j)}{Var(Z_j)} = \frac{Cov(\theta_j, Z_j)}{Var(Z_j)} + \frac{Cov(\hat{\theta}_j - \theta_j, Z_j)}{Var(Z_j)} = \beta_1. \quad (33)$$

This is a consequence of the fact that classical measurement error on the left-hand side of a regression does not lead to bias. In contrast, shrinkage on the left introduces non-classical measurement error that does generate bias. The slope coefficient from a regression of θ_j^* on Z_j is

$$\frac{Cov(\theta_j^*, Z_j)}{Var(Z_j)} = \frac{Cov\left(\left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s^2}\right)\hat{\theta}_j + \left(\frac{s^2}{\sigma_\theta^2 + s^2}\right)\mu_\theta, Z_j\right)}{Var(Z_j)} = \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s^2}\right)\beta_1. \quad (34)$$

Shrinkage of the dependent variable therefore attenuates the resulting coefficient toward zero. This combination of results suggests a rule-of-thumb for EB shrinkage in a regression context: shrinkage on the right fixes bias, but shrinkage on the left causes bias.

2.3.3 Long regression or mean residuals? Correlated versus uncorrelated random effects

The initial estimation step in many empirical Bayes value-added analyses is an OLS regression like (5), which includes a set of group indicator variables along with a vector of controls X_i . Some studies instead start by partialing controls out of the outcome variable, then apply empirical Bayes methods to group means of the resulting residuals (e.g., [Chetty et al., 2014a; Jackson et al., 2020](#)). How should we interpret differences in results between these approaches?

This question can be answered with the familiar OLS omitted variable bias formula. Consider a mean residual strategy applied to school value-added estimation in Example 2. Let $\gamma^s = E[X_i X_i']^{-1} E[Y_i]$ denote the vector of coefficients from a short regression of Y_i on only the controls X_i , omitting the school indicators D_{ij} . This short regression coefficient is linked to the long regression coefficient γ from [equation \(5\)](#) by the relation

$$\begin{aligned} \gamma^s &= \gamma + E[X_i X_i']^{-1} E\left[X_i \left(\sum_j D_{ij} \theta_j\right)\right] \\ &= \gamma + E[X_i X_i']^{-1} E[\bar{X}_{\mathbf{J}(i)} \theta_{\mathbf{J}(i)}], \end{aligned} \quad (35)$$

where $\mathbf{J}(i) = \{j : D_i = j\}$ encodes the school attended by student i , $\bar{X}_j = E[X_i | D_i = j]$ is the mean of X_i at school j , and the second line uses the

law of iterated expectations. The school means of the short regression residuals $Y_i - X'_i \gamma^s$ are then

$$\begin{aligned}\theta_j^s &= E[Y_i - X'_i \gamma^s | D_i = j] \\ &= E[\theta_j - X'_j (\gamma^s - \gamma) | D_i = j] \\ &= \theta_j - \bar{X}'_j E[X_i X'_i]^{-1} E[\bar{X}_{J(i)} \theta_{J(i)}].\end{aligned}\quad (36)$$

Equation (36) shows that the value-added parameters from the long regression and mean residual approaches coincide when $\bar{X}'_j E[X_i X'_i]^{-1} E[\bar{X}_{J(i)} \theta_{J(i)}] = 0 \forall j$. One scenario satisfying this condition is when the averages of the control variables do not vary across groups, so $\bar{X}_j = E[X_i] = 0 \forall j$ (recall that the mean of X_i is normalized to zero). This is unlikely in realistic applications since the motivation for including controls in the first place is typically to adjust for group differences in the covariates. In the school value-added context, this would require every school to enroll a student population with the same demographics and mean past achievement.

A second possible justification for the mean residual approach is an assumption that $E[\bar{X}_{J(i)} \theta_{J(i)}] = 0$. This allows for the possibility that the school average covariates \bar{X}_j differ across schools but rules out a relationship between these averages and school value-added θ_j . When X_i is a lagged student test score, for example, this restriction requires schools enrolling students with higher average past achievement to be no better (or worse) in terms of causal value-added. The mean residual approach is therefore implicitly based on an *uncorrelated random effects* model in which the θ_j 's are independent of \bar{X}_j . In contrast, the long regression approach is compatible with a *correlated random effects* (CRE) model in which the distribution of θ_j may depend arbitrarily on \bar{X}_j (Chamberlain, 1982).⁹ Since there is typically no reason to rule out a relationship between θ_j and \bar{X}_j , the long regression approach is likely to be preferable in most applications. The uncorrelated random effects assumption can also be tested with a Hausman (1978) test comparing coefficients from the long and short regressions.¹⁰

2.4 Generalizations of linear shrinkage

This subsection considers extensions of the basic empirical Bayes recipe to cases incorporating observed unit characteristics, multiple parameters per unit, and multiple estimates of each unit's parameter.

⁹ Note that the long regression control coefficient γ can be obtained either by regressing Y_i on X_i and the D_{ij} , or by regressing Y_i on X_i controlling for $\bar{X}_{J(i)}$ (Mundlak, 1978). The long regression value-added coefficients θ_j are means of the residual $Y_i - X'_i \gamma$.

¹⁰ This test is sometimes described as comparing “fixed effects” and “random effects” models, terminology that mixes the question of whether the θ_j 's are treated as fixed or random with the question of whether they are related to \bar{X}_j .

2.4.1 Adding covariates

Suppose we observe a vector Z_j of covariates for each unit j (including a constant). Consider a conditional normal/normal model of the form:

$$\hat{\theta}_j | \theta_j, s_j, Z_j \sim N(\theta_j, s_j^2), \quad (37)$$

$$\theta_j | s_j, Z_j \sim N(Z'_j \mu, \sigma_r^2), \quad (38)$$

where μ describes the relationship between effect sizes θ_j and observed characteristics Z_j , and σ_r^2 measures residual variance in θ_j not explained by Z_j . By the logic of [Section 2.3](#), a regression of $\hat{\theta}_j$ on Z_j yields an unbiased estimate $\hat{\mu}$. We can then estimate the unexplained variance in θ_j by deconvolving the resulting residuals with the estimator $\hat{\sigma}_r^2 = J^{-1} \sum_j [(\hat{\theta}_j - Z'_j \hat{\mu})^2 - s_j^2]$.

The posterior mean for θ_j implied by [equations \(37\)](#) and [\(38\)](#) is:

$$\theta_j^* = \left(\frac{\sigma_r^2}{\sigma_r^2 + s_j^2} \right) \hat{\theta}_j + \left(\frac{s_j^2}{\sigma_r^2 + s_j^2} \right) Z'_j \mu. \quad (39)$$

Plugging $\hat{\mu}$ and $\hat{\sigma}_r^2$ into [equation \(39\)](#) yields an EB posterior mean that shrinks $\hat{\theta}_j$ toward an estimated linear index $Z'_j \hat{\mu}$ rather than toward a constant. This approach borrows more strength from units that are similar on observed dimensions, potentially yielding further reductions in MSE relative to $\hat{\theta}_j$. The resulting EB posterior places less weight on the noisy $\hat{\theta}_j$ if the covariates explain more of the variance in θ_j , reflected in a smaller residual variance σ_r^2 . It is straightforward to allow the residual variance σ_r^2 to depend on Z_j as well.

2.4.2 Multivariate EB

Suppose we expand the school value-added setup in Example 2 so that schools may affect each of K outcomes for a given student, labeled Y_{i1}, \dots, Y_{iK} . These might include test scores as well as longer-term outcomes like educational attainment, criminal justice involvement, or earnings (e.g., [Jackson et al., 2020](#); [Beuermann et al., 2022](#); [Rose et al., 2022](#)). This yields a separate version of the VAM regression in [equation \(5\)](#) for each outcome:

$$Y_{ik} = \sum_j \theta_{jk} D_{ij} + X'_i \gamma_k + \epsilon_{ik}, \quad k \in \{1, \dots, K\}. \quad (40)$$

School j 's quality is now described by a $K \times 1$ vector $\theta_j = (\theta_{j1}, \theta_{j2}, \dots, \theta_{jK})'$ collecting its value-added on all outcomes. Seemingly Unrelated Regression (SUR) estimation of system [\(40\)](#) yields an estimate $\hat{\theta}_j$ of this vector, along with a $K \times K$ sampling variance matrix V_j for each school. This matrix has squared standard errors s_{jk}^2 for each $\hat{\theta}_{jk}$ along its diagonal and off-diagonal terms measuring sampling covariances between $\hat{\theta}_{jk}$ and $\hat{\theta}_{jm}$ for $m \neq k$. These covariances will generally be non-zero if the underlying student-level outcomes are correlated.

Extending the normal/normal model to the multivariate case, we have:

$$\hat{\theta}_j | \theta_j, V_j \sim N(\theta_j, V_j), \quad (41)$$

$$\theta_j | V_j \sim N(\mu_\theta, \Sigma_\theta), \quad (42)$$

where the mixing distribution in (42) is now parameterized by a $K \times 1$ mean μ_θ and $K \times K$ variance matrix Σ_θ . The off-diagonal elements of Σ_θ describe how schools' effects on different outcomes are related. For example, do schools that boost test scores also boost high school graduation? The deconvolution step estimates these hyperparameters with:

$$\hat{\mu}_\theta = J^{-1} \sum_j \hat{\theta}_j, \quad \hat{\Sigma}_\theta = J^{-1} \sum_j [(\hat{\theta}_j - \hat{\mu}_\theta)(\hat{\theta}_j - \hat{\mu}_\theta)' - V_j]. \quad (43)$$

As before, the bias-corrected variance matrix estimator $\hat{\Sigma}_\theta$ removes excess variability in the $\hat{\theta}_{jk}$'s due to sampling error, as well as the contribution of noise to the covariances of estimates across outcomes.

Finally, in the shrinkage step, the multivariate oracle posterior mean implied by equations (41) and (42) is a precision-matrix-weighted average of the unbiased estimate $\hat{\theta}_j$ and the prior mean:

$$\theta_j^* = (V_j^{-1} + \Sigma_\theta^{-1})^{-1}(V_j^{-1}\hat{\theta}_j + \Sigma_\theta^{-1}\mu_\theta). \quad (44)$$

A multivariate EB posterior mean $\hat{\theta}_j^*$ plugs hyperparameter estimates from (43) into (44). This estimator borrows strength across multiple outcomes as well as across the ensemble of schools when predicting any one of the outcome-specific value-added parameters θ_{jk} . For instance, if effects on test scores (outcome 1) and high school graduation (outcome 2) are highly correlated across schools, then the posterior mean graduation effect $\hat{\theta}_{j2}^*$ may place substantial weight on the estimated test score effect $\hat{\theta}_{j1}$, especially if $\hat{\theta}_{j2}$ is much noisier than $\hat{\theta}_{j1}$. Other applications of multivariate EB include value-added models with heterogeneous effects, so that each unit is associated with treatment effect parameters for multiple subgroups (e.g., [Abdulkadiroğlu et al., 2020](#); [Avivi, 2024](#)); and instrumental variables settings with imperfect compliance, with each unit characterized by multiple parameters governing outcomes and selection into treatment (e.g., [Raudenbush et al., 2012](#); [Walters, 2015](#)).

2.4.3 Combining estimators

The empirical Bayes approach extends naturally to cases with multiple estimates of each parameter, some possibly biased. In the school value-added example, consider a population OLS regression of Y_i on school indicators with controls for X_i :

$$Y_i = \sum_{j=1}^J \alpha_j D_{ij} + X_i' \Gamma + v_i. \quad (45)$$

This equation differs from the causal value-added model (5) because the OLS residual v_i satisfies $E[D_{ij}v_i] = 0 \forall j$ by definition, while the error term ϵ_i from the causal model may be correlated with the D_{ij} 's if selection-on-observables fails to hold. This possibility is represented by a list of bias parameters $b_j = \theta_j - \alpha_j$ measuring deviations between causal value-added parameters and OLS regression coefficients.

Let $\{\hat{\alpha}_j\}_{j=1}^J$ denote OLS value-added estimates based on fitting [equation \(45\)](#) in a random sample of students. Suppose we also have access to a second set of estimates $\{\hat{\theta}_j\}_{j=1}^J$ from an alternative research design that is not contaminated by selection bias. For example, the $\hat{\theta}_j$'s may be instrumental variables (IV) estimates derived from randomized school entrance lotteries. The $\hat{\theta}_j$'s are assumed to be (asymptotically) unbiased estimates of the causal θ_j 's, but are likely to be noisier than the OLS $\hat{\alpha}_j$ estimates. We then have the model:

$$(\hat{\alpha}_j, \hat{\theta}_j)' | \theta_j, b_j, s_{j\alpha}, s_{j\theta}, c_j \sim N\left((\theta_j + b_j, \theta_j)', \begin{bmatrix} s_{j\alpha}^2 & c_j \\ c_j & s_{j\theta}^2 \end{bmatrix}\right). \quad (46)$$

Here $s_{j\alpha}^2$ and $s_{j\theta}^2$ are the sampling variances of the OLS and IV estimates, and c_j is their sampling covariance. A [Hausman \(1978\)](#)-style overidentification test comparing OLS and IV is a test of the null hypothesis that $b_j = 0 \forall j$ ([Angrist et al., 2016](#)).¹¹

If this overidentification test rejects, we may conclude that the OLS estimates are contaminated by selection bias. The biased $\hat{\alpha}_j$'s may be more precise than the unbiased $\hat{\theta}_j$'s, however, so throwing away OLS completely may discard useful information about school quality. An empirical Bayes approach incorporates this information by forming a best guess of value-added for each school that trades off the bias of OLS against the imprecision of IV.

Consider a model of the causal-valued parameters θ_j and bias parameters b_j as draws from a multivariate normal mixing distribution:

$$(\theta_j, b_j)' | s_{j\alpha}, s_{j\theta}, c_j \sim N\left((\mu_\theta, \mu_b)', \begin{bmatrix} \sigma_\theta^2 & \sigma_{\theta b} \\ \sigma_{\theta b} & \sigma_b^2 \end{bmatrix}\right). \quad (47)$$

The variance parameters σ_θ^2 and σ_b^2 describe the variability of causal value-added and selection bias across schools, while the covariance $\sigma_{\theta b}$ determines whether OLS tends to over-rate or under-rate higher value-added schools. These hyperparameters can be estimated by deconvolving the joint distribution of $(\hat{\alpha}_j, \hat{\theta}_j)$. For example, we can estimate the variance of bias across schools as $\hat{\sigma}_b^2 = J^{-1} \sum_j [(\hat{\alpha}_j - \hat{\theta}_j - \hat{\mu}_b)^2 - (s_{j\alpha}^2 + s_{j\theta}^2 - 2c_j)]$, where $\hat{\mu}_b = J^{-1} \sum_j (\hat{\alpha}_j - \hat{\theta}_j)$. The estimator $\hat{\sigma}_b^2$ looks for overdispersion in the difference between OLS and

¹¹ Under the classical Gauss-Markov assumptions and with no selection bias OLS is the efficient linear estimator of the θ_j 's, in which case $c_j = s_{j\alpha}^2$ so that $Var(\hat{\theta}_j - \hat{\alpha}_j) = s_{j\theta}^2 - s_{j\alpha}^2$.

IV beyond what should be expected from sampling error, implying variation in selection bias.

Posterior means $\theta_j^* = E[\theta_j | \hat{\theta}_j, \hat{\alpha}_j, s_{j\alpha}, s_{j\theta}, c_j]$ based on model (46)–(47) provide minimum MSE forecasts of value-added for each school using all available information. In a scenario where the OLS estimates are extremely precise ($s_{j\alpha} \approx 0$), the oracle posterior mean for θ_j is given by:

$$\theta_j^* = \left(\frac{(1 - R^2)\sigma_\theta^2}{(1 - R^2)\sigma_\theta^2 + s_{j\theta}^2} \right) \hat{\theta}_j + \left(\frac{s_{j\theta}^2}{(1 - R^2)\sigma_\theta^2 + s_{j\theta}^2} \right) (\rho(\hat{\alpha}_j - \mu_b) + (1 - \rho)\mu_\theta), \quad (48)$$

where $\rho = \text{Cov}(\hat{\alpha}_j, \theta_j)/\text{Var}(\hat{\alpha}_j) = (\sigma_\theta^2 + \sigma_{\theta b})/(\sigma_\theta^2 + \sigma_b^2 + 2\sigma_{\theta b})$ is the slope coefficient from a regression of θ_j on $\hat{\alpha}_j$ (the *reliability* of OLS), and R^2 is the R-squared from this regression. This linear shrinkage formula can be seen as a version of equation (39) treating the OLS estimate $\hat{\alpha}_j$ as a covariate that predicts θ_j . An empirical implementation plugs deconvolution estimates of ρ , R^2 , and other hyperparameters into (48) to produce an EB posterior mean. Chetty and Hendren (2018) use such a strategy to combine quasi-experimental estimates of neighborhood effects based on cross-neighborhood moves with observational estimates based on levels of permanent resident outcomes. Angrist et al. (2017; 2024a) generalize this approach to estimate school value-added in an under-identified scenario where lotteries are unavailable for some schools.

2.5 EB decision rules

We have seen that linear shrinkage delivers posterior mean estimates with low mean squared error. It is often of interest to consider goals other than minimizing MSE. In the school value-added example, suppose a school district administrator seeks to select schools with value-added below a cutoff c for reform or closure, and incurs different costs for erroneously selecting high-performing schools (type I errors) or failing to select low-performing schools (type II errors). The objective of the administrator is formalized in a compound decision problem of the form of (26), with component-wise loss function

$$\ell(\delta_j, \theta_j) = \delta_j 1\{\theta_j \geq c\} + (1 - \delta_j) 1\{\theta_j < c\} \xi, \quad (49)$$

where $\delta_j \in \{0, 1\}$ indicates selection of school j , ξ is the cost of a type II error, and the cost of a type I error is normalized to one. Gu and Koenker (2023b) discuss empirical Bayes tail selection problems with loss functions of this type.

Since the θ_j 's are unknown, the administrator chooses a decision rule $\delta(\hat{\theta}_j, s_j)$ to minimize *risk*(expected loss). With J schools the risk of a decision rule δ is given by.

$$\mathcal{R}(\delta) = \sum_{j=1}^J E[\ell(\delta(\hat{\theta}_j, s_j), \theta_j) | s_j]$$

$$= \sum_{j=1}^J \int \int \ell(\delta(\hat{\theta}, s_j), \theta) \frac{1}{s_j} \phi\left(\frac{\hat{\theta} - \theta}{s_j}\right) d\hat{\theta} dG(\theta). \quad (50)$$

The risk-minimizing decision rule is $\delta^* = \arg \min_{\delta \in \mathcal{D}} \mathcal{R}(\delta)$, where \mathcal{D} is the set of functions mapping $(\hat{\theta}_j, s_j)$ to binary decisions. With the loss function in (49), the optimal decision rule is

$$\delta^*(\hat{\theta}_j, s_j) = 1 \left\{ \Pr[\theta_j < c | \hat{\theta}_j, s_j] \geq \frac{1}{1 + \xi} \right\}. \quad (51)$$

That is, the administrator selects schools with sufficiently high posterior probability of falling below the cutoff c , where the confidence necessary to make a selection depends on the relative costs of type I and type II errors. With the normal mixing distribution in (7) this decision rule becomes:

$$\delta^*(\hat{\theta}_j, s_j) = 1 \left\{ \left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + s_j^2} \right) \hat{\theta}_j + \left(\frac{s_j^2}{\sigma_\theta^2 + s_j^2} \right) \mu_\theta + \sqrt{\frac{\sigma_\theta^2 s_j^2}{\sigma_\theta^2 + s_j^2}} \Phi^{-1} \left(\frac{1}{1 + \xi} \right) \leq c \right\}. \quad (52)$$

An EB decision rule $\hat{\delta}^*(\hat{\theta}_j, s_j)$ plugs estimated hyperparameters $\hat{\mu}_\theta$ and $\hat{\sigma}_\theta^2$ into equation (52).

Equation (52) reveals that the solution to this tail selection problem will not generally select schools based on the posterior mean. The optimal decision is a cutoff in the $(1 + \xi)^{-1}$ posterior quantile, which adds an adjustment to the posterior mean based on the posterior standard deviation $\sqrt{\sigma_\theta^2 s_j^2 / (\sigma_\theta^2 + s_j^2)}$. This means schools with the same posterior mean θ_j^* will be treated differently based on their standard errors s_j . Whether this adjustment rewards or punishes schools with large standard errors depends on whether $\Phi^{-1}(1/(1 + \xi))$ is positive or negative, which depends in turn on whether type I or type II errors are more costly (note that this function crosses zero at $\xi = 1$). When the administrator is especially concerned about mistakenly selecting high-performing schools (so ξ is small), she gives the benefit of the doubt to schools with poor posterior means but large standard errors. When the administrator is more concerned about failing to select low performers (so ξ is large), she penalizes schools with large s_j since there is a reasonable chance a school with a very noisy estimate belongs to the lower tail even if its posterior mean appears favorable.

This simple example demonstrates that the right posterior prediction to report in the EB shrinkage step depends on the goal of the analysis and associated loss function. A clear statement of the loss function is therefore a central part of any coherent EB shrinkage analysis. Specifically, what is the purpose of reporting unit-specific posterior estimates, and how do we expect them to be used? Reporting posterior means is sensible if the aim is to maximize average outcomes by directing consumers to units with high

value-added. If the goal is to find units in the tail of the distribution while limiting the likelihood of mistakes, a posterior quantile may be more appropriate. Ordered lists of EB posterior predictions may lead to scrutiny of the highest- and lowest-ranked performers, a tendency that [Gu and Koenker \(2023b\)](#) term the “league table mentality.” As I discuss further in [Section 3.5](#), neither posterior means nor posterior quantiles are generally optimal if the objective is to accurately convey information about relative ranks.

2.6 Precision-dependence

So far I have focused on models assuming independence of the effect sizes θ_j and sampling variances s_j^2 across units. In practice these parameters may be correlated. Potential sources of such precision-dependence can be seen in the normal-means problem of Example 1. Recall that the sampling variance of $\hat{\theta}_j$ in this case equals $s_j^2 = \sigma_\theta^2/n_j$. A correlation between θ_j and s_j^2 will arise if the effect size θ_j is related to group size n_j , the within-group outcome variance σ_j^2 , or both.

There are often economic reasons to expect such correlations. In the teacher value-added context, a large body of research establishes that teaching skill improves with experience ([Staiger and Rockoff, 2010](#)). Lower-quality teachers may also exit the teaching profession faster ([Goldhaber et al., 2011](#)). Both of these forces suggest more data will be available to estimate value-added for higher-quality teachers, implying a negative correlation between θ_j and s_j^2 . Research on hospital value-added shows that higher-quality hospitals tend to have higher market shares and gain market share over time, potentially generating a positive relationship between hospital quality and sample size ([Chandra et al., 2016,2023](#)). In empirical Bayes meta-analyses of the distribution of research findings across studies, precision-dependence may arise because studies are screened on the basis of statistical significance (publication bias) or because researchers concerned with statistical power choose sample sizes with an eye toward expected effect sizes ([Sterne and Narbord, 2004](#)).

To understand the implications of precision-dependence, consider the independent normal/normal model defined by equations (1) and (7). In this model the simple hyperparameter estimates in equations (8) and (9) are not efficient when s_j^2 varies across units. Instead, the following precision-weighted averages provide asymptotically efficient estimates of the mean and variance of G :

$$\hat{\mu}_\theta = \sum_{j=1}^J \left(\frac{[\sigma_\theta^2 + s_j^2]^{-1}}{\sum_{k=1}^J [\sigma_\theta^2 + s_k^2]^{-1}} \right) \hat{\theta}_j, \quad (53)$$

$$\hat{\sigma}_\theta^2 = \sum_{j=1}^J \left(\frac{[\sigma_\theta^2 + s_j^2]^{-2}}{\sum_{k=1}^J [\sigma_\theta^2 + s_k^2]^{-2}} \right) [(\hat{\theta}_j - \hat{\mu}_\theta)^2 - s_j^2]. \quad (54)$$

The weights in these formulas depend on the unknown mixing variance σ_θ^2 .¹² A one-step maximum likelihood approach continuously updates σ_θ^2 and μ_θ to simultaneously estimate the parameters and efficient weights, while a two-step approach substitutes in weights based on a first-step estimate of σ_θ^2 (e.g. the unweighted estimator in (9)). A simpler alternative is to use weights that are proportional to n_j or s_j^{-2} , which captures some of the gains to precision-weighting without the need to estimate the optimal weights.

Under the assumption that θ_j and s_j^2 are independent, these precision-weighting approaches will produce more precise estimates of the hyperparameters of G than the unweighted averages in (8) and (9). When θ_j and s_j^2 are correlated, however, precision-weighting changes the estimand, so the unequally-weighted averages in (53) and (54) will not generally be consistent for μ_θ and σ_θ^2 . The choice of whether or not to precision-weight when estimating the hyperparameters of G in the deconvolution step therefore involves a classic robustness/efficiency tradeoff.

Precision-dependence also affects the shrinkage step of the EB recipe. The standard EB linear shrinkage estimator in equation (15) is based on a prior distribution assuming independence between θ_j and s_j . Its performance will degrade if this assumption is false, and the shrinkage estimate may even perform worse than the unshrunk estimate $\hat{\theta}_j$ with some forms of precision-dependence.

This problem can be seen in a simple example. Suppose the true mixing distribution has mean zero and variance proportional to s_j^2 , so that $\theta_j|s_j \sim \mathcal{N}(0, s_j^2 \sigma^2)$. The oracle posterior mean in this model is $E[\theta_j|\hat{\theta}_j, s_j] = (\sigma^2/[1 + \sigma^2])\hat{\theta}_j$, which uses a constant shrinkage factor that does not depend on s_j^2 and therefore preserves the ordering of the raw $\hat{\theta}_j$'s. Because true effects are more variable for units with higher standard errors, we should not shrink noisier estimates more. A conventional linear shrinkage estimator of the form of (15) will instead apply more shrinkage to units with larger s_j^2 , changing the ordering of units relative to $\hat{\theta}_j$ (and the oracle posterior mean). This means that if we aim to select units with high average value-added, the unadjusted $\hat{\theta}_j$ performs better than standard linear shrinkage in this case.¹³ I next consider strategies for avoiding such problems by relaxing the assumption that effect sizes and standard errors are independent.

2.6.1 Testing and modeling precision-dependence

The assumption that θ_j and s_j are independent is testable. It is convenient to test this assumption with regression-based procedures checking whether the

¹² Equations (1) and (7) imply the marginal variance of $\hat{\theta}_j$ is $\sigma_\theta^2 + s_j^2$, while the marginal variance of $(\hat{\theta}_j - \mu_\theta)^2$ is $2(\sigma_\theta^2 + s_j^2)^2$. The estimators in equations (53) and (54) are therefore inverse-variance-weighted averages.

¹³ See Xie et al. (2012) for an analysis of the performance of various linear shrinkage rules in models with heteroskedasticity.

conditional mean and variance of θ_j depend on s_j . If θ_j and s_j are independent we have $E[\hat{\theta}_j|s_j] = \mu_\theta$ and $E[(\hat{\theta}_j - \mu_\theta)^2 | s_j] = \sigma_\theta^2$. Regressions of $\hat{\theta}_j$ and $(\hat{\theta}_j - \hat{\mu}_\theta)^2 - s_j^2$ on functions of s_j should therefore yield coefficients of zero. Such tests provide a practical first check for the importance of precision-dependence.

If dependence is present, we can incorporate it into the deconvolution step by estimating a conditional mixing distribution that depends on s_j . Chen (2023) studies a class of conditional location scale estimators (CLOSE) that allow the mean and variance of the prior to depend on precision. As a simple parametric version of such a strategy, consider the specification:

$$\theta_j = \psi_0 + \psi_1 \log s_j + s_j^{\psi_2} r_j, \quad (55)$$

with $r_j|s_j \sim N(0, \sigma_r^2)$. This model treats s_j as a covariate that shifts the mean and variance of G , and assumes features of the mixing distribution beyond the first two moments do not depend on precision.

Equation (55) implies the moment conditions $E[\hat{\theta}_j|s_j] = \psi_0 + \psi_1 \log s_j$ and $E[(\hat{\theta}_j - \psi_0 - \psi_1 \log s_j)^2 - s_j^2 | s_j] = s_j^{2\psi_2} \sigma_r^2$. The deconvolution step leverages these restrictions to estimate the hyperparameters of the conditional mixing distribution. A simple two-step approach is to estimate ψ_0 and ψ_1 with a linear regression of $\hat{\theta}_j$ on $\log s_j$, then estimate ψ_2 and σ_r^2 with a non-linear least squares regression of $[(\hat{\theta}_j - \hat{\psi}_0 - \hat{\psi}_1 \log s_j)^2 - s_j^2]$ on $s_j^{2\psi_2} \sigma_r^2$. Alternatively, we can estimate all the parameters in one step by embedding both moment conditions in a single generalized method of moments (GMM) procedure.

The shrinkage step incorporates precision-dependence by forming EB posteriors for the residuals r_j and transforming these residuals to produce posteriors for θ_j . Let $\hat{r}_j = (\hat{\theta}_j - \hat{\psi}_0 - \hat{\psi}_1 \log s_j)/s_j^{\psi_2}$ denote a noisy estimate of r_j constructed using hyperparameter estimates from the deconvolution step. Equations (1) and (55) imply $\hat{r}_j|r_j, s_j \sim N(r_j, s_j^{2(1-\psi_2)})$. This motivates the residual linear shrinkage estimator:

$$\hat{r}_j^* = \left(\frac{\hat{\sigma}_r^2}{\hat{\sigma}_r^2 + s_j^{2(1-\psi_2)}} \right) \hat{r}_j. \quad (56)$$

An EB posterior mean for θ_j accounting for precision-dependence is then given by:

$$\hat{\theta}_j^* = \hat{\psi}_0 + \hat{\psi}_1 \log s_j + s_j^{\psi_2} \hat{r}_j^*. \quad (57)$$

When effect sizes and precision are correlated, this estimator should be expected to improve aggregate MSE relative to both the unbiased estimate $\hat{\theta}_j$ and the conventional linear shrinkage estimator in (15) which neglects precision dependence.

2.6.2 Variance-stabilizing transformations

In some cases it is possible to transform the estimates $\hat{\theta}_j$ to have constant sampling variance rather than estimating the relationship between θ_j and s_j . Such *variance-stabilizing transformations* (VSTs) eliminate concerns about precision-dependence since effect sizes cannot be correlated with precision if precision is constant. In the normal noise model (1), suppose the squared standard error s_j^2 is a known function of the unobserved effect θ_j :

$$s_j^2 = v(\theta_j). \quad (58)$$

Consider a transformation of the form:

$$t(\theta) = \eta \int_{-\infty}^{\theta} v(u)^{-1/2} du \quad (59)$$

for a constant η . The delta method implies

$$\begin{aligned} \text{Var}(t(\hat{\theta}_j)|\theta_j) &\approx t'(\theta_j)^2 s_j^2 \\ &= \eta(v(\theta_j)^{-1/2})^2 v(\theta_j) \\ &= \eta, \end{aligned} \quad (60)$$

where the second line follows from Leibniz's rule. The transformed variable $\hat{t}_j = t(\hat{\theta}_j)$ then has the same sampling variance for all units, and its mixing distribution can be recovered by deconvolution without worrying about precision-dependence.

A VST is an attractive option for handling precision-dependence in settings where a known function $v(\cdot)$ satisfying (58) is available. This approach is less appealing when the right VST is unknown, though in some cases it may be possible to estimate $v(\cdot)$. This involves modeling s_j^2 as a deterministic function of θ_j as in equation (58), rather than modeling θ_j as a (random) function of s_j as in equation (55).¹⁴

Example 3: Binomial mixtures

Suppose the microdata Y_i are generated by independent Bernoulli trials with a different success probability θ_j for each group j . This implies the group-specific success counts $C_j = \sum_i D_{ij} Y_i$ follow a mixture of Binomial distributions:

$$C_j | \theta_j, n_j \sim \text{Bin}(n_j, \theta_j). \quad (61)$$

The variance of the binomial distribution is given by $\text{Var}(C_j|\theta_j, n_j) = n_j \theta_j (1 - \theta_j)$. The variance of the success rate $\hat{\theta}_j = n_j^{-1} C_j$ is then deterministically linked to the success probability as $s_j^2 = n_j^{-1} \theta_j (1 - \theta_j)$.

With binomial noise, sampling variance is stabilized by the Bartlett (1936) arcsine-square-root transformation:

¹⁴ Holland (1973) shows that a VST may not exist in the multivariate case where θ_j is a vector with more than one element.

$$t(\theta) = \sin^{-1}(\sqrt{\theta}). \quad (62)$$

The derivative of this transformation is $t'(\theta) = (2\sqrt{\theta(1-\theta)})^{-1}$, so the asymptotic variance of $\hat{t}_j = \sin^{-1}(\sqrt{\hat{\theta}_j})$ is given by

$$\begin{aligned} \text{Var}(\hat{t}_j | \theta_j, n_j) &\approx \left(\frac{1}{2\sqrt{\theta_j(1-\theta_j)}} \right)^2 \left(\frac{\theta_j(1-\theta_j)}{n_j} \right) \\ &= \frac{1}{4n_j}. \end{aligned} \quad (63)$$

The variance of \hat{t}_j no longer depends directly on θ_j , though some precision-dependence may remain if the sample sizes n_j vary and are correlated with θ_j . [Kline et al. \(2024\)](#) apply the arcsine-square-root transformation to analyze variation in callback rates across first names in a correspondence experiment. [Brown \(2008\)](#) considers an extended class of related binomial VSTs with improved finite-sample properties.

2.6.3 Noisy standard errors

While the standard errors s_j^2 are often treated as known, in practice it is usually necessary to estimate the sampling variance of each $\hat{\theta}_j$. Recall from Example 1 that with normal microdata the sample mean $\hat{\theta}_j$ is normally distributed with mean θ_j and variance $s_j^2 = \sigma_j^2/n_j$. If σ_j^2 is unknown it must be estimated from the data, typically with the unbiased variance estimator:

$$\hat{\sigma}_j^2 = \frac{1}{n_j - 1} \sum_i D_{ij}(Y_i - \hat{\theta}_j)^2. \quad (64)$$

The standard error estimate for unit j is then $\hat{s}_j^2 = \hat{\sigma}_j^2/n_j$, which includes error due to noise in $\hat{\sigma}_j^2$.

In principle we can account for this noise using a model for the sampling distribution of \hat{s}_j^2 . When the microdata Y_i are normally distributed, for example, the scaled sum of squares $(n_j - 1)\hat{\sigma}_j^2/\sigma_j^2$ follows a χ^2 distribution with $n_j - 1$ degrees of freedom. This implies \hat{s}_j^2 follows a Gamma distribution with shape parameter $(n_j - 1)/2$ and scale parameter $2\sigma_j^2/[n_j(n_j - 1)] = 2s_j^2/[n_j - 1]$:

$$\hat{s}_j^2 | s_j^2, n_j \sim \Gamma\left(\frac{n_j - 1}{2}, \frac{2s_j^2}{n_j - 1}\right). \quad (65)$$

This model can be used to implement a bivariate deconvolution that estimates the joint distribution of (θ_j, s_j^2) accounting for the noise in both $\hat{\theta}_j$ and \hat{s}_j^2 (see, e.g., [Gu and Koenker, 2017, 2023a](#)). Multivariate deconvolution may be empirically challenging, however, especially if precision-dependence is present so that the effect sizes θ_j and group-specific variances σ_j are correlated with one another or with sample size n_j .

Three considerations suggest the sampling error in \hat{s}_j^2 may not be consequential in typical applications. First, noise in estimates of sampling variance does not compromise estimates of some features of the mixing distribution. For example, the bias-corrected variance estimate in (9) is consistent for σ_θ^2 whether the true noise variance s_j^2 or an unbiased estimate \hat{s}_j^2 is used for bias correction (see Section 3.1 for elaboration on this point). Second, the noise in \hat{s}_j^2 will often be approximately independent of the noise in $\hat{\theta}_j$, ruling out more pernicious forms of measurement error in \hat{s}_j^2 . With normal data the sample mean and sample variance are independent.¹⁵ Finally, the noise in \hat{s}_j^2 disappears with sample size at a faster rate than the noise in $\hat{\theta}_j$. Equation (65) implies \hat{s}_j^2 has variance $2s_j^4/(n_j - 1)$, which will typically be much smaller than s_j^2 itself as long as n_j is reasonably large. This motivates the standard approach of treating the noise in \hat{s}_j^2 as negligible relative to the noise in $\hat{\theta}_j$.

2.7 Connections to machine learning

Empirical Bayes methods are closely connected to machine learning (ML) approaches to model selection and regularization. ML refers to a battery of empirical tools used to reduce overfitting in high-dimensional settings where the number of available predictors is large relative to sample size (see Varian, 2014; Mullainathan and Spiess, 2017; Athey and Imbens, 2019 for econometrically-oriented reviews). The idea of penalizing model complexity to limit overfitting bears close resemblance to the use of EB shrinkage to reduce variance. I next develop this connection further by highlighting well-known equivalences between EB shrinkage and simple machine learning approaches.

Return to the setup of Example 1 with normally-distributed microdata, and simplify further by assuming a common standard deviation of outcomes across groups ($\sigma_j = \sigma_y \forall j$). Suppose the θ_j 's are drawn from the normal mixing distribution in (7) and set the prior mean μ_θ to zero. The oracle posterior density for θ_j can be written

$$p(\theta_j | \mathbf{D}, \mathbf{Y}) = \frac{\left[\prod_{i=1}^N \left\{ \frac{1}{\sigma_y} \phi\left(\frac{Y_i - \theta_j}{\sigma_y} \right) \right\}^{D_{ij}} \right] \frac{1}{\sigma_\theta} \phi\left(\frac{\theta_j}{\sigma_\theta} \right)}{\int \left[\prod_{i=1}^N \left\{ \frac{1}{\sigma_y} \phi\left(\frac{Y_{ij} - \theta}{\sigma_y} \right) \right\}^{D_{ij}} \right] \frac{1}{\sigma_\theta} \phi\left(\frac{\theta}{\sigma_\theta} \right) d\theta}, \quad (66)$$

where \mathbf{D} and \mathbf{Y} are vectors collecting the group indicators D_{ij} and outcomes Y_i for all observations. We have already seen by equation (11) that this posterior distribution is normal with mean given by the linear shrinkage formula θ_j^* in (12). Since the mean and mode of a normal distribution coincide, we can

¹⁵ Independence of $\hat{\theta}_j$ and \hat{s}_j^2 does not generally hold for estimators other than the sample mean. For instance, Keane and Neal (2023) note that instrumental variables estimation produces correlated noise in estimates and standard errors.

equivalently represent θ_j^* as the maximizer of the posterior density in (66), which implies.

$$\begin{aligned} (\theta_1^*, \dots, \theta_J^*) &= \arg \max_{\theta_1, \dots, \theta_J} \sum_{j=1}^J \log p(\theta_j | \mathbf{D}, \mathbf{Y}) \\ &= \arg \max_{\theta_1, \dots, \theta_J} \sum_{j=1}^J \sum_{i=1}^N D_{ij} \log \phi\left(\frac{Y_{ij} - \theta_j}{\sigma_y}\right) + \sum_{j=1}^J \log \phi\left(\frac{\theta_j}{\sigma_\theta}\right), \end{aligned} \quad (67)$$

where I have dropped constants that do not depend on the θ_j 's. The maximizer of the posterior density is called a *maximum a posteriori* (MAP) estimate. Plugging normal densities into equation (67) and simplifying yields.

$$\begin{aligned} (\theta_1^*, \dots, \theta_J^*) &= \arg \max_{\theta_1, \dots, \theta_J} - \sum_{j=1}^J \sum_{i=1}^N D_{ij} \frac{(Y_{ij} - \theta_j)^2}{2\sigma_y^2} - \sum_{j=1}^J \frac{\theta_j^2}{2\sigma_\theta^2} \\ &= \arg \min_{\theta_1, \dots, \theta_J} \sum_{j=1}^J \sum_{i=1}^N D_{ij} (Y_{ij} - \theta_j)^2 + \lambda h(\theta_1, \dots, \theta_J), \end{aligned} \quad (68)$$

where $\lambda = \sigma_y^2 / \sigma_\theta^2$ and $h(\theta_1, \dots, \theta_J) = \sum_j \theta_j^2$.

Equation (68) shows that the linear shrinkage posteriors θ_j^* solve a regularized least squares problem with an L2 (quadratic) penalty function $h(\cdot)$, a simple ML procedure known as *ridge regression*. This implies we can equivalently think of ridge regression as a Bayesian procedure based on an independent mean-zero normal prior over the parameters. An EB version of this procedure plugs in an estimate of the prior, which means using the data to choose λ . The EB posterior means $\hat{\theta}_j^*$ are therefore ridge regression estimates based on a data-dependent value of the ridge penalty.

ML regularization procedures can often be reinterpreted through an EB lens, with the specific form of regularization determined by choices of prior distribution and loss function. This EB view helps to clarify the implicit distributional assumptions and objectives underlying ML procedures. For example, if we choose a Laplace (double-exponential) prior distribution rather than a normal distribution for G , an analogous derivation yields a MAP estimator that replaces the L2 penalty in equation (68) with an L1 (absolute value) penalty $h(\theta_1, \dots, \theta_J) = \sum_j |\theta_j|$. Regularized least squares with an L1 penalty is another basic ML procedure called the *least absolute shrinkage and selection operator* (lasso; Tibshirani, 1996). The use of lasso can therefore be justified by a choice of Laplace prior combined with a choice to report MAP estimates rather than posterior means.¹⁶ Abadie and Kasy (2019) consider the risk properties of ridge, lasso, and other common ML procedures in an EB framework.

¹⁶ Tibshirani (1996) noted this Bayesian interpretation when first introducing the lasso.

2.8 Linear shrinkage application: school value-added in Boston

The basic empirical Bayes recipe is illustrated here by estimating school value-added in Boston based on data from [Angrist et al. \(2017\)](#). I focus on 2014 math value-added estimates for 46 Boston middle schools. Thirty of these 46 schools operate within the Boston Public Schools (BPS) district, while the remaining 16 are charter schools, which are publicly-funded schools that operate outside BPS and have more freedom than traditional public schools to set curricula and make staffing decisions. The charters in this sample mostly follow the “No Excuses” educational model, a package of practices that has been shown to generate large achievement gains for students in Boston and elsewhere ([Abdulkadiroğlu et al., 2011](#); [Angrist et al., 2013](#); [Dobbie and Fryer, 2013](#); [Walters, 2018](#)). Further details on the characteristics of Boston schools and students are available in [Angrist et al. \(2017\)](#).

[Fig. 1](#) summarizes the three steps of the empirical Bayes recipe for Boston middle schools. I implement step 1 with an OLS regression of sixth-grade math test scores on school indicators and controls as in [equation \(5\)](#). Math scores Y_i are standardized to have mean zero and standard deviation (σ) one in the Boston student population. The covariate vector X_i includes fifth-grade math and reading scores along with indicators for sex, race, free or reduced price lunch status, special education, and English language learner status. I center the school coefficient estimates to have mean zero so that each $\hat{\theta}_j$ is interpretable as an estimate of the effect of school j relative to the average school. Standard errors s_j are computed with the [White \(1980\)](#) heteroskedasticity-robust variance estimator. The open bars in panel A of [Fig. 1](#) display a histogram of the estimated $\hat{\theta}_j$'s, with blue bars representing traditional BPS schools and red bars representing charter schools. These value-added estimates range from roughly -0.4σ to 0.5σ with a standard deviation of 0.221σ .

Some of the variation in OLS VAM estimates in [Fig. 1](#) comes from statistical noise in the estimated $\hat{\theta}_j$'s. Step 2 of the empirical Bayes recipe adjusts for this noise to recover an estimate of the underlying distribution of school quality. The average s_j^2 in this sample is 0.010, which (using [equation \(9\)](#)) results in a bias-corrected standard deviation estimate equal to $\hat{\sigma}_{\theta} = \sqrt{0.221^2 - 0.010} = 0.197\sigma$. The black curve in panel A of [Fig. 1](#) plots a normal distribution with standard deviation $\hat{\sigma}_{\theta}$, which is the deconvolution estimate of G under the normal model for the mixing distribution in [\(7\)](#). The raw standard deviation of average test scores across Boston schools is 0.5σ , so the estimated value-added distribution implies that only about $(0.2/0.5)^2 \times 100 = 16\%$ of the observed variance in school performance is due to causal contributions of schools, with the remaining 84 % explained by selection bias.

The standard errors s_j vary across schools, so I next probe for dependence between effect sizes and precision as discussed in [Section 2.6](#). This investigation suggests little relationship between school value-added and sampling variance. A regression of $\hat{\theta}_j$ on $\log s_j$ yields a slope coefficient of 0.246 with a

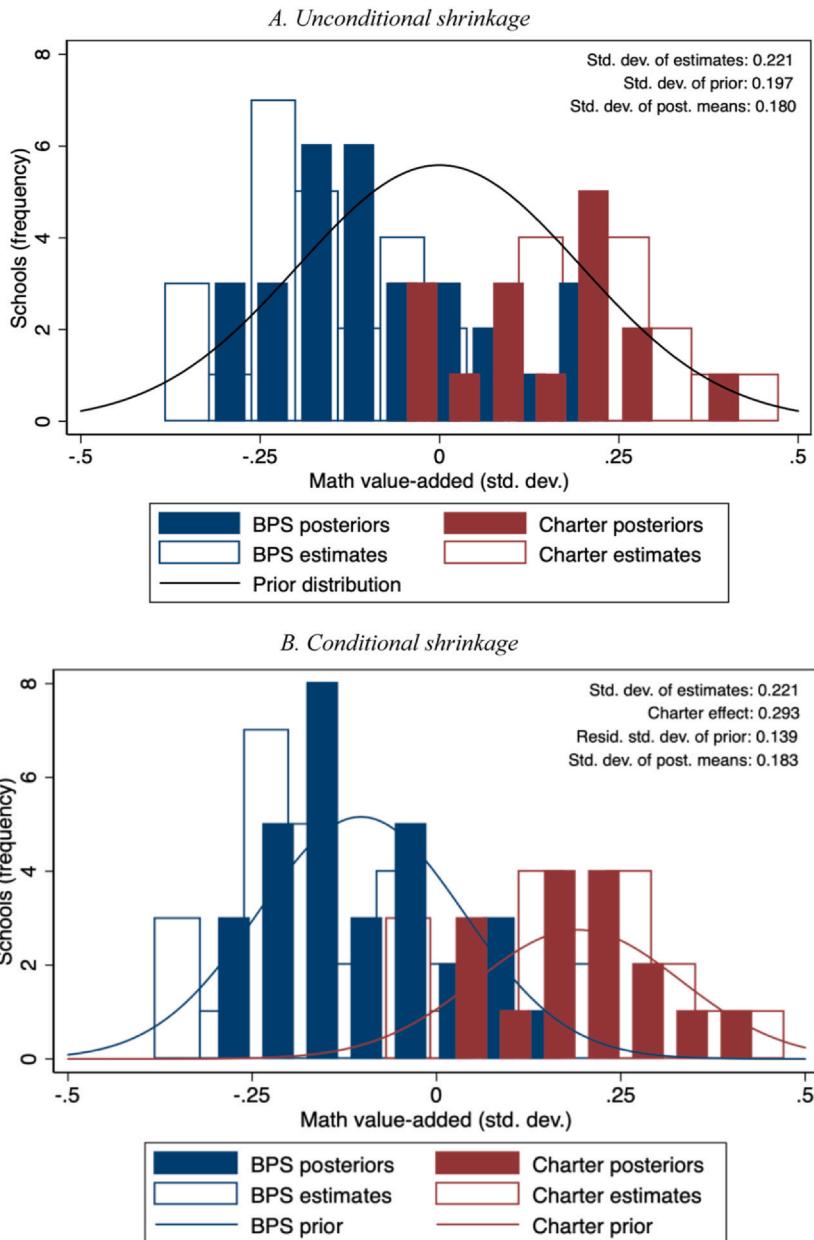


FIG. 1 Empirical Bayes estimates of school value-added in Boston. Notes: This figure displays estimates of value-added for 46 Boston middle schools. Outcomes are sixthgrade math scores in 2014 scaled to have mean zero and standard deviation one among all Boston students. Estimates come from regressions of math scores on school indicators with controls for fifth grade math and reading scores, sex, race, subsidized lunch, special education, and English language learner status.

robust standard error of 0.231, while a regression of $\hat{\theta}_j^2 - s_j^2$ on $\log s_j$ produces a slope coefficient of 0.012 with a standard error 0.047. In view of these weak relationships between estimates and standard errors, I maintain an independent prior in step 3 of the EB recipe and form linear shrinkage posterior means based on equation (15) for each school.

The resulting linear shrinkage estimates are plotted in the solid histogram in panel A of Fig. 1. As expected based on the discussion in Section 2.3, the standard deviation of shrunk posteriors (0.180σ) is smaller than the standard deviation of the prior distribution (0.197σ), which is in turn smaller than the standard deviation of the unadjusted VAM estimates (0.221σ). To understand the gains from shrinkage, note that MSE for $\hat{\theta}_j$ equals the average squared standard error (0.010), while MSE for the posterior mean equals the difference between the variance of the mixing distribution and the variance of posterior means.¹⁷ This comparison indicates that shrinkage reduces aggregate MSE for school value-added estimates by $(1 - [(0.197^2 - 0.180^2)/0.010]) \times 100 = 36\%$.

The upper tail of the estimated value-added distribution in panel A of Fig. 1 is disproportionately composed of charter schools. This suggests that incorporating differences in effectiveness across school sectors may result in improved estimates of school quality. I account for school sector with a conditional prior of the form of (38), allowing a charter-sector location shift in the mean of the mixing distribution and assuming a common within-sector variance σ_r^2 .

Following the logic of Section 2.3, I estimate the charter sector effect by regressing the unbiased $\hat{\theta}_j$'s on a charter indicator, which yields a coefficient of 0.293σ . This implies that attending an average charter school boosts achievement by nearly one-third of a standard deviation relative to an average BPS school. The standard deviation of residuals from this regression is 0.194σ , which generates a deconvolution estimate of the residual variance in school quality equal to $\hat{\sigma}_r = \sqrt{0.194^2 - 0.010} = 0.139\sigma$. The smaller value

School coefficients are centered to have mean zero across all Boston schools. Open histograms plot the distribution of raw value-added estimates with Boston Public Schools (BPS) schools in blue and charter schools in red. Solid curves plot estimated priors based on normal models for the mixing distribution. The prior standard deviation in panel A is calculated by subtracting the average squared standard error from the sample variance of estimates and taking the square root. The charter effect in panel B comes from a regression of value-added estimates on a charter indicator, and the residual standard deviation of the prior is calculated by subtracting the average squared standard error from the raw residual variance then taking the square root. Solid histograms plot linear shrinkage posterior means constructed using the estimated mixing distributions as priors.

¹⁷ By the law of total variance we have $E[(\theta_j - E[\theta_j|\hat{\theta}_j, s_j])^2] = Var(\theta_j) - Var(E[\theta_j|\hat{\theta}_j, s_j])$.

of $\hat{\sigma}_r$ relative to the unconditional $\hat{\sigma}_\theta$ reveals that charter status explains half of the variation in effectiveness across Boston schools: the implied R-squared from a regression of unobserved school quality θ_j on a charter indicator is $1 - (0.139/0.197)^2 = 0.502$. Linear shrinkage posteriors incorporating charter status shrink the $\hat{\theta}_j$'s toward a sector-specific mean using equation (39).

Panel B of Fig. 1 summarizes this conditional EB procedure by plotting the histogram of unadjusted VAM estimates, separate estimated prior distributions for BPS and charter schools, and a histogram of the resulting EB posteriors. A comparison to panel A illustrates how conditional shrinkage treats schools differently based on charter sector status. In panel A, for example, shrinkage toward the mean has little effect on charter schools with unbiased $\hat{\theta}_j$'s near zero since these schools are already estimated to be about average. The shrinkage in panel B accounts for the fact that such schools come from a high-performing sector and pulls estimates for these schools up above zero. Incorporating charter status into the shrinkage procedure yields further improvements in mean squared error, reducing aggregate MSE by an estimated 47 % relative to the unbiased $\hat{\theta}_j$'s.

3 Non-parametric empirical Bayes

This section extends the EB framework by relaxing the normality and independence assumptions maintained for much of Section 2. I first discuss methods for estimating the variance of parameters across units under minimal assumptions, building on the variance component estimation framework of Kline et al. (2020). I then proceed to cover non-parametric estimation of priors and posteriors, partial identification, and other important special cases of non-parametric EB methods. The section concludes with an application of non-parametric EB to a labor market correspondence experiment studied by Kline et al. (2022) and Kline et al. (2024).

3.1 Bias-corrected variance estimation

As before, consider a grouped data structure with J groups, each associated with a group-specific parameter θ_j . Collect these parameters in the $J \times 1$ vector $\Theta = (\theta_1, \dots, \theta_J)'$. Suppose we can form an unbiased estimate $\hat{\theta}_j$ of each group's parameter, and collect these estimates in the vector $\hat{\Theta} = (\hat{\theta}_1, \dots, \hat{\theta}_J)'$. A $J \times J$ matrix V describes the sampling variance of $\hat{\Theta}$. This matrix has the squared standard errors s_j^2 of the individual $\hat{\theta}_j$'s along its diagonal. Unlike the simpler framework of Section 2.1, I do not assume the noise in $\hat{\theta}_j$ is normally distributed, and I do not assume the $\hat{\theta}_j$'s are independent (i.e. the off-diagonal elements of V may not be zero). Moreover, I assume the elements of V are unknown, but an unbiased variance estimator \hat{V} is available. This setup is formalized as:

$$E[\hat{\Theta}|\Theta, V] = \Theta, \quad (69)$$

$$E[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)'|\Theta, V] = V, \quad (70)$$

$$E[\hat{V}|\Theta, V] = V. \quad (71)$$

The conditioning on (Θ, V) in (69)–(71) emphasizes that these objects are treated as fixed parameters rather than random variables for now.

Suppose we are interested in a quadratic form in the vector Θ , given by:

$$Q(\Theta, A) = \Theta' A \Theta, \quad (72)$$

where A is a known non-random weighting matrix. An important special case uses the matrix

$$A_0 = \frac{1}{J-1} \left(I_J - \frac{1}{J} \mathbf{1}_J \mathbf{1}'_J \right), \quad (73)$$

where I_J is the $J \times J$ identity matrix and $\mathbf{1}_J$ is a $J \times 1$ vector of 1's. With this choice of weighting matrix, the quadratic form $Q(\Theta, A_0)$ is the sample variance of θ_j 's:

$$Q(\Theta, A_0) = \frac{1}{J-1} \sum_{j=1}^J (\theta_j - \bar{\theta})^2, \quad (74)$$

with $\bar{\theta} = J^{-1} \sum_j \theta_j$. Parameter $Q(\Theta, A_0)$ summarizes variation in the unknown θ_j 's across the finite set of J observed groups. The general quadratic form in (72) also nests other useful special cases, such as covariances between value-added parameters across multiple outcomes.¹⁸

An obvious starting point for estimating $Q(\Theta, A)$ is a plug-in estimator $Q(\hat{\Theta}, A)$, which substitutes the unbiased estimate $\hat{\Theta}$ for the true Θ in equation (72). This estimator is biased due to the non-linearity of the quadratic form $Q(\cdot)$. Specifically, we have

$$\begin{aligned} E[Q(\hat{\Theta}, A)|\Theta, V] &= E[\hat{\Theta}' A \hat{\Theta}|\Theta, V] \\ &= \Theta' A \Theta + E[(\hat{\Theta} - \Theta)' A (\hat{\Theta} - \Theta)|\Theta, V] \\ &= Q(\Theta, A) + \text{tr}(AV), \end{aligned} \quad (75)$$

where the last equality uses standard properties of the trace operator.¹⁹ Equation (75) formalizes the intuition from Section 2.1 that noise inflates the sample

¹⁸ In the multivariate value-added framework of Section 2.4, collect value-added parameters for all schools and outcomes in the $JK \times 1$ vector $\Theta = (\theta_{11}, \dots, \theta_{J1}, \theta_{12}, \dots, \theta_{JK})'$, and let A_{km} denote a $JK \times JK$ matrix with A_{00} at its $k-m$ th $J \times J$ block and zeros elsewhere. Then $Q(\Theta, A_{km}) = (J-1)^{-1} \sum_j (\theta_{jk} - \bar{\theta}_k)(\theta_{jm} - \bar{\theta}_m)$ with $\bar{\theta}_k = J^{-1} \sum_j \theta_{jk}$.

¹⁹ A scalar equals its trace, trace is a linear operator, and a trace is invariant to cyclic permutations.

variance of estimated $\hat{\theta}_j$'s relative to the variability of the true θ_j 's. With the weighting matrix in (73) and a diagonal V , the bias in the plug-in variance estimate is given by $tr(A_0 V) = J^{-1} \sum_j s_j^2$.

The bias in the plug-in estimator can be corrected using the variance estimate \hat{V} . The result in equation (75) motivates a bias-corrected estimator of the form:

$$\hat{Q}_{BC} = Q(\hat{\Theta}, A) - tr(A\hat{V}). \quad (76)$$

When \hat{V} is unbiased for V , the second term of (76) is an unbiased estimator of the bias term in (75). This implies \hat{Q}_{BC} is an unbiased estimator of $Q(\Theta, A)$:

$$\begin{aligned} E[\hat{Q}_{BC} | \Theta, V] &= E[Q(\hat{\Theta}, A) | \Theta, V] - E[tr(A\hat{V}) | \Theta, V] \\ &= Q(\Theta, A) + tr(AV) - tr(AE[\hat{V} | \Theta, V]) \\ &= Q(\Theta, A). \end{aligned} \quad (77)$$

This bias correction approach may be implemented with various choices of the variance estimate \hat{V} . The properties of the resulting \hat{Q}_{BC} will depend on this choice as well as the variance structure of the underlying microdata. For the case where $\hat{\Theta}$ is a vector of linear regression coefficients, Andrews et al. (2008) consider bias-correction with a \hat{V} that is unbiased with homoskedasticity but biased and inconsistent with heteroskedasticity. A \hat{V} formed based on conventional White (1980) heteroskedasticity-robust (HC0) standard errors is consistent with heteroskedasticity but biased in finite samples. The HC2 robust variance modification proposed by MacKinnon and White (1985) is unbiased when the microdata are homoskedastic and biased but consistent with heteroskedasticity. Kline et al. (2020) propose a leave-out variance estimator that yields a finite-sample unbiased \hat{V} with arbitrary heteroskedasticity. This property makes the Kline et al. (2020) estimator an appealing choice for bias-correction in value-added studies with few observations per unit. In cases where groups are large enough for the standard errors of the individual $\hat{\theta}_j$'s to be accurate, alternative heteroskedasticity-robust variance matrices are likely to produce similar results.

3.1.1 Unbiased estimation of the mixing variance

Equation (77) establishes that we can use $\hat{\Theta}$ and \hat{V} to construct an unbiased estimate of any quadratic form $Q(\Theta, A)$, treating Θ and V as unknown fixed parameters. If we instead view the θ_j 's as random effects drawn from a mixing distribution, the law of iterated expectations implies \hat{Q}_{BC} is an unbiased estimate of the expectation of $Q(\Theta, A)$. We have

$$E[\hat{Q}_{BC}] = E[E[\hat{Q}_{BC} | \Theta, V]] = E[Q(\Theta, A)], \quad (78)$$

where the outer expectations treat Θ and V as random. Importantly, this implies that in the random effects model (6) with θ_j 's drawn independently from G , we can form an unbiased estimate of the mixing variance as:

$$\hat{\sigma}_\theta^2 = \hat{\Theta}' A_0 \hat{\Theta} - \text{tr}(A_0 \hat{V}). \quad (79)$$

This estimator is unbiased for σ_θ^2 because

$$\begin{aligned} E[\hat{\Theta}' A_0 \hat{\Theta} - \text{tr}(A_0 \hat{V})] &= E[Q(\Theta, A_0)] \\ &= E\left[\frac{1}{J-1} \sum_{j=1}^J (\theta_j - \bar{\theta})^2\right] \\ &= \sigma_\theta^2, \end{aligned} \quad (80)$$

where the last equality follows from the fact that the sample variance of θ_j 's is an unbiased estimate of the population variance. When \hat{V} is a diagonal matrix composed of unbiased squared standard error estimates \hat{s}_j^2 , the bias-corrected mixing variance estimator simplifies to $\hat{\sigma}_\theta^2 = (J-1)^{-1} \sum_j [(\hat{\theta}_j - \hat{\mu}_\theta)^2 - J^{-1}(J-1)\hat{s}_j^2]$. This estimator modifies equation (9) with degrees-of-freedom corrections to yield a finite-sample unbiased estimate of σ_θ^2 .

In summary, subtracting off an average squared standard error as in equation (9) is a simple and transparent starting point for bias-corrected mixing variance estimation. Using the more comprehensive bias-correction formula in equation (79) is likely to make a difference when J is small and when the off-diagonal elements of V are large. In a value-added regression like equation (5) with independent student observations, correlation in the $\hat{\theta}_j$'s across schools comes from error in the estimated control coefficient $\hat{\gamma}$.²⁰ This implies the off-diagonal elements of V can be safely ignored when the control coefficients are estimated very precisely, but may be important in scenarios with many fixed effects or other high-dimensional controls. Other approaches to value-added estimation, such as “movers designs” leveraging individuals switching between groups, may also generate important correlations in noise across units, leading to a non-diagonal V (Bonhomme and Denis, 2024). The full bias correction in (79) extracts the signal variance of value-added from correlated noise in such cases. In applications where correlated noise is a potential concern, a

²⁰ As noted in Section 2.3, the $\hat{\theta}_j$'s are school means of the long regression residual $Y_i - X'_i \hat{\gamma}$, which are uncorrelated across schools if there is negligible noise in $\hat{\gamma}$. The nature of correlation in estimates will also depend on how the reference category in a value-added regression is defined. Equation (5) includes a dummy for every school and no constant term, which makes $\hat{\theta}_j$ an estimate of an average covariate-adjusted outcome level at school j . If we instead include a constant and omit a dummy for school 1, the resulting value-added estimates for schools 2 through J equal $\hat{\Delta}_j = \hat{\theta}_j - \hat{\theta}_1$, which are likely to be correlated since these are differences relative to the same base school.

comparison of mixing variance estimates based on equations (9) and (79) is a useful robustness check.

Example 4: Worker and firm effects

A prominent application of bias-corrected variance estimation comes from two-way fixed effects (TWFE) models for worker and firm effects on earnings. Starting with [Abowd et al. \(1999\)](#), a large literature studies linear regressions of the form:

$$\log Y_{it} = \alpha_i + \psi_{\mathbf{J}(i,t)} + X'_{it}\beta + \epsilon_{it}, \quad (81)$$

where Y_{it} is earnings for worker i in year t and X_{it} is a vector of time-varying controls. The function $\mathbf{J}(i, t)$ returns the identity of the firm employing worker i in year t . The worker effect α_i represents a permanent component of earnings specific to worker i that is constant across employers, while the firm effect ψ_j reflects an effect of working for firm j that is constant across workers and over time. Under an exogenous mobility assumption the firm effects in [equation \(81\)](#) capture causal effects of particular employers on pay, which may reflect factors such as market power, compensating differentials, wage setting policies, or rent sharing ([Card et al., 2018](#)). Analyses of such additive worker and firm effect models include [Gruetter and Lalive \(2009\)](#); [Card et al. \(2013\)](#); [Card et al. \(2015\)](#); [Song et al. \(2018\)](#); [Lachowska et al. \(2023a\)](#), and [Lachowska et al. \(2023b\)](#).

The parameters of [equation \(81\)](#) are identified based on movements of workers across firms, so TWFE estimation must be restricted to a connected set of employers linked by worker mobility. Even within this set there may be few workers linking some groups of firms, yielding noisy estimated worker and firm effects $\hat{\psi}_j$ and $\hat{\alpha}_i$. The errors in these estimates are also likely to be correlated: overestimating a worker's effect will tend to lead to underestimating the effect of his or her firm, resulting in a negative sampling covariance between $\hat{\alpha}_i$ and $\hat{\psi}_{\mathbf{J}(i,t)}$. The influence of this noise on the variation in estimated worker and firm effects is commonly referred to as "limited mobility bias" ([Andrews et al., 2008](#)), a special case of the plug-in bias illustrated in [equation \(75\)](#).

The variances and covariances of the worker and firm effects from [equation \(81\)](#) can be written as special cases of the quadratic form in (72). [Bonhomme et al. \(2023\)](#) compare estimates of these variance components using multiple strategies for bias correction including the "homoskedasticity-only" correction of [Andrews et al. \(2008\)](#), the heteroskedasticity-robust bias correction of [Kline et al. \(2020\)](#), and random effects methods building on [Woodcock \(2008\)](#) and [Bonhomme et al. \(2019\)](#). Their results for several countries show that bias-correction substantially reduces the estimated magnitude of variation in firm effects and often flips the sign of the estimated covariance between α_i and $\psi_{\mathbf{J}(i,t)}$ from negative to positive. They also find that alternative methods for bias-correction tend to produce similar results. Adopting an EB view, we can think of the parameters of [equation \(81\)](#) as (correlated) random effects drawn

from a mixing distribution describing worker and firm heterogeneity, and interpret bias-corrected variance components as estimates of the second moments of this distribution.

3.2 Non-parametric priors and posteriors

The preceding section shows that we can estimate the variance of G under minimal assumptions. If we are content to quantify heterogeneity across units with the mixing variance and form linear shrinkage posteriors with low mean squared error, estimating the mean and variance of G may be enough. In some cases, however, it is useful to develop a more complete picture of the value-added distribution. To flexibly estimate the full mixing distribution G , I return to a hierarchical random effects setup with mutually independent normally distributed estimates and effect sizes independent of standard errors:

$$\hat{\theta}_j | \theta_j, s_j \sim N(\theta_j, s_j^2), \quad (82)$$

$$\theta_j | s_j \sim G. \quad (83)$$

The mixing distribution is non-parametrically identified in this model ([Kotlarski, 1967](#); [Evdokimov and White, 2012](#)), which motivates estimators that avoid imposing a functional form for G . I consider two approaches to flexible estimation of G : a non-parametric maximum likelihood estimator (NPMLE), and a log-spline deconvolution estimator proposed by [Efron \(2016\)](#).

3.2.1 Non-parametric maximum likelihood

Non-parametric maximum likelihood is a classic approach to flexibly estimating distributions of unobserved heterogeneity. NPMLE was outlined in an abstract by [Robbins \(1950\)](#) and developed in detail by [Kiefer and Wolfowitz \(1956\)](#). Within labor economics, [Heckman and Singer \(1984\)](#) applied NPMLE to estimate heterogeneity distributions in mixed proportional hazards models. The NPMLE estimator picks \hat{G} to maximize the likelihood of the observed data over all possible mixing distributions. This estimator is given by:

$$\hat{G}_{NPMLE} = \arg \max_{G \in \mathcal{G}} \sum_{j=1}^J \log \left(\int \frac{1}{s_j} \phi \left(\frac{\hat{\theta}_j - \theta}{s_j} \right) dG(\theta) \right), \quad (84)$$

where \mathcal{G} is the set of all cumulative distribution functions. The \hat{G}_{NPMLE} that solves this problem is a discrete distribution with at most J mass points. While maximizing over all possible distribution functions may appear to be a formidable empirical task, [Koenker and Mizera \(2014\)](#) outline computationally-efficient procedures that quickly approximate the NPMLE using convex optimization methods.²¹ [Gilraine et al. \(2020\)](#) apply these methods to non-parametrically estimate distributions of teacher value-added.

²¹ The **REBayes** R package implements these methods ([Koenker and Gu, 2017](#)).

NPMLE imposes no restrictions on the mixing distribution, and EB decision rules using \hat{G}_{NPMLE} as prior have been shown to provide a close approximation to the decisions of an oracle who knows G (Jiang and Zhang, 2009; Jiang, 2020). However, the discrete distribution function produced by NPMLE can be unweildy in some applications. Posterior distributions using \hat{G}_{NPMLE} as prior will also be discrete, which may be awkward in settings where exact ties of θ_j 's are implausible. Koenker (2020) notes that this discreteness can lead to narrow EB credible intervals (differences in posterior quantiles) with poor frequentist coverage.²² These issues make it attractive to consider adding some smoothness restrictions in non-parametric deconvolution.

3.2.2 Log-spline deconvolution

Efron (2016) proposes a flexible deconvolution approach that approximates the mixing distribution with a smooth log density parameterized by a natural cubic spline. For a grid of M support points $(\bar{\theta}_1, \dots, \bar{\theta}_M)$, suppose the probability mass at point m is given by:

$$g_m(\alpha) = \exp\left(S'_m \alpha - \log\left(\sum_{k=1}^M \exp(S'_k \alpha)\right)\right), \quad (85)$$

where $S_m = S(\bar{\theta}_m)$ is a $K \times 1$ vector of natural cubic spline basis functions with K knots and α is a $K \times 1$ parameter vector. The parameters of the model are estimated by penalized maximum likelihood as follows:

$$\hat{\alpha} = \arg \max_{\alpha} \sum_{j=1}^J \log\left(\sum_{m=1}^M g_m(\alpha) \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \bar{\theta}_m}{s_j}\right)\right) - \lambda \sqrt{\alpha' \alpha}, \quad (86)$$

where $\lambda \geq 0$ is a tuning parameter. The resulting log-spline deconvolution estimate of G is $\hat{G}_{LS}(\theta) = \sum_m 1\{\bar{\theta}_m \leq \theta\} g_m(\hat{\alpha})$.

The Efron (2016) log spline deconvolution estimator is straightforward to compute and (when M is large) yields a smoother mixing distribution estimate than NPMLE.²³ These advantages make it an appealing option for flexible deconvolution in practice. However, the estimator in (86) requires choosing several tuning parameters including the number of spline knots K ; the support limits, spacing, and number of the support points $\bar{\theta}_m$; and the penalization parameter λ . By the logic of Section 2.7, the penalty term $\lambda \sqrt{\alpha' \alpha}$ may be interpreted as a second-level prior that pushes the estimated mixing distribution toward a uniform distribution to manage overfitting. Kline et al. (2022) suggest

²² Armstrong et al. (2022) propose EB confidence intervals with average coverage guarantees regardless of the form of G .

²³ The **deconvolveR** R package provides software for log spline deconvolution (Narasimhan and Efron, 2020).

choosing λ so that the variance of the resulting \hat{G}_{LS} matches a bias-corrected variance estimate based on the methods of [Section 3.1](#). Tuning \hat{G}_{LS} to match low-order moment estimates in this way ensures that the deconvolved mixing distribution reproduces basic features of G estimated with simpler methods ([Efron and Tibshirani, 1996](#)).

3.2.3 Non-parametric posteriors

Non-parametric EB shrinkage uses an NPMLE or log-spline estimate \hat{G} when forming EB posterior distributions $\mathcal{P}(\theta|\hat{\theta}_j, s_j; \hat{G})$ for each unit. The resulting distribution can be used to compute any posterior feature of interest. The posterior mean of an oracle who knows G is

$$\hat{\theta}_j^* = \int \theta d\mathcal{P}(\theta|\hat{\theta}_j, s_j; G), \quad (87)$$

with the posterior \mathcal{P} defined as in [\(10\)](#). A non-parametric estimate of this quantity derived from log-spline deconvolution is:

$$\hat{\theta}_j^* = \int \theta d\mathcal{P}(\theta|\hat{\theta}_j, s_j; \hat{G}_{LS}) = \frac{\sum_{m=1}^M \bar{\theta}_m \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \bar{\theta}_m}{s_j}\right) g_m(\hat{\alpha})}{\sum_{m=1}^M \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \bar{\theta}_m}{s_j}\right) g_m(\hat{\alpha})}, \quad (88)$$

where $\hat{\alpha}$ is calculated according to [equation \(86\)](#).

Changing notation relative to [Section 2.1](#), we can contrast the non-parametric shrinkage estimate $\hat{\theta}_j^*$ with a linear shrinkage estimate given by:

$$\hat{\theta}_j^{lin} = \left(\frac{\hat{\sigma}_\theta^2}{\hat{\sigma}_\theta^2 + s_j^2} \right) \hat{\theta}_j + \left(\frac{s_j^2}{\hat{\sigma}_\theta^2 + s_j^2} \right) \hat{\mu}_\theta, \quad (89)$$

where $\hat{\mu}_\theta$ and $\hat{\sigma}_\theta^2$ are estimates of the mean and variance of G . These hyper-parameters can be calculated either using simple mean and variance estimators as in [Sections 2.1](#) and [3.1](#), or based on the first two moments of a non-parametric \hat{G} .

The choice between $\hat{\theta}_j^{lin}$ and $\hat{\theta}_j^*$ mirrors the usual tradeoff between the robustness and simplicity of linear estimators versus potential efficiency gains from non-linear estimators. If the mixing distribution is far from a normal distribution the non-parametric approach will leverage the higher moments of G to produce a richer posterior distribution, reducing mean squared error relative to linear shrinkage. On the other hand, non-parametric deconvolution requires estimating extra parameters, and the higher moments of G may be poorly estimated. Linear shrinkage provides a minimum MSE linear approximation to the posterior mean without the need to estimate these higher moments. In applications of non-parametric EB for value-added estimation, comparing non-parametric and linear shrinkage posteriors is a useful reality check on the output of non-linear shrinkage procedures.

3.2.4 Incorporating precision-dependence

The general mixing distribution model in [equation \(83\)](#) assumes independence between effect sizes and standard errors. The same precision-dependence issues discussed in [Section 2.6](#) apply to non-parametric estimation of this model. Like the precision-weighted mean and variance estimators in [equations \(53\)](#) and [\(54\)](#), the maximum likelihood estimators in [\(84\)](#) and [\(86\)](#) will leverage all the restrictions implied by independence of θ_j and s_j to increase efficiency. Imposing these restrictions will enhance precision if they are satisfied but may compromise consistency of the resulting \hat{G} if not. The performance of non-parametric EB posteriors may also suffer if the prior erroneously rules out precision-dependence.

Paralleling the parametric approach of [Section 2.6](#), we can build precision-dependence into non-parametric EB by estimating a model of the relationship between effect sizes and standard errors, then deconvolving residuals. A non-parametric extension of the conditional location/scale model in [equation \(55\)](#) is given by:

$$\theta_j = \mu(s_j) + \sigma(s_j)r_j, \quad r_j|s_j \sim G_r, \quad (90)$$

with $E[r_j|s_j] = 0$ and $Var(r_j|s_j) = 1$. [Chen \(2023\)](#) proposes a CLOSE-NPMLE estimator that estimates the conditional mean and variance functions $\mu(s_j)$ and $\sigma(s_j)$ by non-parametric (local linear) regression, then applies NPMLE to the resulting residuals to estimate G_r . His theoretical results establish that this estimator yields EB decision rules that approximate the decisions of an oracle who knows the full conditional mixing distribution, even if the conditional location/scale model is misspecified. A tractable semi-parametric alternative (sacrificing some flexibility in the first step) is to specify functional forms for $\mu(s_j)$ and $\sigma(s_j)$, estimate the unknown parameters of these functions, then non-parametrically deconvolve residuals with NPMLE or log-spline deconvolution. I provide an example of such an approach in [Section 3.7](#).

3.3 Partial identification

The normal noise model [\(82\)](#) is usually justified as an asymptotic approximation with a growing number of observations in each group. With few observations per group the distribution of the noise in $\hat{\theta}_j$ will depend on the distribution of the underlying microdata Y_i , and the mixing distribution G may not be non-parametrically identified. However, it may still be possible to recover useful features of the prior and posterior distributions with a partial identification approach. I illustrate such an approach through an example drawn from [Kline and Walters' \(2021\)](#) study of job-specific employment discrimination.

Example 5: Job-level employment discrimination

[Kline and Walters \(2021\)](#) analyze data from resume correspondence experiments sending fictitious applications to real job vacancies ([Bertrand and](#)

Mullainathan, 2004; Arceo-Gomez and Campos-Vasquez, 2014; Nunley et al., 2015). To manipulate employers' perceptions of race, resumes in these experiments are assigned racially-distinctive names. Suppose each vacancy $j \in \{1, \dots, J\}$ in a study receives L applications, with race assignment stratified so that L_w have distinctively-white names and $L_b = L - L_w$ have distinctively-Black names. For example, Bertrand and Mullainathan (2004) sent four applications per job with two in each racial group, so $L_w = L_b = 2$.

Let $D_i \in \{1, \dots, J\}$ denote the vacancy that received application i , let $R_i \in \{w, b\}$ represent the racial valence of the name assigned to this application (white or Black), and let $Y_i \in \{0, 1\}$ denote an indicator equal to one if the application was called back by the employer. Assume callback outcomes are generated by Bernoulli trials with stable job-by-race callback probabilities:

$$Y_i | D_i = j, R_i = r \sim \text{Bernoulli}(p_{jr}). \quad (91)$$

The unknown unit-specific parameter in this case is the 2×1 vector $\theta_j = (p_{jw}, p_{jb})'$. As in Example 3, this model of independent Bernoulli trials implies the success counts $C_{jr} = \sum_{i=1}^N \mathbb{1}\{D_i = j\} \mathbb{1}\{R_i = r\} Y_i$ follow $\text{Bin}(L_r, p_{jr})$ distributions for $r \in \{w, b\}$.

Adopting an empirical Bayes view, suppose the two job-specific success probabilities are drawn randomly from a bivariate mixing distribution G :

$$(p_{jw}, p_{jb})' \sim G. \quad (92)$$

This mixing distribution describes heterogeneity in discrimination across jobs in the population. We would like to use the distribution of observed success counts to learn about G . However, it is clear that an asymptotic normal approximation does not apply with only two trials per group, so the assumptions underlying the deconvolution methods discussed in Section (3.2) do not hold.

What can we learn about the mixing distribution in this case? With binomial trials, the likelihood of a particular callback configuration (c_w, c_b) conditional on the callback probabilities at job j is:

$$\begin{aligned} f(c_w, c_b | p_{jw}, p_{jb}) &\equiv \Pr[C_{jw} = c_w, C_{jb} = c_b | p_{jw}, p_{jb}] \\ &= \binom{L_w}{c_w} \binom{L_b}{c_b} p_{jw}^{c_w} (1 - p_{jw})^{L_w - c_w} p_{jb}^{c_b} (1 - p_{jb})^{L_b - c_b}. \end{aligned} \quad (93)$$

Combined with the mixing distribution model in (92), equation (93) implies the share of jobs with this callback configuration is given by:

$$\begin{aligned} \bar{f}(c_w, c_b) &\equiv \Pr[C_{jw} = c_w, C_{jb} = c_b] \\ &= \int f(c_w, c_b | p_w, p_b) dG(p_w, p_b) \end{aligned}$$

$$= \binom{L_w}{c_w} \binom{L_b}{c_b} \sum_{k=0}^{L_w} \sum_{m=0}^{L_b} (-1)^{k+m} \binom{L_w - c_w}{k} \binom{L_b - c_b}{m} \mu(c_w + k, c_b + m), \quad (94)$$

where the function $\mu(x, y) = \int p_w^x p_b^y dG(p_w, p_b)$ describes non-central moments of the bivariate distribution G . Evaluating this expression for all observed values of (c_w, c_b) yields a linear system of the form $\bar{f} = B\mu_G$, which links a vector of callback frequencies \bar{f} to a vector of moments μ_G of the mixing distribution via a known invertible matrix B of binomial coefficients. We can then solve for $\mu_G = B^{-1}\bar{f}$ to recover a set of moments of G from the observed callback frequencies.

This argument shows that all moments of G involving powers of (p_w, p_b) up to (L_w, L_b) are identified. As a result, useful measures of heterogeneity in discrimination are identified even with few trials per job. For example, the variance of the white/Black gap in callback probabilities involves moments up to order two:

$$\text{Var}(p_{jw} - p_{jb}) = \mu(2, 0) - \mu(1, 0)^2 + \mu(0, 2) - \mu(0, 1)^2 - 2[\mu(1, 1) - \mu(1, 0)\mu(0, 1)]. \quad (95)$$

The moments in equation (95) are identified in an experiment that sends at least two applications per group (i.e. $\min\{L_w, L_b\} \geq 2$), as in Bertrand and Mullainathan (2004). Such an experiment can therefore be used to distinguish a scenario in which a positive average effect of white names comes from an equal advantage for white applicants at all jobs (in which case $\text{Var}(p_{jw} - p_{jb}) = 0$) from a scenario in which some jobs favor white applicants much more than others (in which case $\text{Var}(p_{jw} - p_{jb})$ is large). Intuitively, we should see more modest imbalances of $C_{jw} - C_{jb} = 1$ in the former case, while the latter case generates both more highly-imbalanced jobs with $(C_{jw}, C_{jb}) = (2, 0)$ and more instances of equal treatment with $C_{jw} = C_{jb}$. Kline and Walters (2021) apply this approach to document substantial variation in discrimination in several correspondence experiments.

In addition to characterizing heterogeneity in discrimination, the moments of G imply bounds on features of the prior and posterior distributions that are not point-identified. Consider jobs that call back both applicants with distinctively-white names and no applicants with distinctively-Black names in the Bertrand and Mullainathan (2004) study. How sure should we be that jobs with this seemingly-suspicious callback configuration are discriminating? The share of such jobs that have different callback rates for white and Black applicants is:

$$\Pr[p_{jw} \neq p_{jb} | C_{jw} = 2, C_{jb} = 0] = \frac{\int f(2, 0 | p_w, p_b) \mathbf{1}\{p_w \neq p_b\} dG(p_w, p_b)}{\bar{f}(2, 0)}. \quad (96)$$

The smallest value of this posterior probability that is consistent with the observed callback frequencies \bar{f} solves the problem

$$\min_{G \in \mathcal{G}} \frac{\int f(2, 0|p_w, p_b) 1\{p_w \neq p_b\} dG(p_w, p_b)}{\bar{f}(2, 0)} \quad \text{s.t. } \bar{f} = B\mu_G, \quad (97)$$

where \mathcal{G} is the set of all bivariate cumulative distribution functions with domain $[0,1]^2$.

[Equation \(97\)](#) defines an optimization problem that is linear in the probability mass function associated with G , which can be solved with linear programming techniques. [Kline and Walters \(2021\)](#) plug empirical estimates of the callback frequencies \bar{f} into this problem and optimize over discrete mixing distributions defined on a fine two-dimensional grid of callback probabilities. Their results show that experiments with few observations per job can generate informative posterior bounds. For instance, in the [Bertrand and Mullainathan \(2004\)](#) experiment, at least 72 % of jobs that call back two white applicants and no Black applicants must be discriminating, even using the most conservative mixing distribution consistent with the experimental data. The corresponding bound for a more recent experiment conducted by [Nunley et al. \(2015\)](#) is 85 %. The idea of using a maximally-conservative empirical prior to bound posterior probabilities in this way is closely related to empirical Bayes approaches to multiple testing, as I discuss next.

3.4 EB for multiple testing: large-scale inference

Non-parametric empirical Bayes methods are tightly connected with multiple testing problems that arise frequently in labor economics and other areas of applied work. [Efron \(2012\)](#) uses “large-scale inference” to refer to EB methods in this context. Consider a list of null hypotheses for each of J units, such as $H_0 : \theta_j = 0$. This collection of hypotheses might concern which subgroups are affected by an intervention, which schools have value-added below some quantile of the distribution (as in [Section 2.5](#)), or which jobs discriminate against racially-distinctive names (as in [Section 3.3](#)). Let $T_j = 1\{\theta_j = 0\}$ denote an indicator equal to 1 if the null is true for unit j . Suppose we conduct an independent test of H_0 for each unit to generate a collection of p -values p_j . For a rejection threshold \bar{p} , indicators $\delta_j = 1\{p_j \leq \bar{p}\}$ describe which null hypotheses are rejected.

Traditional hypothesis testing seeks to control the probability of type I error (size) for a single test. In other words, we limit the likelihood of a mistaken rejection when the null is true by adopting a decision rule such that $\Pr[\delta_j = 1 | T_j = 1] \leq \alpha$ for a tolerance α . With p -values that are uniformly distributed under the null hypothesis, this size is controlled by setting a rejection threshold of $\bar{p} = \alpha$. A concern with multiple tests is that applying this rule may lead to scenarios in which many rejected hypotheses are true. In the extreme case where all null

hypotheses are true, all rejected hypotheses will be true, but we will be very likely to reject some if \bar{p} is fixed and J is large. This motivates approaches that control alternative notions of aggregate error such as the family-wise error rate (FWER), which is the probability of at least one mistaken rejection: $FWER = \Pr[\sum_{j=1}^J T_j \delta_j \geq 1]$. A standard Bonferroni correction uses a modified rejection threshold $\bar{p} = \alpha/J$ which shrinks with J and thereby controls FWER at level α .

Approaches that control FWER (or generalizations such as k -FWER; [Lehmann and Romano, 2005](#)) are natural when each mistaken rejection is very costly. In settings with large numbers of tests, however, controlling the absolute number of mistakes is stringent, and it is often more natural to control the *rate* of mistakes. The *false discovery proportion* (FDP) is given by:

$$FDP = \begin{cases} \frac{\sum_{j=1}^J \delta_j T_j}{\sum_{j=1}^J \delta_j}, & \sum_{j=1}^J \delta_j > 0; \\ 0, & \sum_{j=1}^J \delta_j = 0. \end{cases} \quad (98)$$

FDP is the share of rejected hypotheses that are true when we reject at least one hypothesis, and is defined to be zero otherwise. [Benjamini and Hochberg \(1995\)](#) propose controlling the *false discovery rate* (FDR), which is the expectation of FDP:

$$FDR = E[FDP]. \quad (99)$$

FDR gives the expected share of true nulls among rejected hypotheses. With a test procedure that controls FDR, therefore, we should expect most rejected null hypotheses to be false.

3.4.1 An EB approach to FDR control

An empirical Bayes approach facilitates FDR control in a setting with many tests. If we adopt the random effects model (6), each true null indicator T_j is a function of a random draw from G . Under this model, the expected share of true nulls among rejected hypotheses (those with $p_j \leq \bar{p}$) is

$$\begin{aligned} FDR &= \Pr[T_j = 1 | \delta_j = 1] \\ &= \frac{\Pr[p_j \leq \bar{p} | T_j = 1] \Pr[T_j = 1]}{\Pr[p_j \leq \bar{p}]} \\ &= \frac{\bar{p} \pi_0}{F(\bar{p})}, \end{aligned} \quad (100)$$

where the second line uses Bayes' rule, the third uses that p -values are uniformly distributed under the null, $F(p) = \Pr[p_j \leq p]$ is the marginal CDF of p -values, and $\pi_0 = \int 1\{\theta = 0\} dG(\theta)$.

[Equation \(100\)](#) shows that controlling FDR requires choosing \bar{p} to limit $\bar{p}\pi_0/F(\bar{p})$. The CDF $F(\cdot)$ can be estimated using the observed p -value

distribution, e.g. with $\hat{F}(\bar{p}) = J^{-1} \sum_j 1\{p_j \leq \bar{p}\}$. The key unknown quantity in [equation \(100\)](#) is π_0 , the population share of true nulls. Importantly, π_0 is a feature of G , and therefore reflects an objective fact about the distribution of parameters in the population being studied. If $\pi_0 = 1$ all hypotheses are true, and any rejection is a mistake. When $\pi_0 = 0$ all rejected hypotheses are false, but so are all hypotheses that are not rejected. In between, π_0 provides the correct prior presumption that the null is true in this population.

[Benjamini and Hochberg \(1995\)](#) propose a conservative approach to FDR control that plugs $\pi_0 = 1$ into [equation \(100\)](#). This approach may still allow FDR control with a \bar{p} greater than zero if the p -values are concentrated toward the origin, so that $F(\bar{p}) \gg \bar{p}$. It is clear, however, that we can do better. It is logically inconsistent to assume $\pi_0 = 1$ while finding that $F(\bar{p}) > \bar{p}$. More generally, the probability π_0 is a feature of G , and we can learn about G via empirical Bayes deconvolution. This suggests leveraging the distribution of results across tests to discipline the choice of π_0 .

The mixing distribution G is non-parametrically identified in the normal noise model [\(82\)–\(83\)](#), which suggests π_0 may be point identified in some applications. However, it is standard to treat this probability as partially identified in the multiple testing context. Intuitively, it is difficult to empirically distinguish between G 's with mass points at exactly zero or only very close to zero. To bound π_0 , note that the marginal density of the p -value distribution at a point p can be written:

$$f(p) = \pi_0 + (1 - \pi_0)f_a(p), \quad (101)$$

where $f_a(p)$ is the density of p -values among false null hypotheses (those with $T_j = 0$). [Equation \(102\)](#) reveals that the observed p -value distribution is a mixture of a uniform density with weight π_0 and an alternative density $f_a(p)$ with weight $1 - \pi_0$ ([Efron et al., 2001](#)). While the exact form of $f_a(p)$ is usually unknown, this density cannot be negative, so the marginal p -value density at any point p provides an upper bound on π_0 :

$$f(p) = \pi_0 + (1 - \pi_0)f_a(p) \geq \pi_0 \quad \forall p. \quad (102)$$

A useful test should generate p -values that are concentrated toward zero when the null is false, so we expect $f_a(p)$ to be lowest (and the resulting bound on π_0 to be tightest) in the upper tail of the p -value distribution. [Storey \(2002\)](#) proposes to bound π_0 with the estimator

$$\hat{\pi}_0 = \frac{\sum_{j=1}^J 1\{p_j > b\}}{(1 - b)J}, \quad (103)$$

where the threshold $b \in [0, 1]$ is a tuning parameter. This amounts to assuming all p -values above b correspond to true nulls and using the share of units in this region to estimate the height of the null density. A higher b results in a tighter bound but noisier estimate. [Storey et al. \(2004\)](#) propose a bootstrap approach to

selecting b that balances this tradeoff to minimize MSE. A simple alternative is to select a value of b a priori since any b provides an upper bound on π_0 .

With an estimated upper bound $\hat{\pi}_0$ in hand, equation (100) gives an upper bound on FDR for any rejection cutoff \bar{p} . Evaluating this expression at each observed p -value yields a list of *q-values*, given by:

$$q_j = \frac{p_j \hat{\pi}_0}{\hat{F}(p_j)}. \quad (104)$$

The *q*-value is an empirical Bayes analogue of the *p*-value (Storey, 2003).²⁴ Rather than controlling $\Pr[\delta_j = 1 | T_j = 1]$ with the *p*-value, we borrow strength from the ensemble of tests to flip the conditioning and control $\Pr[T_j = 1 | \delta_j = 1]$ with the *q*-value. If we reject all hypotheses with *p*-values less than p_j , we should expect at most a share q_j of rejections to be mistakes.

3.5 Ranking problems

Empirical Bayes and related methods are increasingly used to build *report cards* that summarize estimates of quality with rankings or coarse grades. Examples include evaluations of K-12 schools, teachers, colleges, hospitals, doctors, and neighborhoods (Angrist et al., 2024b; Bergman and Hill, 2018; Chetty et al., 2017; Kolstad, 2013; Pope, 2009; Chetty and Hendren, 2018). Report cards of this sort lead naturally to comparisons of top and bottom performers (the “league table mentality” discussed by Gu and Koenker, 2023b). However, the estimators and decision rules I have considered so far aim for good average performance across units, not for accurate pairwise comparisons. Can we be sure that units assigned top grades in a value-added report card are actually among the best?

One approach to this question is to analyze relative rankings of units in a standard frequentist inference framework. Treating the θ_j 's as unknown fixed parameters, the rank of unit j is given by $rank_j = \sum_{k=1}^J 1\{\theta_j \leq \theta_k\}$. Mogstad et al. (2023) develop methods to construct confidence sets for any individual $rank_j$ as well as simultaneous confidence sets for the rankings of all J units. Andrews et al. (2023) propose tools for inference on the value-added of the highest-ranked unit, accounting for upward bias in the maximum estimate due to the ranking step.

An alternative strategy is to formalize the objective of a report card system in an EB compound decision framework. In some cases this makes clear that mis-ranking units may not be a problem. If our goal is to communicate reliable information on absolute quality, ranking units with an absolute performance metric such as an EB posterior mean may lead to good decisions, even if such a

²⁴The approach to *q*-value estimation outlined here is implemented in the **qvalue** R package (Storey, 2015).

ranking yields a high rate of mistakes in relative comparisons (i.e. a low correlation between reported rankings and the true $rank_j$'s). Such mistakes will be hardest to avoid when the θ_j 's are very close together, in which case misrankings may be of little consequence for outcomes. In other cases we may be specifically interested in getting the rankings right, which requires a loss function tailored to this goal.

Kline et al. (2024) use such a loss function to construct a report card summarizing firm-specific discrimination estimates.²⁵ Consider a decision-maker tasked with assigning a grade $\delta_j \in \{1, \dots, J\}$ to each of J units. Let $\delta = (\delta_1, \dots, \delta_J)'$ denote the vector of grades for all units, and let $\Theta = (\theta_1, \dots, \theta_J)'$ denote the vector of true parameters. Suppose the decisionmaker seeks to minimize the loss function

$$\begin{aligned} \mathcal{L}(\delta; \Theta) = & \binom{J}{2}^{-1} \sum_{j=2}^J \sum_{k=1}^j [1\{\theta_j > \theta_k\} 1\{\delta_j < \delta_k\} + 1\{\theta_j < \theta_k\} 1\{\delta_j > \delta_k\} \\ & - \lambda(1\{\theta_j > \theta_k\} 1\{\delta_j > \delta_k\} + 1\{\theta_j < \theta_k\} 1\{\delta_j < \delta_k\})]. \end{aligned} \quad (105)$$

This function assigns a loss of 1 for each pair of units that are mis-ordered (*discordances*), and a gain of $\lambda \in [0,1]$ for each pair that is correctly ordered (*concordances*). Assigning two units the same grade ($\delta_j = \delta_k$) guarantees a loss of zero for the pair.

By rearranging terms in equation (105), we can represent this loss function as

$$\mathcal{L}(\delta; \Theta) = (1 - \lambda)DP(\delta; \Theta) - \lambda\tau(\delta; \Theta), \quad (106)$$

where $DP(\delta; \Theta) = \binom{J}{2}^{-1} \sum_{j=2}^J \sum_{k=1}^j [1\{\theta_j > \theta_k\} 1\{\delta_j < \delta_k\} + 1\{\theta_j < \theta_k\} 1\{\delta_j > \delta_k\}]$ is the *discordance proportion* – the share of pairwise comparisons that are incorrect – and $\tau(\delta; \Theta)$ is Kendall's τ measure of rank correlation between grades δ_j and parameters θ_j (the share of concordances minus the share of discordances). This expression shows that a decision-maker who weights concordances and discordances equally ($\lambda = 1$) will seek to maximize the rank correlation between true parameters and grades. A λ below one can be interpreted as *discordance aversion* that penalizes mistakes over and above their effect on the rank correlation. Such discordance aversion creates an incentive to coarsen rankings and create a report card with fewer than J grades. If ranking mistakes are very costly, it may be optimal to group units together and avoid making assertions about relative performance of units that cannot be distinguished with sufficient certainty.

²⁵ See Gu and Koenker (2022) and Gu and Koenker (2023b) for other recent approaches to ranking and selection in an EB framework.

Following the approach of [Section 2.5](#), an empirical Bayes grading system treats the θ_j 's as random draws from a mixing distribution G , and minimizes risk based on a deconvolution estimate \hat{G} . Optimal decisions for an oracle that knows G are given by

$$\delta^* = \arg \min_{\delta} (1 - \lambda) DR(\delta) - \lambda \bar{\tau}(\delta), \quad (107)$$

where the *discordance rate* $DR(\delta)$ is the posterior expectation of the discordance proportion $DP(\delta; \Theta)$, and $\bar{\tau}(\delta)$ is the posterior expectation of the rank correlation $\tau(\delta; \Theta)$. Under model [\(82\)–\(83\)](#) with a continuous G , these quantities can be written

$$\bar{\tau}(\delta) = \binom{J}{2} \sum_{j=2}^J \sum_{k=1}^j [2\pi_{jk} - 1][1\{\delta_j > \delta_k\} - 1\{\delta_j < \delta_k\}], \quad (108)$$

$$DR(\delta) = \binom{J}{2} \sum_{j=2}^J \sum_{k=1}^j [\pi_{jk} 1\{\delta_j < \delta_k\} + (1 - \pi_{jk}) 1\{\delta_j > \delta_k\}], \quad (109)$$

where $\pi_{jk} = \Pr[\theta_j > \theta_k | \hat{\theta}_j, \hat{\theta}_k, s_j, s_k]$ is the posterior probability that value-added for unit j exceeds that of unit k given the estimates and standard errors for both units:

$$\pi_{jk} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^t \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - t}{s_j}\right) \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_k - u}{s_k}\right) dG(u) dG(t)}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - t}{s_j}\right) \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_k - u}{s_k}\right) dG(u) dG(t)}. \quad (110)$$

[Equations \(108\) and \(109\)](#) show that the decision-relevant features of the posterior distribution for this ranking problem are the pairwise posterior probabilities π_{jk} . With two units the optimal grading rule sets $\delta_j > \delta_k$ if $\pi_{jk} > (1 + \lambda)^{-1}$, sets $\delta_j < \delta_k$ if $\pi_{jk} < \lambda(1 + \lambda)^{-1}$, and declares a tie ($\delta_j = \delta_k$) otherwise. With more than two units such pairwise decisions may lead to [Condorcet \(1785\)](#)-style cycles that violate transitivity – for example, for three units j , k , and m , the decision-maker may want to set $\delta_j > \delta_k$ but $\delta_j = \delta_m$ and $\delta_k = \delta_m$. [Kline et al. \(2024\)](#) represent problem [\(107\)](#) as an integer linear programming problem with transitivity constraints that rule out such cycles. An EB solution $\hat{\delta}^*$ plugs empirical posterior probabilities $\hat{\pi}_{jk}$ into this problem based on a mixing distribution estimate \hat{G} from the deconvolution step. The resulting report card classifies units into coarse grades that balance information content (captured by $\bar{\tau}(\hat{\delta}^*)$) against ranking mistakes (captured by $DR(\hat{\delta}^*)$), with the strength of this tradeoff determined by the preference parameter λ .

3.6 Compound decisions and shrinkage strategies

I next synthesize themes from throughout the chapter in a simple compound decision framework.²⁶ Consider a decision-maker who aims to select units with high values of a parameter θ_j , which is assumed to be positive for all units ($\theta_j \geq 0$). For example, θ_j might be value-added of unit j relative to a baseline or status quo reference category. The decision-maker earns utility $\theta_j^{1-\omega}$ if she selects a unit with parameter θ_j and pays a constant cost κ for each selection. Parameter $\omega \in [0, 1]$ governs the decision-maker's risk aversion, with $\omega = 0$ corresponding to risk-neutrality. The component-wise loss function for this decision-maker is given by:

$$\ell(\theta_j, \delta_j) = -\delta_j(\theta_j^{1-\omega} - \kappa), \quad (111)$$

where $\delta_j \in \{0, 1\}$ indicates selection of unit j .

Suppose the decision-maker has access to estimates and standard errors $(\hat{\theta}_j, s_j)$ for each unit to use in making decisions. These estimates are unbiased and normally distributed as in equation (82). The underlying parameters θ_j are drawn from a distribution G according to model (83). With J units, the risk of a decision rule δ is then

$$\mathcal{R}(\delta; G, \omega) = - \sum_{j=1}^J \int \int \delta(\hat{\theta}_j, s_j)(\theta_j^{1-\omega} - \kappa) \frac{1}{s_j} \phi\left(\frac{\hat{\theta}_j - \theta_j}{s_j}\right) d\hat{\theta}_j dG(\theta_j). \quad (112)$$

I have written $\mathcal{R}(\delta; G, \omega)$ as a function of G and ω to emphasize that risk depends on the mixing distribution as well as the decision-maker's preferences. The optimal decision rule for an oracle that knows G is $\delta^*(G, \omega) = \arg \min_{\delta \in \mathcal{D}} \mathcal{R}(\delta; G, \omega)$, where \mathcal{D} is the set of candidate decision rules mapping $(\hat{\theta}_j, s_j)$ to binary selection decisions. It is straightforward to show that such an oracle selects units if and only if the posterior expectation of $\theta_j^{1-\omega}$ exceeds the cost κ :

$$\delta^*(\hat{\theta}_j, s_j; G, \omega) = 1 \left\{ \int \theta^{1-\omega} d\mathcal{P}(\theta | \hat{\theta}_j, s_j; G) \geq \kappa \right\}, \quad (113)$$

where the posterior distribution \mathcal{P} is defined as in equation (10).

Equation (113) facilitates comparison of shrinkage approaches distinguished by different degrees of risk aversion. When $\omega = 0$ the decision-maker selects units based on a cutoff in the posterior mean θ_j^* . As ω grows, the expectation that determines the decision rule becomes increasingly sensitive to the lower tail of the posterior distribution. In the limit as ω approaches 1, we have

²⁶ See Section XI.A of Kline et al. (2022) for related discussion in the context of auditing to detect discrimination.

$$\lim_{\omega \rightarrow 1} \delta^*(\hat{\theta}_j, s_j; G, \omega) = 1 \{ \Pr[\theta_j = 0 | \hat{\theta}_j, s_j] \leq 1 - \kappa \}. \quad (114)$$

The probability $\Pr[\theta_j = 0 | \hat{\theta}_j, s_j] = \int 1 \{ \theta = 0 \} d\mathcal{P}(\theta | \hat{\theta}_j, s_j; G)$ is the share of units with $\theta_j = 0$ among those with a particular estimate and standard error, also known as a *local false discovery rate* (Efron et al., 2001). Equation (114) shows that a maximally risk-averse decision-maker selects units based on a cutoff in LFDR. We can therefore think of decisions based on posterior means and false discovery rates as endpoints of a continuum of shrinkage strategies traced out by varying risk aversion.

A decision-maker who does not know G must estimate the posterior expectation in equation (113) to implement a feasible EB decision rule. When G is point-identified as in model (82)-(83), it is sensible to estimate the mixing distribution with the deconvolution methods described in Section 3.2. When the mixing distribution is partially identified as in Section 3.3, there are multiple G 's that are consistent with the distribution of the observed data, and it is no longer clear which one to use for decision-making. *Minimax* decisions provide an important conservative benchmark in this case (Wald, 1945; Savage, 1951; Manski, 2000). Formally, a minimax decision rule is given by:

$$\delta^{mm}(\omega) = \arg \min_{\delta \in \mathcal{D}} \max_{G \in \mathcal{G}_I} \mathcal{R}(\delta; G, \omega), \quad (115)$$

where \mathcal{G}_I is the identified set for the mixing distribution (i.e. the set of G 's consistent with the population distribution of the observed data). Problem (115) can be motivated by an adversarial setup in which an opponent observes the decision rule selected by the analyst and chooses the maximally-damaging mixing distribution in response. An EB minimax rule plugs a deconvolution estimate of the identified set \mathcal{G}_I into (115).

This decision framework clarifies the relationship between the EB posterior mean emphasized for much of this chapter and other shrinkage strategies such as the multiple testing approach developed in Section 3.4. The conventional approach of listing units ordered by EB posterior means mimics the selection decisions of a risk-neutral decision-maker who forms posteriors based on a point estimate of G . In contrast, ordering units by q -values is conservative in two senses. First, this approach corresponds to the decisions of a maximally risk-averse decision-maker, as shown in equation (114). Second, by using an estimated upper bound on the prior probability π_0 , the q -value approach adopts a worst-case empirical prior that is as favorable as possible to the null hypothesis among those consistent with the data, in the spirit of the minimax rule in (115). Whether one of these two shrinkage strategies (or something in between) is preferable depends on the context and economic goal – is the aim to select units that will improve average outcomes, or to recommend units that are likely to generate improvements even in a worst-case scenario? Section 3.7 presents an empirical contrast of these two types of shrinkage.

3.7 Non-parametric EB application: firm-level labor market discrimination

This subsection applies non-parametric empirical Bayes methods to study variation in race and gender discrimination across large US employers. My analysis revists a large resume correspondence experiment conducted by Kline, Rose and Walters (2022, 2024), which extended the analysis of Kline and Walters (2021) to study variation in discrimination across entire firms rather than individual jobs. This experiment submitted applications to multiple entry-level job vacancies nested within 108 Fortune 500 employers. Up to 125 vacancies were sampled for each firm, with each vacancy for a given firm in a different US county. Following Bertrand and Mullainathan (2004), resumes were randomly assigned distinctive names to convey race and gender to the employer, with race assignment stratified so that each vacancy received 4 distinctively-Black and 4 distinctively-white names. Male and female names were each assigned to 50 % of applications with no stratification. The primary outcome is an indicator for whether the employer attempted to contact an applicant within 30 days by phone or e-mail (I sometimes use “callbacks” as shorthand for these contacts). Following Kline et al. (2024), I focus on 97 firms with at least 40 sampled vacancies and overall callback rates above 3 %. This results in a sample of 78,910 applications to 10,453 jobs nested within 97 firms. Full details on the experimental design and sample characteristics are available in Kline et al. (2022) and Kline et al. (2024).

My analysis mostly follows Kline et al. (2022), with some departures to illustrate issues discussed in the preceding sections. In step 1 of the EB recipe, I estimate firm-level discrimination parameters and corresponding standard errors for each employer. These estimates come from firm-specific OLS regressions of callbacks on race, which can be written:

$$Y_i = \sum_{j=1}^J D_{ij} [\alpha_j + \theta_j 1\{R_i = w\}] + e_i, \quad (116)$$

where Y_i indicates a callback for application i , D_{ij} indicates that application i was sent to a vacancy at firm j , and $R_i \in \{w, b\}$ denotes the racial distinctiveness of the name assigned to the application. Parameter α_j measures the callback rate for applications with distinctively-Black names at firm j , while θ_j captures the gap in contact rates between distinctively-white and distinctively-Black names at this firm. Since race was randomly assigned we can interpret θ_j as an average treatment effect of distinctively-white names on callbacks at firm j .²⁷ Corresponding OLS estimates $\hat{\theta}_j$ provide unbiased estimates of these causal parameters. Standard errors s_j are clustered by job to account for job-level

²⁷ See the Appendix to Kline et al. (2022) for a potential outcomes framework formalizing this interpretation.

differences in overall contact rates along with stratification of race assignments by job. Models for gender replace the race indicator with an indicator for a distinctively-male name in [equation \(116\)](#).

3.7.1 Distributions of discrimination

Step 2 of the EB recipe uses the firm-specific estimates to summarize the distribution of discrimination across firms. I first report means and bias-corrected standard deviations of race and gender contact gaps. The estimated mean $\hat{\mu}_\theta$ is the average of $\hat{\theta}_j$'s as in [equation \(8\)](#), while the estimated standard deviation is the square root of the bias-corrected variance estimate from [equation \(80\)](#). Since estimates are independent across firms the mixing variance estimator simplifies to $\hat{\sigma}_\theta^2 = (J - 1)^{-1} \sum_j [(\hat{\theta}_j - \hat{\mu}_\theta)^2 - J^{-1}(J - 1)s_j^2]$ in this case. As in [Kline et al. \(2022\)](#), I report standard errors for $\hat{\mu}_\theta$ and $\hat{\sigma}_\theta$ based on a job-clustered weighted bootstrap procedure that draws *iid* exponential weights for each job and reweights the regressions used to produce $(\hat{\theta}_j, s_j)$ in each bootstrap iteration.

Firms favor distinctively-white names over distinctively-Black names on average, and this gap is highly variable across firms. This can be seen in column (1) of [Table 1](#), which shows estimated means and standard deviations of firm-specific OLS race coefficients. The mean level gap in column (1) demonstrates that on average, firms call applications with distinctively-white names 2.1 % points more often than applications with distinctively-Black names, an estimate that is highly statistically significant (t -statistic > 11). Subsequent rows compare unadjusted and bias-corrected standard deviations of contact gaps. Bias correction reduces the standard deviation of gaps from 2.4 % points to 1.7 % points, which implies that $(1 - (0.017/0.024)^2) \times 100 = 50$ % of the variance in $\hat{\theta}_j$'s is due to statistical noise rather than true firm heterogeneity. Nonetheless, the bias-corrected estimate reveals large differences in discrimination even after accounting for this noise: a firm that is one standard deviation above the mean penalizes distinctively-Black names 80 % more than the average firm.

In contrast to the results for race, column (3) of [Table 1](#) shows that the mean gender coefficient is a precisely-estimated zero. However, the bias-corrected standard deviation estimate is even larger for gender than for race (0.031 versus 0.017). The combination of a zero mean and a large standard deviation implies that there must be mass both above and below zero – some firms favor men, while others favor women. This finding highlights the value of an EB analysis of firm heterogeneity: a focus on the mean would suggest little gender discrimination in this experiment, but the second moment of the distribution reveals substantial discrimination operating in each direction.

I next characterize the full mixing distribution of firm-specific discrimination with the log-spline deconvolution estimator of [Efron \(2016\)](#). Unlike the mean and variance estimates in [Table 1](#), this deconvolution

TABLE 1 Variation in race and gender contact gaps across firms.

	Race gaps (white – Black)		Gender gaps (male – female)	
	Estimate	Std. err.	Estimate	Std. err.
	(1)	(2)	(3)	(4)
Mean	0.021	0.002	-0.001	0.003
Std. devs.: Uncorrected	0.024	0.002	0.042	0.003
Bias-corrected	0.017	0.003	0.031	0.005
Number of firms	97		97	

Notes: This table reports estimated means and standard deviations of race and gender contact gaps across 97 large US employers. Estimates come from OLS regressions of a callback indicator on an intercept and an indicator for a distinctively-white or distinctively-male name, separately for each firm. Columns (1) and (2) show results for race gaps (white names minus Black names), while columns (3) and (4) display results for gender gaps (male names minus female names). Standard errors for firm-specific OLS coefficients are calculated with a job-clustered covariance matrix. Mean estimates are averages of firm-specific gaps. The uncorrected standard deviation is the square root of the sample variance of firm gap estimates. Bias corrected standard deviations subtract the average squared standard error before taking the square root. Standard errors in columns (2) and (4) come from a job-clustered weighted bootstrap procedure with 1000 iterations. Each bootstrap iteration draws an *iid* exponential weight for each job and reweights the regressions used to produce firm-specific gaps and standard errors.

procedure requires taking a stand on the relationship between effect sizes and precision. The basic tests discussed in Section 2.6 reject independence in this case, which indicates that accounting for precision-dependence is likely to be important. Specifically, a regression of the race level gap $\hat{\theta}_j$ on $\log s_j$ yields a coefficient of 0.034 with a robust standard error of 0.005. This indicates strong precision-dependence in the conditional mean of the race gap. A corresponding regression of the male/female gap on its log standard error yields an insignificant coefficient of -0.005 (SE = 0.018), but a regression of $(\hat{\theta}_j - \hat{\mu}_\theta)^2 - s_j^2$ on $\log s_j$ yields a marginally significant coefficient of 0.0035 (SE = 0.0019), suggesting potential dependence in the conditional variance.

Motivated by these tests, I estimate models of dependence between effect sizes and standard errors and deconvolve residuals from these models. Seventy-eight of the 97 estimated racial gaps favor distinctively-white names, and Kline et al. (2022) test and cannot reject the hypothesis that the few observed negative estimates are attributable to sampling error. Following Kline et al. (2024), I therefore adopt a model of precision-dependence that implies white/Black contact gaps are positive for each value of s_j :

$$\hat{\theta}_j = \exp(\psi_1 + \psi_2 \log s_j) r_j, \quad r_j | s_j \sim G_r, \quad (117)$$

where r_j has positive support and $E[r_j] = 1$. This model implies $E[\hat{\theta}_j|s_j] = \exp(\psi_1 + \psi_2 \log s_j)$. I estimate ψ_1 and ψ_2 in a first-step non-linear least squares regression, which yields estimates of $\hat{\psi}_1 = 2.52$ (SE = 0.80) and $\hat{\psi}_2 = 1.56$ (SE = 0.21). I then form residuals $\hat{r}_j = \hat{\theta}_j / \exp(\hat{\psi}_1 + \hat{\psi}_2 \log s_j)$, and estimate G_r by applying the log-spline deconvolution estimator to these residuals, assuming $\hat{r}_j|r_j, s_j \sim N(r_j, \exp(-2\hat{\psi}_1)s_j^{2(1-\hat{\psi}_2)})$ and constraining the mean of the deconvolved distribution to equal 1. The log-spline penalty in equation (86) is calibrated to match a bias-corrected estimate of the variance of r_j . I choose the other log-spline tuning parameters by fixing the number of spline knots at $K = 5$ and using $M = 1,000$ equally-spaced support points between zero and the empirical maximum of \hat{r}_j .

Model (117) is inappropriate for gender since we know from Table 1 that some gender gaps are positive while others are negative. The tests discussed above also indicate dependence in the conditional variance but not the conditional mean. I accommodate these findings with the alternative model

$$\theta_j = \psi_0 + s_j^{\psi_2} r_j, \quad r_j|s_j \sim G_r, \quad (118)$$

where $E[r_j] = 0$ and $Var(r_j) = \sigma_r^2$. This model implies $E[\hat{\theta}_j|s_j] = \psi_0$ and $E[(\hat{\theta}_j - \psi_0)^2 - s_j^2|s_j] = \sigma_r^2 s_j^{2\psi_2}$. I estimate ψ_0 with the mean of the $\hat{\theta}_j$'s, then estimate ψ_2 by non-linear least squares based on the second moment condition. The resulting estimates are $\hat{\psi}_1 = -0.001$ (SE = 0.005) and $\hat{\psi}_2 = 0.854$ (SE = 0.361). I then deconvolve the residuals $\hat{r}_j = (\hat{\theta}_j - \hat{\psi}_1)/s_j^{\hat{\psi}_2}$ assuming $\hat{r}_j|r_j, s_j \sim N(r_j, s_j^{2(1-\hat{\psi}_2)})$ and constraining the mean of the deconvolved distribution to equal zero. Following the analysis for race, I use a five-knot spline with 1,000 equally-spaced support points between the empirical minimum and maximum of the estimated residuals.

Results of these flexible deconvolutions appear in Fig. 2, with log-spline estimates for race in panel A and estimates for gender in panel B. Each panel displays the deconvolved distribution of the residual r_j along with the resulting marginal distribution of θ_j , which comes from applying a change-of-variables to the distribution of residuals combined with the empirical distribution of standard errors.²⁸ The first two moments of these distributions are close to the

²⁸ Estimates of the precision-dependence parameters and residual distribution from equations (117) and (118) imply a conditional distribution of θ_j for each observed value of s_j . I construct a smoothed estimate of the marginal distribution of θ_j by creating an equally-spaced grid of M support points between the minimum and maximum support limits of the conditional distributions, and assigning the mass at each point of the conditional distributions to the closest support point in the marginal grid (as measured by the absolute value of the difference between support points).

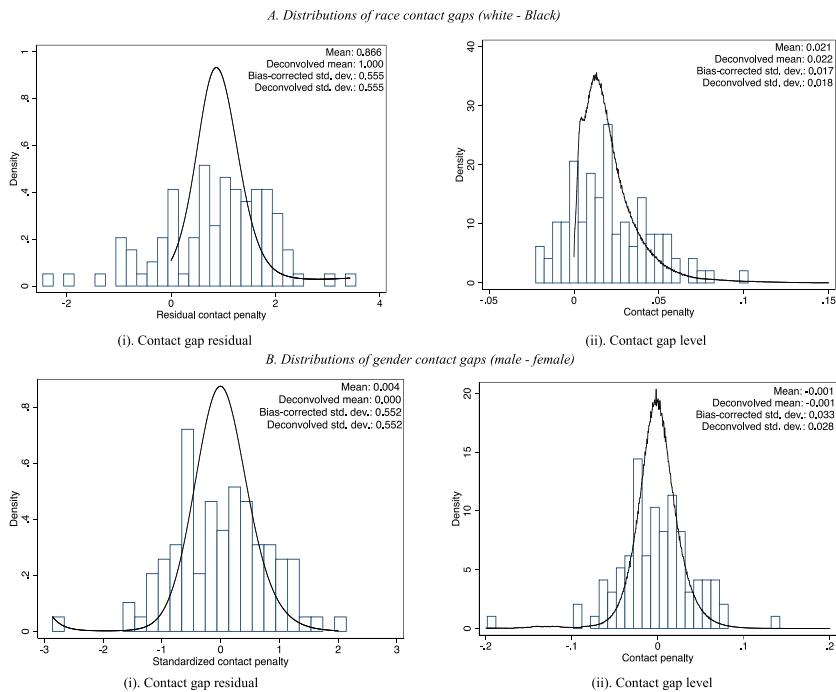


FIG. 2 Deconvolution estimates of discrimination distributions. Notes: This figure displays log-spline deconvolution estimates of distributions of differences in contact rates by race and gender across firms. Panel A displays estimated distributions of race gaps (white – Black), and panel B shows estimated distributions of gender gaps (male – female). Blue bars display histograms of observed estimates, and black curves show estimated prior distributions. In each panel, display (i) shows the deconvolved distribution of residuals transformed to eliminate precision-dependence, and display (ii) shows the resulting distribution of contact gap levels, which is constructed by applying a change of variables to the residual distribution and empirical distribution of standard errors. Log-spline estimates come from a penalized maximum likelihood procedure with penalty term calibrated so that bias-corrected and log-spline estimates of residual standard deviations match.

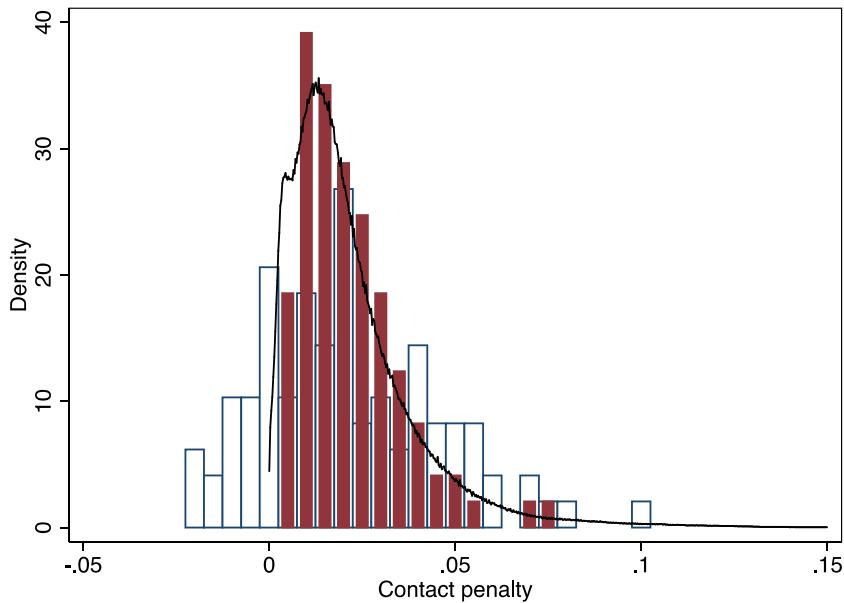
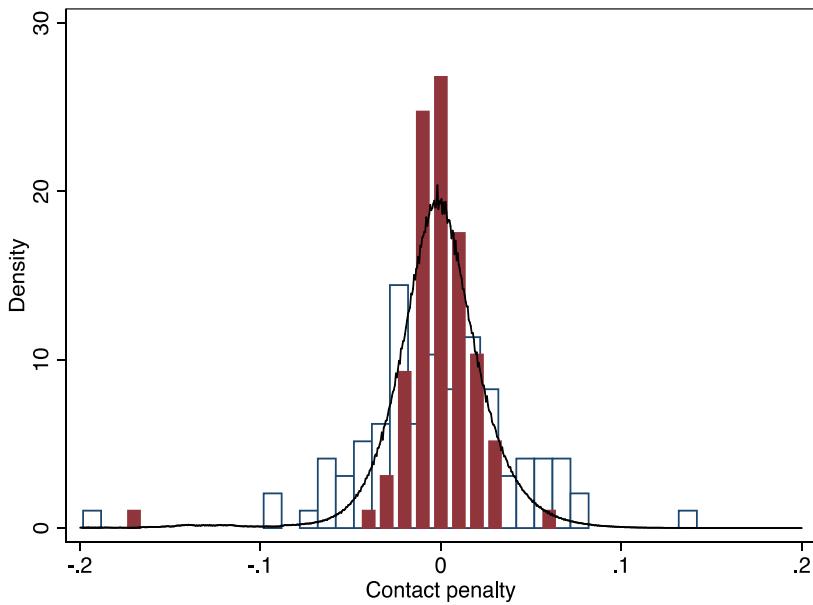
means and variances reported in Table 1, but the non-parametric deconvolutions also reveal more subtle features of the distributions of race and gender discrimination. The distribution of race gaps in panel A(ii) is asymmetric with a long right tail, suggesting that while all firms weakly favor distinctively-white names, the average race gap is driven by a small share of heavy discriminators. The gender gap distribution in panel B(ii) is symmetric with a sharp peak at zero and fat tails in both directions. This indicates that most firms display little preference for male or female names, but a few discriminate substantially against each group.

3.7.2 Posterior predictions of discrimination

Moving to step 3 of the EB recipe, I next consider the prospects for learning about firm-specific discrimination parameters via empirical Bayes shrinkage. Fig. 3 plots posterior mean race and gender gaps $\hat{\theta}_j^*$ derived from the log-spline prior estimates, overlaid on the prior distribution and a histogram of the unbiased $\hat{\theta}_j$ estimates. Posteriors for each θ_j are constructed by computing EB posterior mean residuals \hat{r}_j^* , then transforming these to predict θ_j according to equations (117) and (118). As usual, the shrunk posteriors are less variable than the mixing distribution, which is less variable than the noisy unbiased estimates. Still, the posterior mean estimates are dispersed enough to generate informative estimates of firm-specific parameters. The standard deviation of the posterior mean estimates for race is 0.014 and the average squared standard error of the $\hat{\theta}_j$'s is 0.0003. Coupled with the mixing standard deviation of 0.018 in Fig. 2, this implies shrinkage reduces MSE by an estimated $(1 - [(0.018^2 - 0.014^2)/0.0003]) \times 100 = 57\%$. For gender, the corresponding reduction in MSE is 75 %.

Fig. 4 assesses differences between non-parametric and linear shrinkage by plotting posterior means $\hat{\theta}_j^*$ against linear shrinkage estimates $\hat{\theta}_j^{lin}$. The linear shrinkage estimates are constructed as in equation (14) using mean and variance hyperparameter estimates from Table 1. The differences between linear and non-linear shrinkage estimates are most evident in the tails of the distribution. As shown in panel A(i), the non-parametric posterior means for race inherit the positive support restriction I imposed on the prior, so the few negative point estimates are shrunk to slightly above zero. The non-parametric and linear shrinkage posteriors also differ noticeably in the upper tail, where the non-parametric procedure generates larger values. This is a consequence of two forces: the upper tail of the non-parametric prior distribution is thicker than that of a normal distribution, and the non-parametric posterior incorporates precision-dependence with effect sizes increasing in standard errors. As a result, large positive point estimates are shrunk less by the non-parametric procedure, particularly those with large standard errors.

I parse these explanations in panel A(ii) by incorporating precision-dependence into the linear shrinkage approach. Specifically, I apply linear rather than non-parametric shrinkage to the estimated residuals \hat{r}_j from equation (117) before transforming the residuals to compute posteriors for θ_j . Posteriors from this conditional shrinkage strategy align better with the non-parametric posterior estimates throughout the distribution. However, non-parametric shrinkage still generates some larger posterior means at the top of the distribution due to the thick tail of the non-parametric prior. Panel B shows that non-parametric shrinkage also generates more extreme posterior mean gender gaps in the tails of the distribution than linear shrinkage, especially for one firm

A. Race contact gaps (white - Black)*B. Gender contact gaps (male - female)*

(CAPTION ON NEXT PAGE)

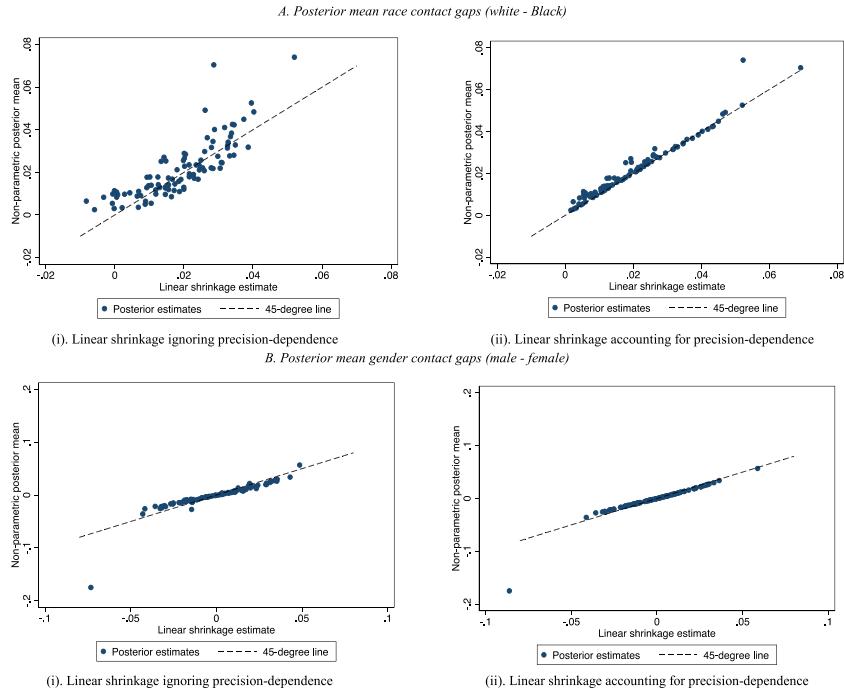


FIG. 4 Non-parametric posterior mean and linear shrinkage estimates of firm-level discrimination. Notes: This figure compares non-parametric and linear shrinkage posterior mean estimates of firm-specific discrimination parameters. Panel A displays posterior mean race gaps (white – Black), and panel B shows posterior mean gender gaps (male – female). Non-parametric posterior means use the deconvolved distributions from Figure 2 as priors. Linear shrinkage posterior means are precision-weighted averages of a firm’s unbiased estimate and an estimated prior mean. In each panel, display (i) shows linear shrinkage estimates assuming effect sizes are independent of standard errors, while display (ii) incorporates precision-dependence by applying linear shrinkage to residuals from models relating effect sizes to standard errors, then transforming the resulting posterior residuals to produce posterior contact gaps. Dashed lines are 45-degree lines.

with a large negative posterior mean indicating substantial discrimination against men. This is a consequence of the extra mass in the left tail of the prior distribution displayed in Fig. 3B(ii).

FIG. 3 Empirical Bayes estimates of firm-level discrimination. Notes: This figure displays distributions of discrimination estimates for 97 US employers. Blue bars show histograms of unbiased contact gap estimates between applicants with distinctively-white and distinctively-Black names (panel A) or distinctively-male and distinctively-female names (panel B). Black curves show log-spline deconvolution estimates of discrimination distributions. Red bars show non-parametric posterior means that use the log-spline estimates as priors.

3.7.3 Multiple testing to detect discrimination

In addition to forming posterior mean estimates with low mean squared error, it is natural ask what we can say with confidence about which firms discriminate against distinctively-Black names at all. I investigate this question with a multiple testing analysis along the lines of [Section 3.4](#). This analysis begins with one-tailed z -tests of the null hypothesis $H_0 : \theta_j = 0$ against the alternative $H_A : \theta_j > 0$, generating p -values $p_j = 1 - \Phi(\hat{\theta}_j/s_j)$ for each firm j .²⁹

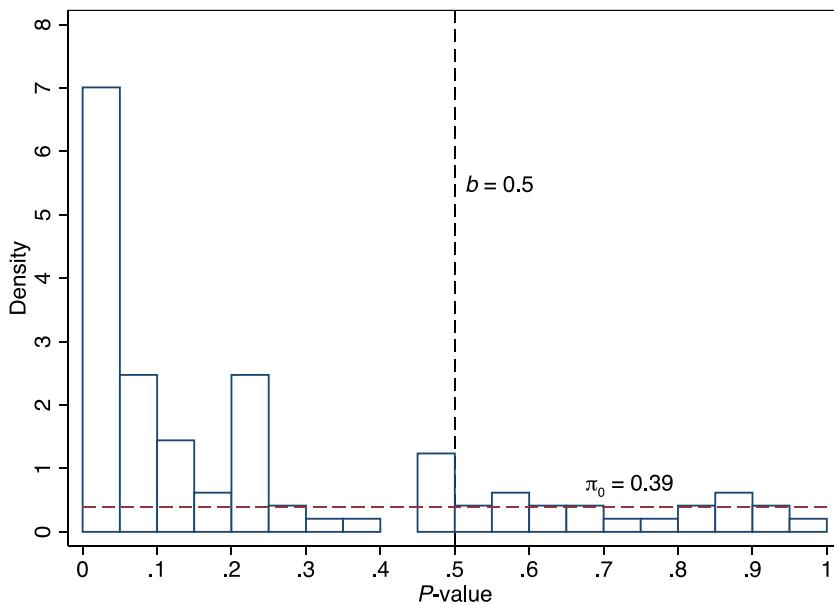
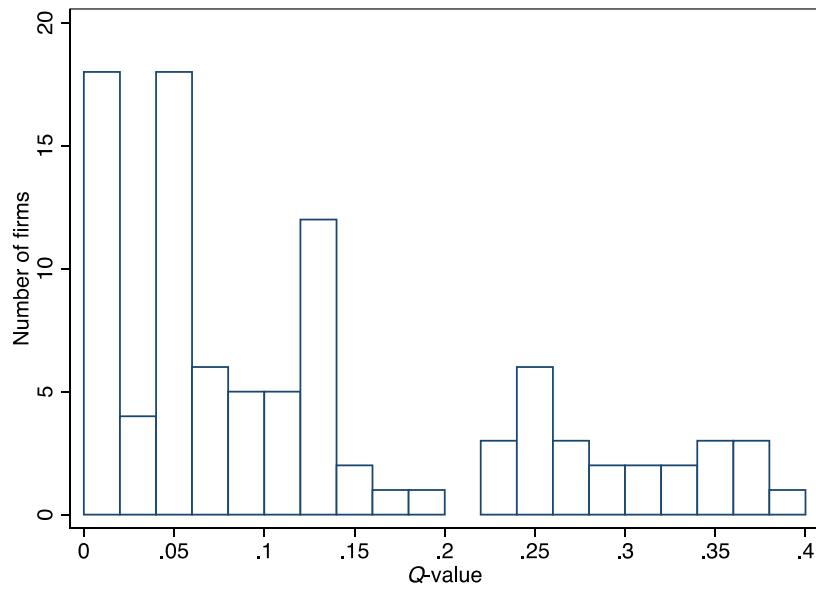
These tests generate small p -values for many firms. Panel A of [Fig. 5](#) displays a histogram of the p -values p_j , which are concentrated toward zero. This is unsurprising since the moment estimates in [Table 1](#) show the mean and variance of G are clearly not zero in this case – this implies that some firms must be discriminating. To obtain a conservative estimate of the share of firms that are *not* discriminating, I set the threshold b in [equation \(103\)](#) to 0.5 (shown as a black vertical line in [Fig. 5A](#)), which generates an estimated bound of $\hat{\pi}_0 = 0.39$ (the red horizontal line). In other words, at least 61 % of firms must be discriminating to rationalize the observed p -value distribution.

This bound on the prior share of true nulls implies that many firms can be reliably classified as discriminating even while controlling the false discovery rate to a low level. I compute q -values by plugging $\hat{\pi}_0$ into [equation \(104\)](#) along with a CDF estimate $\hat{F}(p)$ based on the empirical distribution function of p -values. A histogram of the resulting q -value distribution appears in panel B of [Fig. 5](#). Twenty-eight firms have q -values below 0.05. Since a q -value threshold of 0.05 controls FDR to 5 %, we should expect at most one out of every twenty firms with $q_j \leq 0.05$ to have $\theta_j = 0$ in repeated applications of this test procedure.

3.7.4 Contrasting shrinkage approaches

[Table 2](#) summarizes the empirical Bayes analysis of this experiment by listing the 97 firms in the sample ordered by their q -values. The table reports each firm's unbiased estimate $\hat{\theta}_j$, standard error s_j , p -value p_j , q -value q_j , non-parametric posterior mean $\hat{\theta}_j^*$, and linear shrinkage posterior mean $\hat{\theta}_j^{lin}$. The various EB shrinkage estimates are broadly aligned across firms. For example, Genuine Parts (Napa Auto) has the smallest q -value along with the largest point estimate, the largest linear shrinkage estimate, and the largest non-parametric posterior mean. However, some notable differences are evident across these measures as well. For instance, Walmart has both a large point

²⁹ [Kline et al. \(2022\)](#) conduct a similar multiple-testing analysis for the full sample of 108 firms based on paired-sample t -tests. I use the subsample of 97 firms with large samples studied by [Kline et al. \(2024\)](#) and employ a simple z -test for illustrative purposes, which tends to yield slightly smaller p -values.

A. P-values for no discrimination against distinctively-Black names*B. Q-values for no discrimination against distinctively-Black names*

(CAPTION ON NEXT PAGE)

estimate (0.070) and a large standard error (0.039). Since the estimated prior model indicates strong positive dependence between effects and standard errors, this company's estimate is shrunk very little, and it receives the second-highest posterior mean on the list (0.070). Given its large standard error, however, Walmart receives a q -value of 0.05, ranking 29th by this metric. Conversely, VFC has a modest point estimate of 0.038 but a small standard error of 0.014, which results in both a low q -value of 0.018 and a low posterior mean of 0.022.

[Fig. 6](#) presents a more systematic investigation of differences between shrinkage approaches by plotting decision frontiers based on posterior means and q -values. Each decision rule selects the 20 % of firms with most extreme estimates of discrimination against distinctively-Black names, corresponding to posterior means above 0.032 and q -values below 0.024. These decision rules can be seen as versions of the selection rule in [equation \(113\)](#), setting $\omega = 0$ for the posterior mean, $\omega \rightarrow 1$ for the q -value, and calibrating the cost κ so that 20 % of firms are selected. Observed combinations of point estimates and log standard errors are denoted with black points, and shaded regions depict hypothetical decisions for combinations not observed in the experiment.

While there is substantial overlap between firms selected based on posterior means and q -values, there is also slippage between these two decision rules. Specifically, 13 of the 19 firms selected by the posterior mean rule are also selected by the q -value rule. Similarly, since each rule selects the same total number by design, 13 of 19 selected by the q -value rule are also selected by the posterior mean rule. Columns (7) and (8) of [Table 2](#) label the firms selected by each decision rule.

Discrepancies between these classifications arise because of the differing shapes of the selection frontiers depicted in [Fig. 6](#). A q -value decision rule defines an upward-sloping frontier in the plane relating point estimates to log standard errors. Since there is more uncertainty for firms with large standard errors, higher point estimates are necessary to classify such firms as discriminating while limiting the false discovery rate. In contrast, the posterior mean selection frontier defines a *downward*-sloping relationship between point

FIG. 5 Empirical Bayes multiple-testing analysis of firm-level discrimination. Notes: This figure displays the results of an empirical Bayes multiple testing analysis of discrimination against distinctively-Black names for 97 firms. Panel A shows p -values from one-tailed z -tests. The black vertical line indicates a threshold of $b = 0.5$ used to bound the share of true nulls, and the red horizontal line displays the resulting estimated bound. Panel B shows a histogram of q -values constructed based on the bound from panel A and the empirical distribution function of p -values.

TABLE 2 Empirical Bayes estimates of discrimination against distinctively-Black names.

	Estimate (1)	Std. err. (2)	P- value (3)	Q- value (4)	Non- par. post. mean (5)	Linear. shrinkage (6)	Bottom 20% (7)	Top 20% post. mean (8)
Genuine Parts (Napa Auto)	0.098	0.020	0.000	0.000	0.074	0.052	Yes	Yes
AutoNation	0.053	0.015	0.000	0.004	0.032	0.039	Yes	Yes
Advance Auto Parts	0.074	0.022	0.000	0.006	0.048	0.040	Yes	Yes
O'Reilly Automotive	0.079	0.024	0.001	0.006	0.053	0.040	Yes	Yes
Rite Aid	0.047	0.016	0.002	0.012	0.028	0.034	Yes	
Ascena (Ann Taylor / Loft)	0.068	0.023	0.002	0.012	0.045	0.037	Yes	Yes
Gap	0.052	0.019	0.003	0.015	0.033	0.035	Yes	Yes
Dick's	0.051	0.021	0.007	0.015	0.034	0.033	Yes	Yes
Murphy USA	0.049	0.020	0.007	0.016	0.033	0.033	Yes	Yes
Dillard's	0.045	0.017	0.003	0.016	0.028	0.033	Yes	
Tractor Supply	0.040	0.016	0.007	0.016	0.025	0.031	Yes	

AutoZone	0.054	0.022	0.006	0.017	0.037	0.033	Yes	Yes
TJX	0.048	0.019	0.007	0.017	0.031	0.033	Yes	
Bed Bath & Beyond	0.040	0.016	0.006	0.018	0.024	0.031	Yes	
Pizza Hut	0.056	0.022	0.006	0.018	0.038	0.034	Yes	
VFC (North Face / Vans)	0.038	0.014	0.005	0.018	0.022	0.031	Yes	
Pilot Flying /	0.060	0.024	0.006	0.020	0.043	0.034	Yes	
Walgreens	0.061	0.024	0.005	0.020	0.042	0.035	Yes	
Universal Health	0.055	0.024	0.013	0.024	0.041	0.032	Yes	
Sherwin-Williams	0.035	0.015	0.012	0.025	0.022	0.028		
Bath & Body Works	0.040	0.018	0.014	0.026	0.028	0.030		
Olive Garden	0.034	0.016	0.016	0.028	0.022	0.028		
J.C. Penney	0.030	0.016	0.028	0.045	0.021	0.026		
Nationwide	0.026	0.014	0.027	0.045	0.017	0.024		
Dollar General	0.039	0.022	0.034	0.048	0.032	0.028	Yes	
Publix	0.047	0.026	0.033	0.049	0.040	0.029	Yes	
Marriott	0.030	0.017	0.038	0.049	0.022	0.025		
Republic Services	0.020	0.011	0.038	0.050	0.012	0.020		
Walmart	0.070	0.039	0.037	0.050	0.070	0.029	Yes	

Continued

Table 2 Empirical Bayes estimates of discrimination against distinctively-Black names.—Cont'd

	Estimate (1)	Std. err. (2)	P- value (3)	Q- value (4)	Non- par. (5)	Linear. (6)	Bottom 20 % (7)	Top 20 % post. mean (8)
Firm name								
US Foods	0.024	0.014	0.044	0.050	0.017	0.023		
Victoria's Secret	0.042	0.023	0.033	0.050	0.034	0.028	Yes	
Goodyear	0.015	0.009	0.042	0.052	0.009	0.016		
Hertz	0.018	0.010	0.044	0.052	0.011	0.019		
Builders	0.028	0.017	0.047	0.053	0.021	0.024		
FirstSource								
Aramark	0.019	0.012	0.051	0.055	0.013	0.020		
Tyson Foods	0.024	0.016	0.061	0.058	0.019	0.023		
Comcast	0.039	0.025	0.057	0.059	0.036	0.027	Yes	
Best Buy	0.028	0.018	0.059	0.059	0.023	0.024		
Cardinal Health	0.034	0.021	0.056	0.059	0.030	0.026		
Dollar Tree	0.030	0.019	0.061	0.059	0.026	0.025		
UGI	0.022	0.015	0.077	0.068	0.018	0.022		
LKQ Auto	0.027	0.019	0.076	0.069	0.023	0.024		

KFC	0.022	0.015	0.075	0.070	0.018	0.022
Dean Foods	0.044	0.032	0.082	0.071	0.049	0.026
Costco	0.022	0.016	0.089	0.073	0.019	0.022
CVS Health	0.015	0.011	0.088	0.074	0.011	0.017
PepsiCo	0.009	0.007	0.102	0.083	0.005	0.011
Macy's	0.012	0.010	0.118	0.093	0.010	0.015
Jones Lang LaSalle	0.018	0.015	0.123	0.095	0.017	0.019
Sears (incl. repair / auto)	0.017	0.015	0.129	0.096	0.016	0.019
Cintas	0.022	0.019	0.128	0.097	0.024	0.021
Kohl's	0.015	0.014	0.143	0.105	0.015	0.017
Edward Jones	0.013	0.012	0.147	0.106	0.012	0.016
CBRE	0.007	0.007	0.155	0.109	0.005	0.009
AECOM	0.005	0.005	0.159	0.110	0.004	0.007
DISH	0.019	0.019	0.166	0.113	0.023	0.020
Nordstrom	0.011	0.014	0.208	0.127	0.013	0.015
Estee Lauder	0.018	0.022	0.206	0.128	0.026	0.020
XPO Logistics	0.010	0.013	0.217	0.129	0.013	0.015
Performance Food Group	0.019	0.023	0.206	0.130	0.028	0.020

Continued

Table 2 Empirical Bayes estimates of discrimination against distinctively-Black names.—Cont'd

	Estimate (1)	Std. err. (2)	p_{-} value (3)	Q - value (4)	Non- par. post. mean (5)	Linear. shrinkage (6)	Bottom 20 % q-value (7)	Top 20 % post mean (8)
International Paper	0.013	0.016	0.216	0.130	0.017	0.017		
Honeywell	0.011	0.013	0.204	0.131	0.013	0.015		
Geico	0.018	0.024	0.227	0.133	0.029	0.020		
GameStop	0.008	0.011	0.238	0.133	0.010	0.012		
US Bank	0.012	0.014	0.204	0.133	0.014	0.016		
Starbucks	0.014	0.019	0.232	0.134	0.021	0.018		
UnitedHealth	0.006	0.008	0.236	0.134	0.006	0.009		
Ross Stores	0.018	0.022	0.203	0.136	0.027	0.020		
AT&T	0.010	0.017	0.281	0.155	0.017	0.016		
Stanley Black & Decker	0.008	0.014	0.290	0.157	0.013	0.013		
Foot Locker	0.006	0.013	0.312	0.167	0.012	0.012		
United Rentals	0.006	0.015	0.352	0.186	0.014	0.013		
Ulta Beauty	0.002	0.015	0.461	0.233	0.014	0.011		

WestRock	0.002	0.019	0.460	0.236	0.018	0.012
Target	0.001	0.011	0.457	0.238	0.009	0.007
CarMax	0.001	0.025	0.482	0.241	0.025	0.015
Home Depot	0.000	0.006	0.508	0.241	0.003	0.002
Kroger	0.000	0.011	0.506	0.243	0.009	0.006
Disney (incl. stores)	0.000	0.015	0.500	0.244	0.014	0.010
Waste Management	0.000	0.015	0.497	0.246	0.013	0.009
Safeway	-0.003	0.027	0.552	0.259	0.027	0.014
McLane Company	-0.005	0.026	0.571	0.261	0.025	0.013
Lab Corp	-0.002	0.014	0.571	0.265	0.011	0.007
Mondelez	-0.005	0.020	0.601	0.272	0.018	0.010
Lowe's	-0.008	0.020	0.648	0.290	0.018	0.009
Sysco	-0.003	0.006	0.675	0.298	0.003	0.000
FedEx	-0.007	0.014	0.687	0.300	0.010	0.004
State Farm	-0.009	0.013	0.741	0.320	0.010	0.003
Ryder System	-0.007	0.009	0.775	0.331	0.005	-0.001
Charter / Spectrum	-0.010	0.012	0.800	0.338	0.008	0.001
JPMorgan Chase	-0.012	0.014	0.820	0.342	0.010	0.001

Continued

Table 2 Empirical Bayes estimates of discrimination against distinctively-Black names.—Cont'd

	Estimate	Std. err.	p- value	Q- value	Non- par.	Linear.	Bottom 20 %	Top 20 %
Firm name	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
J.B. Hunt	-0.017	0.015	0.861	0.356	0.011	0.001		
Avis-Budget	-0.017	0.014	0.877	0.358	0.010	-0.001		
Kindred	-0.021	0.017	0.895	0.362	0.011	0.000		
Healthcare								
Hilton	-0.018	0.013	0.916	0.367	0.008	-0.003		
Quest Diagnostics	-0.009	0.005	0.947	0.375	0.002	-0.006		
Dr Pepper	-0.023	0.012	0.973	0.381	0.007	-0.008		

Notes: This table reports the results of an empirical Bayes analysis of discrimination against distinctively-Black names among 97 employers. Firms are ordered by q-values from a multiple testing procedure. Column (1) shows each firm's estimated difference in contact rates between distinctively-white and distinctively-Black names, and column (2) displays corresponding standard errors. Column (3) shows p-values from one-tailed tests of the null hypothesis of no discrimination against distinctively-Black names, and column (4) shows q-values. Column (5) displays non-parametric posterior means using log-spline deconvolution estimates as priors, incorporating dependence between effect sizes and standard errors. Column (6) shows linear shrinkage posterior means ignoring precision-dependence. Column (7) labels firms in the bottom 20% of the q-values from column (4). Column (8) labels firms in the top 20% of the non-parametric posterior means from column (5).

estimates and log standard errors.³⁰ This pattern is driven by the strong positive relationship between effect sizes and standard errors evident in the scatter plot of black points, which is built into the prior distribution used to construct posterior means. As a result, among firms with intermediate point estimates, the q -value rule selects firms with smaller standard errors (the yellow region) while the posterior mean rule selects firms with larger standard errors (the green region).

While the q -value and posterior mean decision rules select different sets of firms, differences in expected outcomes between these approaches turn out to be modest. Average posterior means among firms selected by the posterior mean and q -value rules are 0.043 and 0.037, implying that a q -value cutoff selects firms with only slightly smaller expected discrimination values. Likewise, the highest q -value among firms selected by the posterior mean rule is only 0.071, suggesting that a cutoff in $\hat{\theta}_j^*$ selects firms where the posterior probability of discriminating is also high. These findings indicate that a risk neutral analyst would pay little price for using the decision rule of a risk-averse decision-maker (and vice versa). More generally, these sorts of contrasts between decision rules can help to trace out the frontier of outcomes available with different shrinkage strategies and assess the robustness of EB shrinkage analyses to the assumed form of decision-maker preferences.

4 Conclusion

Empirical research in labor economics increasingly focuses on variation in quality or conduct across large sets of units like firms, schools, or neighborhoods. This chapter has reviewed empirical Bayes methods for quantifying heterogeneity, estimating unit-specific parameters, and making statistical decisions in such studies. The EB recipe outlined here proceeds by estimating each unit's parameter, using the ensemble of parameter estimates to construct an empirical prior distribution, and forming posteriors based on this prior combined with the unit-specific estimates. This EB shrinkage approach can be applied in service of a variety of statistical and economic goals, including reducing aggregate mean squared error, directing workers or consumers to units with favorable expected outcomes, ranking units, and making selection decisions while limiting the likelihood of mistakes.

³⁰The posterior mean selection rule becomes extremely non-linear for log standard errors below -4.5 , suggesting that an enormous point estimate is required to warrant selection in this region. This phenomenon is due to the upper bound on the support of residuals imposed in Panel A(i) of Fig. 2. The maximum of the empirical residuals \hat{r}_j used to set this support is 3.4, which implies θ_j can be no bigger than $\exp(\hat{\psi}_1 + \hat{\psi}_2 \log s_j) \times 3.4$. With $\hat{\psi}_1 = 2.52$, $\hat{\psi}_2 = 1.56$, and a decision cutoff of $\hat{\theta}_j^* \geq 0.032$, a firm with $\log s_j \leq -4.61$ cannot be selected regardless of its point estimate. This issue has little impact on empirical selection decisions since no firms have estimates near the highly non-linear portion of the decision frontier.

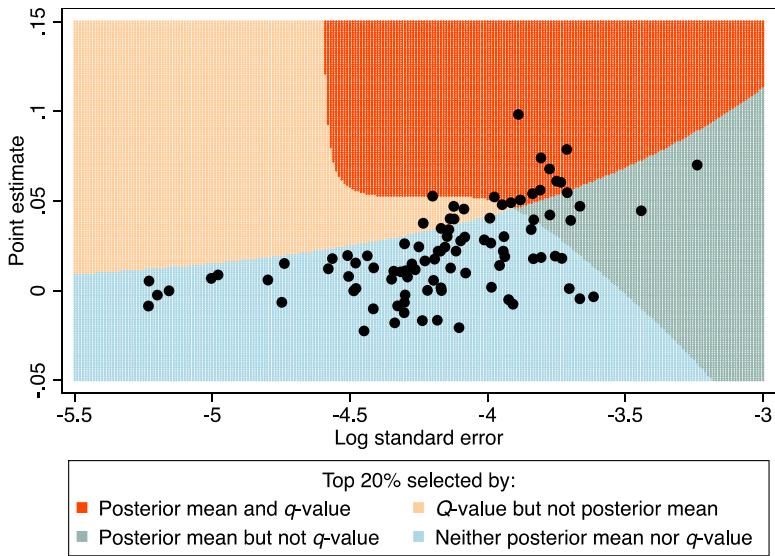


FIG. 6 Selection frontiers for empirical Bayes decision rules. Notes: This figure contrasts empirical Bayes decision rules that select firms based on posterior means and q -values for discrimination against distinctively-Black names. Posterior means use log-spline deconvolution estimates as priors, allowing for dependence between effect sizes and precision. Q -values come from a multiple testing analysis that bounds the prior share of firms that do not discriminate. Decision rules select the 20 % of firms with most extreme discrimination estimates based on either the posterior mean or the q -value. The red region shows combinations of point estimates and standard errors that are selected by both posterior mean and q -value decision rules. The yellow region shows combinations selected by a q -value decision rule but not a posterior mean decision rule. The green region shows combinations selected by a posterior mean decision rule but not a q -value decision rule. The blue region shows combinations selected by neither decision rule. Black points indicate the point estimates and standard errors for the 97 firms in the experiment.

Thanks to the increasing availability of large-scale administrative labor market data, potential applications of EB methods in labor economics should continue to grow. This is likely to generate avenues for answering novel economic questions as well as new methodological challenges. Most EB value-added analyses in labor economics to date have employed simple James/Stein-style linear shrinkage strategies. However, realistic empirical applications often bear little resemblance to the stylized James/Stein framework with normally-distributed estimates, homogeneous variances, and independence across units. The non-parametric EB methods discussed in the second half of this chapter provide tools for quantifying heterogeneity and implementing shrinkage in more general settings. Applying these methods to account for the complexities of real-world labor market data and research designs is a fruitful direction for empirical work.

A second promising direction is to tighten the link between economic objectives and econometric estimation when applying EB methods. This chapter has emphasized the connection between EB shrinkage and decision problems that aim to minimize various forms of aggregate error across units. Conventional linear shrinkage is useful for reducing mean squared error and making risk-neutral selection decisions, which may not correspond to how EB estimates will be used in practice. An explicit statement of the relevant loss function clarifies the appropriate shrinkage strategy and injects economic reasoning into value-added analysis. Such an approach has the potential to make EB applications in labor economics more useful and actionable for policymakers, workers, and households.

References

- Abadie, A., Kasy, M., 2019. Choosing among regularized estimators in empirical economics: the risk of machine learning. *The Review of Economics and Statistics* 101, 743–762.
- Abaluck, J., CaceresBravo, M., Hull, P., Starc, A., 2021. Mortality effects and choice across private health insurance plans. *The Quarterly Journal of Economics* 136, 1557–1610.
- Abdulkadiroğlu, A., Pathak, P.A., Schellenberg, J., Walters, C.R., 2020. Do parents value school effectiveness? *American Economic Review* 110, 1502–1539.
- Abdulkadiroğlu, A., Angrist, J.D., Dynarski, S., Kane, T.J., Pathak, P.A., 2011. Accountability and flexibility in public schools: evidence from Boston's charters and pilots. *Quarterly Journal of Economics* 126 (2), 699–748.
- Abowd, J.M., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. *Econometrica* 67, 251–333.
- Andrews, I., Kitagawa, T., McCloskey, A., 2023. Inference on winners. *The Quarterly Journal of Economics* 139, 305–358.
- Andrews, M.J., Gill, L., Schank, T., Upward, R., 2008. High wage workers and low wage firms: negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171, 673–697.
- Angrist, J., Hull, P., Pathak, P.A., Walters, C., 2024a. Credible school value-added with undersubscribed school lotteries. *The Review of Economics and Statistics* 106, 1–19.
- Angrist, J., Hull, P., Pathak, P.A., Walters, C.R., 2024b. Race and the mismeasure of school quality. *American Economic Review: Insights* 6, 20–37.
- Angrist, J., Hull, P., Walters, C., 2023. Chapter 1 – Methods for measuring school effectiveness. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), *Handbook of the Economics of Education*. vol. 7. Elsevier, pp. 1–60.
- Angrist, J.D., Hull, P.D., Pathak, P.A., Walters, C.R., 2016. Interpreting tests of school VAM validity. *American Economic Review: Papers & Proceedings* 106, 388–392.
- Angrist, J.D., Hull, P.D., Pathak, P.A., Walters, C.R., 2017. Leveraging lotteries for school value-added: testing and estimation. *Quarterly Journal of Economics* 132, 871–919.
- Angrist, J.D., Pathak, P.A., Walters, C.R., 2013. Explaining charter school effectiveness. *American Economic Journal: Applied Economics* 5, 1–27.
- Arceo-Gomez, E.O., Campos-Vasquez, R.M., 2014. Race and marriage in the labor market: a discrimination correspondence study in a developing country. *American Economic Review: Papers & Proceedings* 104, 376–380.

- Armstrong, T.B., Kolesár, M., Plagborg-Møller, M., 2022. Robust empirical Bayes confidence intervals. *Econometrica* 90, 2567–2602.
- Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. *Annual Review of Economics* 11, 685–725.
- Avivi, H., 2024. One land, many promises: assessing the consequences of unequal childhood location effects, Working paper.
- Bartlett, M.S., 1936. The square root transformation in analysis of variance. Supplement to the *Journal of the Royal Statistical Society* 3, 68–78.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57, 289–300.
- Bergman, P., Hill, M.J., 2018. The effects of making performance information public: regression discontinuity evidence from Los Angeles teachers. *Economics of Education Review* 66, 104–113.
- Bertrand, M., Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94, 991–1013.
- Beuermann, D.W., Jackson, C.K., Navarro-Sola, L., Pardo, F., 2022. What is a good school, and can parents tell? Evidence on the multidimensionality of school output. *The Review of Economic Studies* 90, 65–101.
- Bock, M.E., 1975. Minimax estimators of the mean of a multivariate normal distribution. *The Annals of Statistics* 3, 209–218.
- Bonhomme, S., Denis, A., 2024. Estimating heterogeneous effects: applications to labor economics, Working paper.
- Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., Setzler, B., 2023. How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics* 41, 291–322.
- Bonhomme, S., Lamadon, T., Manresa, E., 2019. A distributional framework for matched employer employee data. *Econometrica* 87, 699–739.
- Brown, L.D., 2008. In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies. *The Annals of Applied Statistics* 2, 113–152.
- Card, D., 1999. The causal effect of education on earnings. *Handbook of Labor Economics*. vol. 3. Elsevier, pp. 1801–1863.
- Card, D., Cardoso, A.R., Heining, J., Kline, P., 2018. Firms and labor market inequality: evidence and some theory. *Journal of Labor Economics* 36, S13–S70.
- Card, D., Cardoso, A.R., Kline, P., 2015. Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics* 131, 633–686.
- Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics* 128, 967–1015.
- Chamberlain, G., 1982. Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.
- Chan, D.C., Gentzkow, M., Yu, C., 2022. Selection with variation in diagnostic skill: evidence from radiologists. *The Quarterly Journal of Economics* 137, 729–783.
- Chandra, A., Dalton, M., Staiger, D.O., 2023. Are hospital quality indicators causal? NBER working paper no. 31789.
- Chandra, A., Finkelstein, A., Sacarny, A., Syverson, C., 2016. Health care exceptionalism? Performance and allocation in the US health care sector. *American Economic Review* 106, 2110–2144.

- Chen, J., 2023. Empirical Bayes when estimation precision predicts parameters, ArXiv working paper 2212.14444.
- Chetty, R., Deming, D., Friedman, J.N., 2023. Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges, NBER working paper no. 31492.
- Chetty, R., Friedman, J.N., Hendren, N., Jones, M.R., Porter, S.R., 2018. The opportunity atlas: mapping the childhood roots of social mobility, Working paper.
- Chetty, R., Friedman, J.N., Rockoff, J.E., 2014a. Measuring the impact of teachers I: evaluating bias in teacher value-added estimates. *American Economic Review* 104, 2563–2593.
- Chetty, R., Friedman, J.N., Rockoff, J.E., 2014b. Measuring the impact of teachers II: teacher value-added and student outcomes in adulthood. *American Economic Review* 104, 2633–2679.
- Chetty, R., Friedman, J.N., Saez, E., Turner, N., Yagan, D., 2017. Mobility report cards: the role of colleges in intergenerational mobility, The Equality of Opportunity Project, January.
- Chetty, R., Hendren, N., 2018. Impacts of neighborhoods on intergenerational mobility II: county-level estimates. *Quarterly Journal of Economics* 133, 1163–1228.
- Condorcet, M.D., 1785. *Essay on the Application of Analysis to the Probability of Majority Decisions*. Imprimerie Royale, Paris.
- Dale, S.B., Krueger, A.B., 2002. Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables. *Quarterly Journal of Economics* 117, 1491–1527.
- Dale, S.B., Krueger, A.B., 2014. Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of Human Resources* 49, 323–358.
- Dobbie, W., Fryer, R.G., 2013. Getting beneath the veil of effective schools: evidence from New York City, *American Economic Journal: Applied Economics* 5, 28–60.
- Efron, B., 2012. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. vol. 1 Cambridge University Press.
- Efron, B., 2016. Empirical Bayes deconvolution estimates. *Biometrika* 103, 1–20.
- Efron, B., Morris, C., 1973a. Combining possibly related estimation problems. *Journal of the Royal Statistical Society. Series B (Methodological)* 35, 379–421.
- Efron, B., Morris, C., 1973b. Stein's estimation rule and its competitors – an empirical Bayes approach. *Journal of the American Statistical Association* 68, 117–130.
- Efron, B., Morris, C., 1975. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association* 70, 311–319.
- Efron, B., Tibshirani, R., 1996. Using specially designed exponential families for density estimation. *The Annals of Statistics* 24, 2431–2461.
- Efron, B., Tibshirani, R., Storey, J.D., Tusher, V., 2001. Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association* 96, 1151–1160.
- Einav, L., Finkelstein, A., Mahoney, N., 2022. Producing health: measuring value added of nursing homes, NBER working paper no. 30228.
- Evdokimov, K., White, H., 2012. Some extensions of a lemma of kotlarski. *Econometric Theory* 28, 925–932.
- Fenizia, A., 2022. Managers and productivity in the public sector. *Econometrica* 90, 1063–1084.
- Frandsen, B., Lefgren, L., Leslie, E., 2023. Judging judge fixed effects. *American Economic Review* 113, 253–277.
- Gilraine, M., Gu, J., McMillan, R., 2020. A new method for estimating teacher value-added, NBER working paper no. 27094.

- Goldhaber, D., Gross, B., Player, D., 2011. Teacher career paths, teacher quality, and persistence in the classroom: are public schools keeping their best? *Journal of Policy Analysis and Management* 30, 57–87.
- Goncalves, F., Mello, S., 2021. A few bad apples? Racial bias in policing. *American Economic Review* 111, 1406–1441.
- Gruetter, M., Lalivé, R., 2009. The importance of firms in wage determination. *Labour Economics* 16, 149–160.
- Gu, J., Koenker, R., 2016. On a problem of Robbins. *International Statistical Review / Revue Internationale de Statistique* 84, 224–244.
- Gu, J., Koenker, R., 2017. Unobserved heterogeneity in income dynamics: an empirical Bayes perspective. *Journal of Business & Economic Statistics* 35, 1–16.
- Gu, J., Koenker, R., 2022. Ranking and selection from pairwise comparisons: empirical Bayes methods for citation analysis. *AEA Papers and Proceedings* 112, 624–629.
- Gu, J., Koenker, R., 2023a. GLVmix: NPMLE of Gaussian location-scale mixture model, <https://rdrr.io/cran/REBayes/src/R/GLVmix.R>.
- Gu, J., Koenker, R., 2023b. Invidious comparisons: ranking and selection as compound decisions. *Econometrica* 91, 1–41.
- Hausman, J.A., 1978. Specification tests in econometrics. *Econometrica* 46, 1251–1271.
- Heckman, J.J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.
- Holland, P.W., 1973. Covariance stabilizing transformations. *The Annals of Statistics* 1, 84–92.
- Jackson, C.K., Porter, S.C., Easton, J.Q., Blanchard, A., Kiguel, S., 2020. School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights* 2, 491–508.
- James, W., Stein, C., 1961. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1, 361–379.
- Jiang, W., 2020. On general maximum likelihood empirical Bayes estimation of heteroscedastic IID normal means. *Electronic Journal of Statistics* 14, 2272–2297.
- Jiang, W., Zhang, C.-H., 2009. General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics* 37, 1647–1684.
- Keane, M., Neal, T., 2023. Instrument strength in IV estimation and inference: a guide to theory and practice. *Journal of Econometrics* 235, 1625–1653.
- Kiefer, J., Wolfowitz, J., 1956. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics* 27, 887–906.
- Kline, P., Rose, E.K., Walters, C.R., 2022. Systemic discrimination among large U.S. employers. *The Quarterly Journal of Economics* 137, 1963–2036.
- Kline, P., Saggio, R., Sølvsten, M., 2020. Leave-out estimation of variance components. *Econometrica* 88, 1859–1898.
- Kline, P., Walters, C., 2021. Reasonable doubt: experimental detection of job-level employment discrimination. *Econometrica* 89, 765–792.
- Kline, P.M., Rose, E.K., Walters, C.R., 2024. A discrimination report card. *American Economic Review* 114, 2472–2525.
- Koenker, R., 2020. Empirical Bayes confidence intervals: an R vinaigrette, Working paper.
- Koenker, R., Gu, J., 2017. REBayes: an R package for empirical Bayes mixture methods. *Journal of Statistical Software* 82, 1–26.
- Koenker, R., Gu, J., 2024. Empirical Bayes for the reluctant frequentist, ArXiv working paper 2404.30422.

- Koenker, R., Mizera, I., 2014. Convex optimization, shape constraints, compound decisions, and empirical Bayes rules. *Journal of the American Statistical Association* 109, 674–685.
- Kolstad, J.T., 2013. Information and quality when motivation is intrinsic: evidence from surgeon report cards. *American Economic Review* 103, 2875–2910.
- Kotlarski, I., 1967. On characterizing the gamma and the normal distribution. *Pacific Journal of Mathematics* 20, 69–76.
- Krueger, A.B., Summers, L.H., 1988. Efficiency wages and the inter-industry wage structure. *Econometrica* 56, 259–293.
- Lachowska, M., Mas, A., Saggio, R., Woodbury, S.A., 2023a. Do firm effects drift? Evidence from Washington administrative data. *Journal of Econometrics* 233, 375–395.
- Lachowska, M., Mas, A., Saggio, R., Woodbury, S.A., 2023b. Work hours mismatch, NBER working paper no. 31205.
- Lai, T.L., Siegmund, D., 2018. Herbert Robbins, 1915–2001: biographical memoir, National Academy of Sciences.
- Lehmann, E.L., Romano, J.P., 2005. Generalizations of the familywise error rate. *The Annals of Statistics* 33, 1138–1154.
- Lindley, D.V., 1962. Discussion of professor steinas paper. *Journal of the Royal Statistical Society: Series B (Methodological)* 24, 285–287.
- MacKinnon, J.G., White, H., 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29, 305–325.
- Manski, C.F., 2000. Identification problems and decisions under ambiguity: empirical analysis of treatment response and normative analysis of treatment choice. *Journal of Econometrics* 95, 415–442.
- Mogstad, M., Romano, J.P., Shaikh, A.M., Wilhelm, D., 2023. Inference for ranks with applications to mobility across neighbourhoods and academic achievement across countries. *The Review of Economic Studies* 91, 476–518.
- Morris, C.N., 1983. Parametric empirical Bayes inference: theory and applications. *Journal of the American Statistical Association* 78, 47–55.
- Mountjoy, J., Hickman, B., 2021. The returns to college(s): relative value-added and match effects in higher education, NBER working paper no. 29276.
- Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives* 31, 87–106.
- Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica* 46, 69–85.
- Narasimhan, B., Efron, B., 2020. deconvolver: a G-modeling program for deconvolution and empirical Bayes estimation. *Journal of Statistical Software* 94, 1–20.
- Nunley, J.M., Pugh, A., Romero, N., Seals, R.A., 2015. Racial discrimination in the labor market for recent college graduates: evidence from a field experiment. *B. E. Journal of Economic Analysis and Policy* 15, 1093–1125.
- Pope, D.G., 2009. Reacting to rankings: evidence from “America’s Best Hospitals. *Journal of Health Economics* 28, 1154–1165.
- Raudenbush, S., Reardon, S., Nomi, T., 2012. Statistical analysis for multisite trials using instrumental variables with random coefficients. *Journal of Research on Educational Effectiveness* 5, 303–332.
- Robbins, H., 1950. A generalization of the method of maximum likelihood: estimating a mixing distribution. *The Annals of Mathematical Statistics* 21, 314–315.
- Robbins, H., 1951. Asymptotically subminimax solutions of compound statistical decision problems. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability I*, 131–139.

- Robbins, H., 1956. An empirical Bayes approach to statistics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 1, 157–163.
- Robbins, H., 1964. The empirical Bayes approach to statistical decision problems. *The Annals of Mathematical Statistics* 35, 1–20.
- Rose, E.K., Schellenberg, J.T., Shem-Tov, Y., 2022. The effects of teacher quality on adult criminal justice contact, NBER working paper no. 30274.
- Savage, L.J., 1951. The theory of statistical decision. *Journal of the American Statistical Association* 46, 55–67.
- Song, J., Price, D.J., Guvenen, F., Bloom, N., vonWachter, T., 2018. Firming up inequality. *The Quarterly Journal of Economics* 134, 1–50.
- Staiger, D.O., Rockoff, J.E., 2010. Searching for effective teachers with imperfect information. *Journal of Economic Perspectives* 24, 97–118.
- Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, 197–206.
- Sterne, J.A.C., Narbord, R.M., 2004. Funnel plots in meta-analysis. *The Stata Journal* 4, 127–141.
- Storey, J.D., 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64, 479–498.
- Storey, J.D., 2003. The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics* 31, 2013–2035.
- Storey, J.D., 2015. qvalue: q-value estimation for false discovery rate control, <https://github.com/StoreyLab/qvalue>.
- Storey, J.D., Taylor, J.E., Siegmund, D., 2004. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66, 187–205.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 267–288.
- Varian, H.R., 2014. Big data: new tricks for econometrics. *Journal of Economic Perspectives* 28, 3–28.
- Wald, A., 1945. Statistical decision functions which minimize the maximum risk. *Annals of Mathematics* 46, 265–280.
- Walters, C.R., 2015. Inputs in the production of early childhood human capital: evidence from Head Start, *American Economic Journal: Applied Economics* 7, 76–102.
- Walters, C.R., 2018. The demand for effective charter schools. *Journal of Political Economy* 126.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.
- Woodcock, S.D., 2008. Wage differentials in the presence of unobserved worker, firm, and match heterogeneity, *Labour Economics*, 15, 771–793, european Association of Labour Economists 19th annual conference/Firms and Employees.
- Xie, X., Kou, S.C., Brown, L.D., 2012. SURE estimates for a heteroscedastic hierarchical model. *Journal of the American Statistical Association* 107, 1465–1479.

Chapter 4

Minimum wages in the 21st century

Arindrajit Dube^a, and Attila Lindner^{b,*}

^aUmass Amherst, NBER, MA, United States, ^bUniversity College London, London, United Kingdom

*Corresponding author. e-mail address: a.lindner@ucl.ac.uk

Chapter Outline

1 Introduction and overview	261	6 Minimum wages in developing countries	348
2 A brief history of minimum wages	264	7 Conclusion and future directions	350
2.1 The rationale for minimum wage policies	264	Appendix A Additional results	352
2.2 The minimum wage debate	268	Appendix B Bias from heterogeneous pre-existing trends: a simulation study	361
3 The wage and employment effects of minimum wages	271	Appendix C Data sources for cross-country Kaitz indices	364
3.1 Wages, employment and labor demand	271	Appendix D Constructing historical QCEW restaurant data	366
3.2 Empirical methods to study the impact of minimum wage policies	274	Appendix E Construction of 60 state-level minimum wage events	368
3.3 Review of the evidence on employment effects	302		
3.4 Effect on total hours	313		
4 Margins of adjustment	314	Appendix F Construction of probability groups using demographic predictors	370
4.1 Review the evidence on various margins of adjustment	314	References	370
4.2 Summary of evidence on margins of adjustment	335	Further reading	383
4.3 Modeling implications and open questions	337		
5 Inequality, distributional implications, and downstream effects	341		

1 Introduction and overview

Minimum wage policies have evolved from their initial introduction, and have become an important tool used by many countries to address various economic and social challenges. The debate and perspectives on minimum

wages have also evolved considerably, influenced by both changing economic theories and empirical research. Although economists were highly skeptical about the benefits of the policy throughout much of the previous century, that view was seriously challenged beginning in the 1990s. Today, in the 21st century, there is widespread interest in making greater use of minimum wages ([Manning, 2021](#)). Countries around the world are introducing or raising their minimum wages at an unprecedented rate. Major international organizations such as the OECD, the IMF, and the World Bank have advocated for the policy as a means to alleviate inequality and even boost employment ([ILO, OECD, IMF, and World Bank, 2012](#)).

In the research community, there has been a proliferation of empirical studies that develop new techniques, take advantage of new data sources, and consider new aspects of the policy. The literature has also expanded from primarily based on evidence from the United States to examining the impact of the policy worldwide. Meanwhile, there have been significant theoretical advances in understanding how labor markets operate and how those models can be applied to better understand the impact of wage floors.

This chapter summarizes the key insights from the literature on minimum wages, with a focus on advancements made over the last two decades. We begin by reviewing the motivation behind the introduction of minimum wages and how these policies have evolved over time and across countries. Next, we provide an overview of the intellectual debate surrounding minimum wages. This topic has been highly controversial among economists and has significantly influenced the field of economics, particularly in economic modeling and research methods.

The relationship between employment and wages is central to testing the neoclassical view of low-wage labor markets. However, the best approach to testing that relationship has been the subject of a long-running debate among marginalist (e.g., ([Stigler, 1946](#))), institutionalist (e.g., ([Lester, 1946](#))), and empiricist (e.g., ([Card and Alan, 1994b](#))) traditions within the discipline. The minimum wage literature has also played a crucial role in shifting economics from a predominantly theoretical discipline to an empirical and data-driven one.

The “new” minimum wage research began in the early 1990s, focusing on studying the policy’s impact using quasi-experimental research designs. In practice, this has involved the use of various difference-in-differences style estimators and their intellectual offshoots. In this chapter, we take stock of these methods and critically examine them, with a focus on the contributions made since the last *Handbook of Labor Economics* review by ([Brown, 1999](#)). The literature has advanced considerably since then.

We discuss the introduction of more sophisticated empirical approaches to studying the impact of minimum wages, such as synthetic control, border

discontinuity design, and refined difference-in-differences estimators like the stacked event-study approach. We also address challenges with common approaches, such as the two-way fixed effects regression, which has been widely used in the literature. Additionally, we cover empirical methods developed outside the U.S. context, particularly those that exploit nationwide variation in minimum wages. Finally, we explore recent advances in understanding the broader impact of the policy on low-wage workers, moving beyond a sole focus on specific groups like teens.

Methodological advances over the past three decades have provided a more nuanced understanding of the effects of minimum wage policies. We begin by offering an up-to-date review of the evidence on employment and wages. Specifically, we summarize estimates of the own-wage employment elasticity, which compares the percentage change in employment to the percentage change in wages—representing the labor demand elasticity in the standard competitive model. We also examine the heterogeneous impacts of the policy across individuals, firms, regions, and depending on the nature of the minimum wage shocks. In this chapter, we use data from U.S. minimum wage events to illustrate some of the empirical challenges in the literature and demonstrate how state-of-the-art approaches can yield reliable estimates for both wages and employment.

However, simply studying the effects of the policy on employment and wages does not fully capture how firms, workers, and markets respond to minimum wage changes. Fortunately, the empirical literature of the 21st century has significantly advanced our understanding of the various margins of adjustment that play a role in response to minimum wage policies. We explore the effects of minimum wages on amenities, other inputs (such as capital and higher-skilled workers), firm entry and exit, output prices and demand, profits, and productivity. While the evidence on some of these margins is still in its early stages, a relatively clear picture has emerged regarding the importance of price pass-through and the productivity-enhancing effects of the policy. We also offer ideas on how these findings could shape future theoretical developments in understanding low-wage labor markets.

Minimum wage policies are a significant tool for combating rising inequality. Considerable attention has been devoted to quantifying how minimum wages shape the wage distribution and their role in the evolution of inequality over time. We review the evidence and methods developed to study the inequality consequences of these policies. Additionally, the minimum wage is often seen as a major redistributive tool outside of the tax and benefit system. In light of this, we discuss the evidence on the distributional implications of the policy.

While the main focus of this review is on understanding the labor market consequences of the policy, there are also broader implications worth considering. An extensive literature has explored the impact of minimum wages on various “downstream” socioeconomic outcomes, such as health, crime, and

education. Although our review cannot delve into the details of these areas, we provide a summary of the state-of-the-art findings regarding these indirect consequences of the policy.

This chapter broadens the scope of existing reviews by examining the impact of minimum wages in various national and economic contexts. The previous *Handbook* chapter on minimum wages by (Brown, 1999) provided an excellent review of the evolution and history of minimum wage policy in the U.S. However, a key feature of 21st-century research is the increasing number of high-quality studies emerging from outside the U.S. This reflects the reality that major policy changes have been instituted globally over the last three decades—including the (re)introduction of minimum wages in places like China (1993), the U.K. (1999), Hong Kong (2010), and Germany (2015)—alongside growing access to high-quality administrative datasets. Throughout this chapter, we highlight evidence from these novel analyses.

Additionally, a growing number of studies are exploring the implications of minimum wage policies in less developed economies. We conclude our review by summarizing the key findings from this emerging literature and discussing how the policy operates in less developed economic contexts.

2 A brief history of minimum wages

2.1 The rationale for minimum wage policies

The motivation behind minimum wage regulations has varied significantly across countries, regions, and historical periods. The first modern-day minimum wages were enacted in New Zealand in 1894 and in the Australian state of Victoria in 1896 (Starr, 1981). The British Parliament adopted legislation in the same spirit in 1909. Initially, these minimum wages focused only on specific industries (e.g., only four trades for Great Britain: chain making by hand, paper-box making, lace finishing, and wholesale tailoring). The main rationale was to prevent and settle industrial disputes and to eliminate “cesweating”—the payment of exceptionally low wages (Webb, 1912; Hammond, 1915; Metcalf, 1999).¹ Proliferation of “sweatshops” in 1890s was a major concern in these countries (Nordlund, 1997). Sweatshops were typically lower-productivity businesses that relied on recruiting cheap labor as their business model, primarily employing women and young workers (including orphans) and paying them substandard wages. More productive companies employing working-class breadwinners were threatened by these business practices and were often supportive of initiatives that sought to reverse these trends.

¹ Although these are the first examples of wage floors in recent history, the concept of a minimum wage goes back much farther. For example, the policy was present in the Hammurabi code (Rositani, 2017).

In the U.S., minimum wage legislation first emerged at the state-level in the early 1910s, beginning in Massachusetts. Concern for workers seen as being the most vulnerable to exploitation by low-paying employers—such as immigrants, women, and children—played a key role in the push for minimum standards through state legislation. However, the United States Supreme Court invalidated most of these laws. Minimum wage legislation was also a part of the National Industrial Recovery Act (NRA) introduced in 1933, but the States Supreme Court later found that legislation unconstitutional.

In 1938, the first federal minimum wage (25 cents per hour, or around \$4.50 in 2023\$) was established under the Fair Labor Standards Act, which also aimed to regulate hours and restrict child labor (Grossman, 1978; Brown, 1999; Fishback and Seltzer, 2021). Although the main objective of the act was to ensure the safety and well-being of workers, the law was also designed to mitigate race-to-the-bottom competition (Newell, 2009). Importantly, the Fair Labor Standards Act had a limited scope and applied only to workers in “inter-state or foreign commerce”, which in practice meant workers in manufacturing and tradable goods sectors. Private sector coverage gradually increased from around 50 % in 1938 to 60 % in the 1960s (Brown, 1999). In 1967, a significant extension in coverage brought in much of the service workforce, such as restaurants, laundries, and retail sectors, reaching almost 80 % of all private sector workers (see (Bailey et al., 2021); Derenoncourt and Montialoux, 2021). With further gradual increases, the coverage eventually reached 90 % of workers (Brown, 1999).²

Latin American countries were also among the first to introduce minimum wage laws, with many, such as Mexico and Brazil, adopting these laws in the 1930s and 1940s (Grimshaw and Miozzo, 2003). Unlike in the United States, constitutions in these countries played a major positive role in establishing pay standards. Many countries in the region had constitutional provisions recognizing the right of workers to receive wages sufficient for a decent standard of living, and explicitly stating the state’s responsibility for setting minimum wages (Collier and Collier, 2002). This state intervention in labor relations was also motivated by an effort to make workers look to the state, rather than to unions, to protect their interests. During this period, Latin American labor unions were often associated with radical political ideologies that challenged state authority (Cook, 2010).

After the Second World War, countries in western Europe started adopting minimum wage policies with examples including Luxembourg (1945), France (1950), Spain (1963) and the Netherlands (1969) (Dolado et al., 1996). This expansion occurred during a period when the idea that workers should have the

² Minimum wage exceptions apply under specific circumstances to workers with disabilities, full-time students, youth under age 20 in their first 90 consecutive calendar days of employment, tipped employees, and student-learners.

right to protection against extremely low wages was gaining traction ([ILO, 2017](#)). However, many European countries with strong union representation, such as Austria, Italy, Denmark, Norway, and Sweden, avoided legislating minimum wages. Instead, these countries opted for collective bargaining, where unions had greater influence in setting labor standards. Unions in these countries were often opposed to minimum wage legislation, fearing that direct state involvement in labor relations could render them redundant ([Meyer, 2016](#)). There were also concerns that state-sponsored wage floors could end up becoming wage ceilings, constraining unions' ability to negotiate better standards. In recent decades, however, the decline in union membership and coverage—especially in the low-wage service sector—has led to a more favorable view of minimum wage regulations among unions in some countries, like Germany, paralleling the introduction of minimum wage legislation.

Minimum wages in Africa were implemented on a significant scale around the 1940s and 1950s. Their pay regulations were closely linked to the colonial powers and were modeled after the wage councils in developed countries where standards were set only for specific trades ([Starr, 1981; Malan, 1978](#)). In Asia, India introduced its minimum wage policy in 1948 (though today provincial-level minimum wages are the relevant standards), Taiwan in 1956, and Japan in 1959. More recently, other countries in the region followed suit, including South Korea in 1988, China in 1993, and Vietnam in 2006. Singapore stands out as somewhat of an outlier that does not have a minimum wage in place. The government has long held the belief that the efficiency cost of a minimum wage is too high and the key to helping low-wage earners is to incentivize them to “climb up the skill ladder” ([Ministry of Manpower in Singapore, 2018](#)).

[Fig. 1](#) plots recent trends in minimum wages for several major economies around the world (Brazil, China, India, Germany, France, Spain, the U.K., and the U.S.). To make minimum wages comparable over time and across countries, we calculate the ratio of the minimum wage to the average wage in each country.³ Unfortunately, average wage data are not always available for past periods, which limits the time frames for which the minimum wage ratio can be calculated. Additionally, we calculate two time series for the United States: one that considers only the federal minimum wage, and another that also accounts for state-level minimum wages.

Among these major economies, the highest minimum wages today are in India and several European countries (France, UK, and Germany), where they stand at around 50–55 % of the average wage. The minimum wage levels in Brazil and

³ The minimum wage is often expressed relative to the median wage for this type of comparison ([Dube, 2019a](#)). Here we use the mean wage instead because it is easier to obtain for several countries. The ratio of the minimum wage to the median (or mean) wage is commonly referred to as the Kaitz index; the original formula also included the coverage rates by sector, but the relevance of that adjustment has greatly diminished as coverage rates today tend to be very high ([Kaitz, 1970](#)).

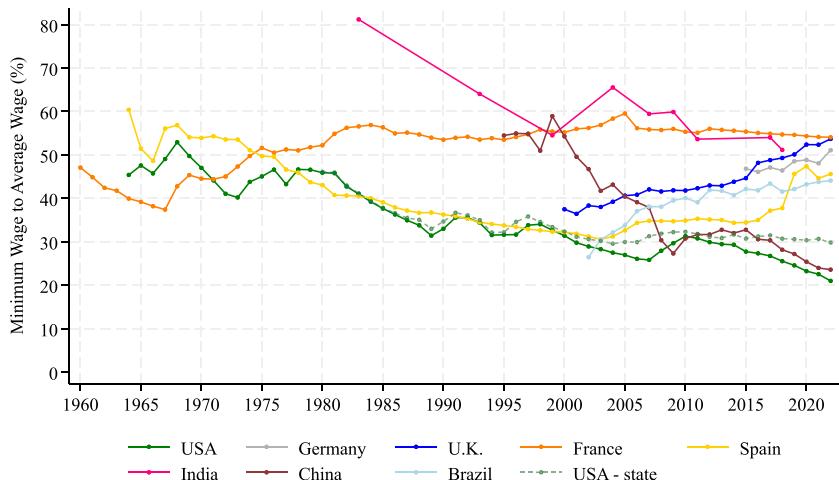


FIG. 1 Evolution of minimum wage-to-average wage ratio (Kaitz index) over time in selected countries. Notes: Data for France, Germany, Spain, UK and US comes from OECD - we scale the reported Kaitz index by 1.1 to account for the difference between average *full-time* wage and overall average wage. The USA-state estimate is a weighted Kaitz index taking into account state minimum wages and populations. For India, we use province-level data used in (Khurana et al., 2023b) and (Khurana and K. Mahajan, 2020), provided to us by the authors, and average these for a national estimate. Data for China is from the Chinese National Bureau of Statistics and the China Statistical Yearbook. Data for Brazil comes from IPEA and the PNAD in later years. See Appendix 10 for detailed explanations.

Spain are slightly lower —around 45 % of the average wage. In contrast, in the two largest economies in the world (China and the U.S.), current minimum wage levels are relatively low, at around 20–30 % of the mean wage.

Interestingly, recent trends differ considerably between these countries. In India, relative minimum wages started to decline in the mid 1980s from a relatively high initial level. Despite this decline, India's minimum wage (relative to the average wage) remains among the highest today. Meanwhile, we have seen a dramatic increase minimum wage levels among a number of European countries (Germany, Spain, UK) and Brazil in recent years. In contrast, the two largest economies (China and the U.S.) have experienced large drops in their levels of minimum wages. Notably, China's minimum wage was around 55 % of the average wage in the mid- 90 s, but it has since fallen close to 20 %. In the U.S., the federal minimum wage was never very high (around 45 % of average wage in mid-1970s) but has fallen dramatically since then. It is important to note that in the U.S., this decline is considerably more muted when factoring in state-level minimum wages, which have become more significant over time. Today, with no nominal increase in the federal minimum wage for over 15 years, the federal minimum wage is largely nonbinding in the U.S., making state and local wage floors the only economically relevant ones in the American economy.

2.2 The minimum wage debate

The intellectual controversies surrounding the minimum wage go back more than a century. At the core of this debate is the question of what the appropriate model for the (low-wage) labor market is and whether government regulations can improve labor market outcomes.

The central debate on the minimum wage revolves around its employment consequences. A long-standing tradition in economics argues that a minimum wage set above the competitive level will reduce employment ([Mill, 1848](#); [Marshall, 1897](#); [Stigler, 1946](#)). These predictions were formalized within the neoclassical framework ([Hicks, 1932](#)). According to this framework, if labor markets are assumed to be 1) competitive with many firms passively accepting prevailing market-level wages, 2) characterized by free entry and exit of firms; and 3) composed of homogeneous workplaces from workers' perspective, then equilibrium employment and wages will be fully efficient.

With two inputs in production (capital and low-skilled labor), the effect of wages on labor demand in this framework can be expressed as follows (see the Hicks-Marshall rule of derived demand in ([Hamermesh, 1995](#))):

$$\frac{\Delta \ln \text{Emp}}{\Delta \ln \text{Wage}} = -(1 - \text{share}_L) * \sigma_{K,L} - \text{share}_L * \eta \quad (1)$$

where share_L is the share of labor in production, $\sigma_{K,L}$ is the elasticity of substitution between capital and labor, and η is the elasticity of output demand. The first term represents the substitution effect: since minimum wages make labor more expensive relative to capital, firms will substitute away from labor. The second term reflects the scale effect. In a competitive environment with free entry and exit, firms have a limited ability to absorb the minimum wage by lowering profits, so they raise prices to cover the increased expenses. This price increase leads to a reduction in consumer demand, which, in turn, decreases the scale of production and employment. The key insight from this framework is that introducing a binding minimum wage will have an unambiguously negative effect on labor demand, leading to lower employment.

Economic thinkers have long pointed out the unintended consequences of minimum wage policies ([Sidgwick, 1886](#); [Marshall, 1897](#)). The minimum wage is often cited as a prime example of how market interference, even with the best of intentions, can do more harm than good ([Stigler, 1946](#)). However, there is also a long tradition of challenging the key prediction of the standard competitive framework. For instance, if firms have considerable market power in setting wages, then there are pre-existing market distortions. In such cases, minimum wages can serve as a second-best policy to alleviate these distortions and move closer to an efficient allocation. A leading example of this is the monopsony framework, where a single firm operates in the labor market. Historically, however, the monopsony model was thought to have limited empirical relevance, as few low-skilled labor markets were characterized by a

single dominant employer with substantial wage-setting power (Stigler, 1946; Brown, 1999).⁴

Another critique came from institutionalist economists such as John Dunlop, Clark Kerr, Richard Lester, and Lloyd Reynolds, who challenged the neoclassical (or “marginalist”) view of the labor market. These economists questioned the notion of unfettered profit maximization and deemed as overly simplistic the view that firms always produce at a point where marginal costs equal the marginal product. Instead, they relied on institutional expertise, detailed case studies, and surveys of management and workers to paint a more empirically-minded picture of firm behavior (Freeman, 1989).

A key example is (Lester, 1946) who surveyed executive officers in various industries. The survey responses overwhelmingly emphasized the significance of current and prospective market demand for the firm's products as the key factor in determining its employment level. In contrast, managers rated the importance of wages as quite low (perhaps surprisingly so). The correlation between concerns about the level of wages and the labor share of production was also quite weak, which goes against the prediction of the neoclassical framework as illustrated in the formula discussed earlier.⁵ Furthermore, (Lester, 1946) pointed out that even though the introduction of federal minimum wages reduced the North-South wage differentials in the manufacturing sector, it did not translate into an employment reduction at Southern firms (if anything, employment increased).

This approach to studying labor demand and the effect of wage shocks faced vehement opposition from neoclassical economists on various grounds (Machlup, 1946; Stigler, 1946; Friedman, 1953). First, it was argued that Lester's survey employed an overly simplistic version of the neoclassical model. The model could be easily extended by incorporating factors such as non-pecuniary benefits and more dynamic decision-making processes on production capacity. Once these modifications are taken into account, the mapping between the neoclassical theory and the survey questions becomes less clear. Second, (Machlup, 1946) and (Friedman, 1953) argued that, similar to how an automobile driver who overtakes a truck, or a professional pool player who strikes a ball, cannot describe the calculations and the physics underlying their behaviors, business executives cannot explain what drives their decisions based on exact

⁴ The idea that wage-setting power can emerge in the presence of strong but imperfect competition began gaining attention in the 1980s. The search literature developed around the notion that while firms compete, search frictions create some job-specific surplus once workers and firms meet. Imperfect substitution between workplaces can also create wage-setting power, much like imperfect substitution between goods in output markets gives firms price-setting power (Card et al., 2018).

⁵ (Lester, 1946) argued that most businesses face a declining marginal cost between 70 % and 100 % of their production capacity, so wage determination is not driven by the marginal cost being equal to marginal benefit. Furthermore, the capacity of production is determined by market demand, and that production cost considerations play only a limited role in that.

mathematical rules. Instead, good business executives, much like a good driver or a good pool player, can make the right decision by sensing profit-maximizing behavior. And if they did not know how to do that—the argument went—they would be replaced by more adept decision-makers. Based on this “as if” reasoning, (Friedman, 1953) argued that as long as a model has good predictive power, the realism of its assumptions does not matter.

By the 1960s, the neoclassical view of the labor market became the dominant approach among economists. Furthermore, the advancement of the human capital theory during the 1960s also led to the presumption that wages simply reflect skills, knowledge, and ability, leaving little scope for the government to effectively intervene in the labor market (Osterman, 2011). This perspective was further supported by a series of empirical studies showing a negative link between employment and wages using time series and plant-level data (Hamermesh, 1995). These findings underscored the predictive power of the neoclassical approach, and led to a consensus regarding the negative implications of minimum wages.

However, this consensus began to dissolve with the emergence of a series of more credible empirical evidence in the early 1990s. Riding the wave of the credibility revolution, a set of papers studied the impact of the policy by applying new and more credible empirical techniques, exploiting state-level variation in minimum wages, and applying difference-in-differences style estimators. Many of these papers found no (Katz and Kevin, 1992) or even positive employment effects (Card, 1992a, 1992b; Card and Alan, 1994b), while others confirmed the negative effect of the policy (Neumark and Wascher, 1992). The previously established consensus broke, and the debate on the employment effects of the policy (re)started.

The initial reaction to the positive—or lack of negative—employment effects in response to the policy was not particularly welcoming.⁶ The vehement opposition to the findings of (Card and Alan, 1994b) reflect that a fundamental issue was at stake (Leonard, 2000). The lack of disemployment effects of the policy challenged a basic prediction of the neoclassical view of the labor market, and suggested that the workhorse model of labor markets might be missing some crucial mechanisms through which labor markets adapt to the policy change.

During the last three decades, there have been three major developments. First, the idea that core theories in economics can be tested and possibly

⁶ A good example of the critical response comes from an opinion piece penned by the Nobel laureate James Buchanan and published in the *Wall Street Journal*, where he wrote: “Just as no physicist would claim that ‘water runs uphill,’ no self-respecting economist would claim that increases in the minimum wage increase employment. Such a claim, if seriously advanced, becomes equivalent to a denial that there is even minimal scientific content in economics, and that, in consequence, economists can do nothing but write as advocates for ideological interests.” (Buchanan, 1996).

rejected has gained greater acceptance within the discipline. As a consequence, the debate on the policy has shifted to a more empirical focus. Second, an extensive literature has emerged on the employment and wage effects of the policy, which we will review in the next section. A fruitful corollary is that this literature has become an arena in which various new empirical techniques have been developed and tested. Finally, new and richer models of low-wage labor markets have emerged (or were rediscovered). This also provided a theoretical basis to studying the impact of the policy in a richer environment, given the ambiguous predictions of the policy on employment and wages. To assess the empirical relevance of these new models, the empirical literature has started to explore outcomes beyond employment, which we discuss in [Section 4](#).

3 The wage and employment effects of minimum wages

3.1 Wages, employment and labor demand

At the risk of understatement, there is an extensive literature studying the impact of minimum wages on employment. In this section, we will provide a guided tour of this literature using the very useful construct of the own-wage employment elasticity (OWE). This elasticity scales the observed (percentage) change in employment by the observed (percentage) change in wages. Therefore, it shows how responsive employment is to a change in wages paid to workers induced by the minimum wage policy—which is equivalent to the elasticity of labor demand in the standard neoclassical framework. However, in some models—such as ones with monopsony power—the labor demand curve is not well defined (or is not binding), even as the OWE remains an economically meaningful measure. For this reason, we use the more catholic employment elasticity terminology in order to avoid imposing model-based interpretations of the empirical evidence.

When quantifying employment responses, the main alternative to using the employment elasticity with respect to own wages is to use the employment elasticity with respect to the minimum wage. The latter approach is common in the literature (e.g., Doucouliagos and Stanley, 2009; [Neumark and Shirley, 2022](#)). However, the OWE offers several important advantages.

First, scaling employment changes by wage changes makes the estimates more comparable across studies and economic contexts. This is important because the same change in the minimum wage may be much more binding for certain economies, demographic groups, industries, or time periods. For this reason, we expect the effect of the minimum wage on employment to depend on the fraction of workers affected by the policy. For example, we anticipate that minimum wages will raise an average teen's wage more than an average prime-age (25 to 54 year old) worker's wage, since teens are relatively more likely to be low-wage workers. Consequently, even if there were *the same* (negative) employment effect of the minimum wage on all affected workers,

the minimum wage employment elasticity for teens would be larger in magnitude than that for prime-age workers. In contrast, scaling employment changes by wage changes accounts for the variation in the bite of the policy across different groups, making the comparison of elasticities across groups much more meaningful.

Second, by focusing on the relative magnitudes and signs of the employment and wage changes, we are better positioned to calculate the policy-relevant trade-off. Policy-makers are interested in the consequences of the policy on workers' wage and employment, but that is difficult to infer from the employment elasticity with respect to the minimum wage. For example, if a 10 % increase in the minimum wage leads to a 1 % reduction employment, it could mean an own-wage employment elasticity of -1 if the minimum wage increases wages by 1 %. This would be considered high. However, the 1 % reduction of employment could be also consistent with an own-wage elasticity of -0.1 if a 10 % increase in the minimum wage leads to a 10 % increase in wages. In the latter case, job loss would be considered small by most, and might be welfare maximizing (see (Lee and Saez, 2012)).⁷

Third, as we explained in the previous section, the minimum wage debate centers around the relationship between employment and wage changes, and their alignment with the prediction of the neoclassical framework. In theoretical discussions, it is often assumed that the link between minimum wage and actual wages is one-to-one, but that assumption rarely holds up in real-world data. Therefore, an additional step is required to shed light on the link between wages and the minimum wage. Formally, this could be conceptualized as a two-step instrumental variable (IV) procedure. Suppose we are interested in estimating the causal effect of wage changes on employment, formally:

$$\Delta \ln \text{emp} = \beta \Delta \ln \text{wage} + \varepsilon$$

We can use the change in the minimum wage policy denoted as ΔMW , which may be either a binary or continuous measure, as an instrument for the change in wages, and implement the following two-step procedure.

$$\begin{aligned} \text{1st stage: } & \Delta \ln \text{wage} = \alpha \Delta MW + \varepsilon \\ \text{2nd stage: } & \Delta \ln \text{emp} = \beta^{IV} \Delta \ln \text{wage} + u \end{aligned}$$

where $\overline{\ln \text{wage}}$ is the predicted wage change ($\alpha \Delta MW$) and β^{IV} identifies the OWE. In the first stage, we estimate the link between the minimum wage and the actual wage earned by a worker. The second stage estimate then provides the link between the wage and employment. The reduced-form relationship

⁷The usual argument in favor of using the employment elasticity with respect to the minimum wage is that the policy-maker can change the minimum wage, but not the wages directly. However, without knowing the sign and magnitude of the wage effects, the policymaker can only evaluate the potential employment costs of the policy without knowing the benefits coming from higher wages.

between the instrument and the employment is given by the following equation.

$$\Delta \ln \text{emp} = \gamma \Delta MW + v$$

If ΔMW is the change in log minimum wage, then γ identifies the employment elasticity with respect to the minimum wage; if it is a binary policy change measure then γ is the reduced form policy effect on employment. This formulation clarifies that the OWE can be thought of as an indirect least squares estimator:

$$\beta^{IV} = \frac{\gamma}{\alpha}$$

With a binary ΔMW , β^{IV} is also the Wald estimate of the LATE (e.g (Angrist et al., 2000)).

$$\beta^{IV} = \frac{\gamma}{\alpha} = \frac{E(\ln \text{emp} | \Delta MW = 1) - E(\ln \text{emp} | \Delta MW = 0)}{E(\ln \text{wage} | \Delta MW = 1) - E(\ln \text{wage} | \Delta MW = 0)} \quad (2)$$

Looking at the OWE through the lens of an IV approach means that the considerations of the validity of the instrument, the role of the compliers, and the measurement error in wages all apply here. For example, whether the minimum wage shock is a valid instrument or not depends on the following assumptions: 1) relevance—there is a strong enough first stage, i.e., there is a clear association between the minimum wage change and change in the average wage of the relevant group,⁸ and 2) exclusion restriction—the minimum wage change only affects employment through wages.

Furthermore, we can also gain insights from the IV literature by recognizing β as a Local Average Treatment Effect (LATE). Consider the case of a binary instrument ΔMW representing a policy change. Here, the treatment is an increase in log wages for “compliers” induced by a minimum wage policy event, and the outcome is log employment.

Note that if we add a set of individuals completely unaffected by the policy to the sample, both the first-stage (α) and the reduced form relationship (γ) weaken, but the ratio of α and γ still stays the same, since the complier share enters (multiplicatively) in both the numerator and the denominator of Eq. (2). This shows that as long as the relevance requirement is satisfied, changing the share of affected workers (say by adding more “never-takers” who are unaffected by minimum wages to the sample) does not alter the fact that β^{IV} is still a valid LATE for the complier sub-population. In contrast, the employment elasticity with respect to the minimum wage (i.e., γ divided by the change in log minimum wage in the case of a binary ΔMW) will be sensitive to adding nontreated workers to the analysis.

⁸ Note that the IV interpretation suggests that the first stage estimate of α needs to be strong enough to avoid a weak instrument bias when estimating the OWE, which is a higher standard than statistical significance at conventional levels.

While recognizing the advantages of the OWE, it is worth noting some important limitations. First, while measuring the impact of the policy on wages helps scale and interpret the employment effects, such a scaling can be imperfect. For instance, changes in non-pecuniary benefits should also be taken into account, both from the policy perspective and the economic debate around the policy. Policymakers presumably care about workers' welfare, and not only about their wages *per se*. A significant decrease in non-wage benefits could fundamentally alter the trade-off between employment and the total (wage and non-wage) compensation of workers. Second, the neoclassical debate focuses on the link between labor cost and employment. However, if firms respond to a minimum wage shock by cutting nonwage benefits, then the change in effective labor costs will be limited. In that case, even if the employment elasticity with respect to *labor costs* is large in magnitude, the employment elasticity with respect to *wages* could be small. We will discuss the evidence on non-wage amenities in [Section 4.1.2](#).

3.2 Empirical methods to study the impact of minimum wage policies

Roadmap. This section covers the main empirical strategies developed to assess the impact of minimum wages on employment and wages (although many of the methodological lessons are relevant for studying other outcomes as well). Until the early 1990s, the predominant empirical strategy to estimate the impact of minimum wages relied on time series analysis, which struggled to isolate causal effects due to its reliance on extrapolating trends without suitable control groups. The limitation of that literature has been discussed in previous reviews, and the presence of serious publication bias has been documented (see Chapter 6 in ([Card and Alan, 1995b](#))). The early 1990s witnessed the emergence of the “new minimum wage literature”—which developed and applied various difference-in-differences style estimation strategies. Needless to say, relying on an explicit control group—instead of only extrapolating past trends—to form a counterfactual was a major innovation in the literature. Still, the choice of control groups is not always straightforward; as we will see, such choices have sometimes been contentious in the literature. Furthermore, the literature is sometimes unclear on exactly what economic object is identified by the proposed difference-in-differences empirical strategy.

3.2.1 Exploiting local variation in the level of minimum wages

The effect of minimum wage policies can be studied most directly by examining local variations in the policy. Since the emergence of the “new minimum wage” literature, the United States has been a fertile ground for research as there is considerable variation in the level of statutory minimums across states and over time. More recently, a new literature has emerged that studies the effect of the policy exploiting variation at an even more granular level (cities or counties).

Case studies. A classic approach in the literature evaluates the impact of a single policy change. The best-known example of such a case study in the modern literature is ([Card and Alan, 1994b](#)), which evaluated the impact of a 19 % increase in the minimum wage that went into effect in New Jersey on April 1st, 1992. Card and Krueger use a first-differenced regression, where the change in employment at fast-food chain restaurants in New Jersey is compared to the change in similar restaurants from neighboring counties in Pennsylvania. More recently, this case study approach has re-emerged in the context of studying city-level minimum wage changes: ([Jardim et al., 2022](#)) study the impact of the policy in Seattle and ([Karabarbounis et al., 2022](#)) in Minneapolis by applying the synthetic control method.

By focusing on one specific event, researchers can often rely on more detailed data sources and implement a clearer identification strategy. For example, ([Card and Alan, 1994b](#)) collect survey data from fast food chains, as detailed firm-level data on employment and wages were not available in the early 1990s. Subsequently, in responding to the critique by ([Neumark and Wascher, 2000](#)), they also access state-level administrative payroll records from New Jersey and Pennsylvania to further improve data quality ([Card and Alan, 2000](#)). In a similar vein, ([Jardim et al., 2022](#)) and ([Karabarbounis et al., 2022](#)) use administrative data on hourly wages that are only available for a few states in the United States. By focusing on a specific event, we may be able to better understand the context of the minimum wage change and consider the biases induced by shocks that coincided with the policy implementation. For instance, in the context of Seattle's minimum wage ordinance, there has been an extensive discussion on how Seattle's exceptionally strong labor market boom may contaminate the estimated employment effects ([Dube and Lindner, 2021](#)).

TWFE-log(MW). At the same time, an individual case study has the obvious limitation that it is just that—a single case. Statistical inference for single cases is challenging, and findings from cases have limited generalizability. Being able to go beyond individual cases, therefore, has obvious appeal.

Parallel to the case-study approach, an extensive literature emerged applying a two-way fixed effect empirical strategy that involves unit- and time-specific fixed effects to control for unobserved factors (e.g., [Neumark and Wascher, 1992](#); [Orrenius and Zavodny, 2008](#); [Allegretto et al., 2011](#); [Meer and West, 2016](#); [Neumark et al., 2014](#)). This is often thought to be a generalization of the case-study approach which pools various individual cases; but as we will see later, this is not quite right, as this approach introduces additional (sometimes opaque) assumptions.

Usually, the following regression is estimated in some form, which we will refer as the TWFE-log(MW):

$$y_{st} = \beta \ln(MW_{st}) + X_{st}\Lambda + \gamma_s + \delta_t + \nu_{st} \quad (3)$$

where y_{st} is typically the employment rate or average wages in state s at time t , in levels or in logs, and $\ln(MW_{st})$ is the log of the binding minimum wage (greater of state or federal minimums). As we shall see below, in most applications, the outcome variables are restricted to certain demographic groups such as teens or young workers, as those groups are the most exposed to the policy. The regression includes state fixed effects (γ_s), time fixed effects (δ_t), and typically time-varying, state-specific control variables (X_{st}).

The inclusion of state and time effects implies that within-state variation in the level of minimum wages over time is used for identification. In principle, this allows researchers to better isolate the causal impact of minimum wage changes by comparing changes within the same state rather than across different states with potentially confounding differences. A key advantage of this design is that it pools together many minimum wage changes. Peculiar shocks coinciding with one particular event are averaged out, and so those shocks have less influence on the estimates. The design also leads to an estimate for an “average” minimum wage shock with greater external validity than the findings based on specific events. Additionally, since the TWFE is just a panel regression, standard inference is available when variation across many minimum wage changes is exploited.

However, it turns out that this design also has some serious problems when used to study minimum wages. One issue is the sensitivity of results to how unobserved time-invariant heterogeneity is controlled for. For example, (Neumark and Wascher, 1992) show that the estimated negative impact in the TWFE specification becomes positive in a first-differenced (FD) specification (Cengiz et al., 2019). (see Appendix G) also find similar differences between the TWFE and FD estimates for the 1979 and 2016 period (though not after 2000). Such differences are surprising since the two models should converge to the same estimated coefficients in large samples (with many states and time periods) (Cengiz et al., 2019). linked it to violations of parallel trends in the early part of the sample. The large discrepancy also highlights that specifications need to be carefully scrutinized and additional falsification tests are required to pick the better estimation model.⁹

Second, a contentious issue emerged about what to control for in the regression Eq. (3). The original study by (Neumark and Wascher, 1992) on the impact of the minimum wage on teens includes estimates with and without the proportion of teens in school. The relationship between minimum wage and employment is positive

⁹ The TWFE and FD estimators are the same if $T = 2$. If $T > 2$, TWFE and FD will not be identical, but both are unbiased and consistent according to standard assumptions. One source of discrepancy could be the presence of lagged effects, which the FD can miss. However, the use of distributed lags does little to resolve the discrepancy, as demonstrated in (Cengiz et al., 2019). When TWFE and FD specifications yield different results, it is difficult to determine which one is more accurate. As a statistical matter, if T is large, and N is small, FD (especially with lags) is likely preferable, as TWFE is sensitive to small violations in assumptions (we will see more on this). If T is small, N is large: relative performance will depend on the autocorrelation of the error term (for more on this, see (Wooldridge, 2012), page 490).

without controls, whereas the sign flips and becomes negative when controlling for that variable. Controlling for the proportion of teens in school might be simply a “bad control” because it picks up a key mechanism through which the minimum wage affects teen labor markets. In any case, the sensitivity of the estimates could also reflect that minimum wage shocks are implemented parallel to other policy changes or in an unusual economic environment. In the presence of such shocks, understanding how those shocks affect low-wage labor markets is essential, and simply controlling for those shocks might not be enough. As we will also see below (Section 3.2.2), controlling for covariates in a difference-in-differences context requires care when the effect of treatment is heterogeneous.

In the follow-up literature, the debate about controlling for time-varying observable factors largely moved to how to control for the unobserved time-varying differences across states, e.g., by including state-specific time trends (Neumark et al., 2014; Allegretto et al., 2011; Manning, 2021). reviewed this debate and estimated seven different TWFE specifications using the Current Population Survey data between 1979 and 2019. He found the employment estimates to be sensitive to the inclusion of state-specific time trends: the estimated teen employment falls without time trends and increases when variation in state time effects is taken into account. Concern about violations of parallel trends led to the use of other ways to form counterfactuals such as synthetic control, factor models, or border discontinuity (Dube, Lester et al., 2010; Neumark et al., 2014; Totty, 2017). However, the TWFE panel regressions were somewhat of a black box, making it difficult to diagnose the problems.

In addition to these issues, a growing number of studies have highlighted the fragility of TWFE estimates under staggered adoption and heterogeneous treatment effects (De Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Goodman-Bacon, 2021). Although the TWFE estimates an average of the treatment effects, the weights are sometimes negative, creating possibly serious bias. In order to address these concerns, (Cengiz et al., 2019) proposed an event study design. As we shall see, an event-based approach also helps shed light on exactly what variation is being used for identification, allowing us to better see where violations of parallel trends occur.

Event study. A pooled event study is the proper generalization of the individual case study approach pioneered by (Card and Alan, 1994b). This can be seen transparently using the stacked event study approach proposed by (Cengiz et al., 2019). They study all sizable minimum wage changes between 1979 and 2016.¹⁰ For each prominent state-level minimum wage change, h , they create an event-specific dataset, which includes observations from the

¹⁰(Cengiz et al., 2019) define prominent state minimum wage changes as being more than \$0.25/hour and affecting more than 2 % of the workforce; there were 138 such events. Note that some of these events are part of multiple year phase-ins of large minimum wage changes. In the our empirical exercise below, we combine these multiple-year phase-in events into one, reaching 60 major events.

treated state $s^*(h)$ around the policy change (e.g., between three years before, \mathcal{T} , and five years after $\bar{\mathcal{T}}$ the date of the policy change). The event-specific dataset also includes observations from all the “clean control” states that did not experience any major changes in minimum wages over the event window. Therefore, just like for an individual case study, a single event-specific dataset includes one treated state, where the counterfactual outcome is obtained by averaging across all untreated (control) states. Here τ represents time relative to the event date, so $\tau = 0$ is when the minimum wage is raised in state $s^*(h)$. Subsequently, all the event-specific data are stacked. Once this stacked dataset is created, a number of approaches can be applied.

First, one can assess the impact of the policy across all events, by running the following regression:

$$y_{hst} = \beta \times \mathbb{I}[\tau \geq 0] \times D_{hs} + \Omega_{hst} \Lambda + \gamma_{hs} + \delta_{ht} + \nu_{hst} \quad (4)$$

where y_{hst} is the outcome in the dataset belonging to event h in state s and at event time τ . The regression includes γ_{hs} , event-specific state fixed effects, and τ_{ht} , event-specific time effects, ensuring that all the identification comes from within an event. The treatment variable equals one for the treated state and zero otherwise, $D_{hs} = \mathbb{1}[s^*(h) = s]$. Hence, the coefficients β show the change in outcome in the treated states relative to the change in outcome in the control states. The Ω_{hst} are covariates. For example, (Cengiz et al., 2019) control for small minimum wage changes that took place in the control states.¹¹ Finally, standard errors should be clustered at the state level; this automatically takes into account that some observations from a state may show up more than once in the stacked dataset.

It is worth highlighting that this event study regression aligns events by event time (τ) and not by calendar time (t), so it is equivalent to a setting where events happen all at once. Moreover, by dropping all states with any events within the 8-year event window from the “clean” control set, we guard against bias due to heterogeneous treatment effects. Therefore, the negative weighting issues from staggered implementation (see, e.g., (De Chaisemartin and d’Haultfoeuille, 2020; Goodman-Bacon, 2021; Borousyak et al., 2024)) do not arise here. The proposed regression estimation provides us with a (positively) weighted average impact of the minimum wage across events. These weights are proportional to the variance of the treatment; therefore, the simple stacking approach produces a variance-weighted ATT (VWATT) (Gardner et al., 2024; Dube et al., 2023). As shown in (Dube et al., 2023), one can obtain exactly the same estimate as the stacked approach by using the local projections DiD (LP-DiD) regression in the original panel (i.e., not stacked) dataset. LP-DiD simply regresses:

$$\overline{A\bar{y}_{st}} = \beta \times D_{st} + \delta_t + \nu_{st} \quad (5)$$

¹¹ (Cengiz et al., 2022) offer a refinement to this approach that discards units with small changes from the control group. This is preferred (when feasible) because it requires fewer assumptions.

where $\overline{\Delta y_{st}} = \frac{1}{6} \sum_{k=0}^5 (y_{s,t+k} - y_{s,t-1})$ is a post- minus pre-treatment difference in outcome averaged over the post-treatment window, and $D_{st} = 1$ equal to one if state s is newly treated at time t .¹² The estimation sample only includes observations that are either newly treated on date t (i.e., $D_{st} = 1$) or “clean control units” as defined above. Below, when we show empirical results for the event study analysis, we implement it using Eq. (5).¹³

Second, being clear about the exact treatment dates, the event window, and the admissible control units (as is the case with the stacked dataset) also allows us to naturally obtain event-by-event estimates. This can be done by separately estimating a β_h for each event, which is equivalent to replacing β with β_h in Eq. (4), or changing the sample to include only one treated unit at a time in Eq. (5). However, given that each β_h is identified from a single treated unit, the usual approach of clustering the standard errors at the state level does not produce the correct statistical inference. This can be addressed by using the (Ferman and Pinto, 2019) standard errors when calculating event-specific estimates. The event-by-event estimates can then be used to transparently study whether the impact of the policy is heterogeneous and/or driven by a few outlier events. One can also explore the impact of different weighting schemes when aggregating events (e.g., instead of weighting by the sample size, variance, etc., equal weights can be applied).

Third, this event study approach allows us to cleanly estimate the changes in outcome around the timing of the policy change by running an event-study type regression:

$$y_{hst\tau} = \sum_{k=\underline{T}}^{k=\bar{T}} \beta_k \mathbb{I}[\tau = k] \times D_{sh} + \Omega_{hst} \Lambda + \gamma_{hs} + \tau_{h\tau} + \nu_{hst} \quad (6)$$

Analogously, separate regressions could be estimated for each k using Eq. 5 with long-differenced outcomes to recover the same dynamic responses:

$$y_{s,t+k} - y_{s,t-1} = \beta_k \times D_{st} + \delta_t + \nu_{st} \quad (7)$$

where the estimation sample only uses observations that are newly treated or are clean controls.

Looking at the coefficients before the reform allows us to test for pre-existing trends, which is the standard approach to assess the parallel trend

¹² We use the following convention when defining the treatment variable D : if D has a time subscript, it equals to one only on the date the state first receives treatment, otherwise it equals to zero. If the treatment variable has no time dimension, as in Eq. (4), then it is only a function of the state, s , and is equal to either zero or one in all periods.

¹³ As also shown in (Dube et al., 2023), it is straightforward to re-weight the observations to produce an (equally weighted) estimate of the ATT; this is available in the Stata and R code of lpdid. The equally weighted version of the LP-DiD is numerically equivalent to the DiD estimate of (Callaway and Sant'Anna, 2021). We also demonstrate this equal weighting below in our empirical implementation.

assumption—the treated and control states would have evolved similarly in absence of the minimum wage change. If there are concerns about the presence of pre-existing trends, one can balance the treatment and control by matching on covariates, or by implementing a synthetic control approach (e.g., see ([Allegretto et al., 2017](#)) for such an application studying the impact of city-level minimum wages). Alternatively, matching and synthetic control methods can be combined as in ([Kellogg et al., 2021](#)).

Until relatively recently, distributed lag versions of TWFE were thought to identify dynamic responses to treatment using a difference-in-differences design, similar to [Eqs. \(6\) or \(7\)](#). However, this is not the case, and the dynamics implied by a distributed lag TWFE model can be quite misleading due to issues like negative weighting ([Sun and Abraham, 2021](#)). This further underscores the importance of using a proper difference-in-differences event study design instead of TWFE-based strategies.

Event-study vs. TWFE-log(MW): an empirical exploration. A key difference between the event-study and the TWFE-log(MW) specification is that event studies focus on employment changes within the event window, while the TWFE-log(MW) compares employment changes across different time-periods, even if these periods are far apart. The TWFE assumes that the parallel trends assumption holds across the full sample; this is a stronger assumption than those made by an event-study design, which assumes parallel trends holds within the event window. These properties are important to understand for why TWFE-log(MW) produced sensitive results, especially for samples that include the 1980s and 1990s. Below we present evidence to underscore this issue and shed some light on the parallel trends assumption.

Much of the variation used for identifying minimum wage effects in the U.S. comes from the comparison between two groups of states: 35 states that have raised their minimum wages at some point between 1980 and 2019 (we will call them the “ever-treated” group), versus 15 that did not (the “never-treated” or the “control” group). To be clear, there are differences in timing in when these 35 states raised their minimums. However, in practice they mostly did so in 3 waves: a small wave in the late 1980s, and two much larger waves in the 2000s. These waves emerged as a natural response to prolonged periods in which the federal minimum wage remained stagnant. As a result, most of the minimum wage “events” happened in the 1998–2006 and post-2013 periods, with a smaller set of events in the late 1980s and a handful in the early 1990s including the famous 1992 New Jersey increase studied by ([Card and Alan, 1994b](#)). Panel A of [Fig. 2](#) plots the nominal minimum wages in the two groups of states, showing the clear patterns of these three waves.

Since much of the identifying variation comes from comparing these two groups of states, a natural question is how similar these groups are. As it turns out, the assumption that the “ever-treated” states have followed parallel trends as compared to the “never-treated” states seems much more plausible since the late 1990s. In contrast, for the period between early 1980 to mid-1990s, labor

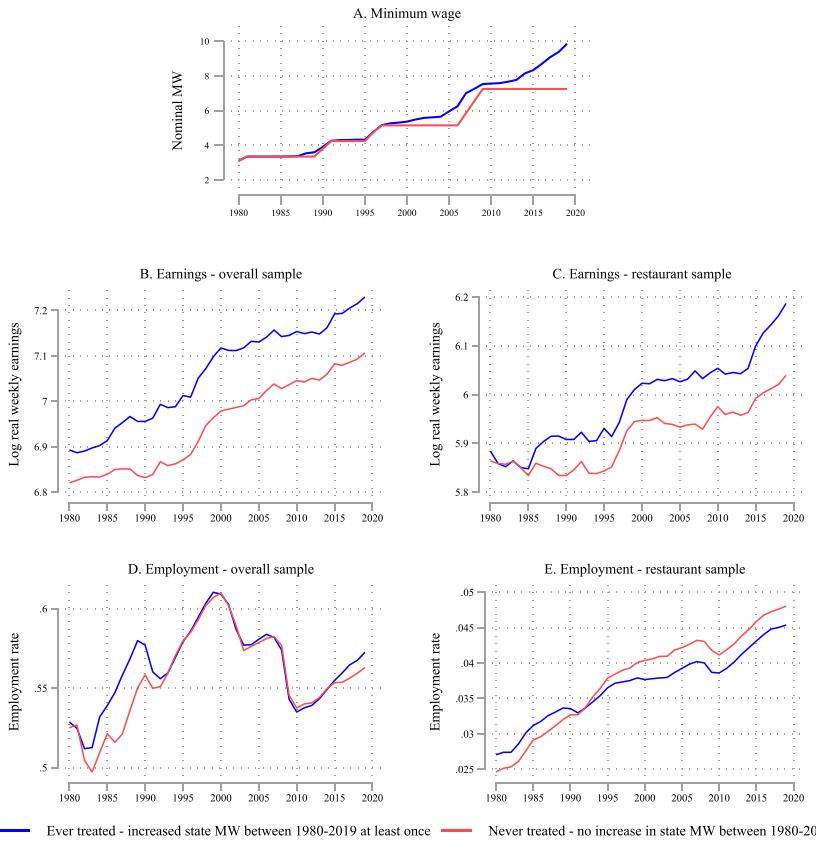


FIG. 2 Minimum wages, earnings and employment in ever-treated and never-treated states. Notes: These figures show the evolution of the minimum wage, earnings, and employment in ever treated and never treated states - the 35 ever treated states had at least one increase in MW above the federal level between 1980–2019, while the 15 never treated states had no such increase. Panel A shows how the nominal minimum wage changed and B and C show the progression of log average weekly earnings (in 2023 dollars) for the full sample and the restaurant sample respectively. Panel D and E show the log employment rate for these samples. All averages for the two groups are weighted by state population. For both earnings and employment in the restaurant sample, post 1990 data is as per the NAICS 3-digit definition. From 1980–1990, the data has been imputed using a scaling factor derived from the SIC/NAICS ratio in 1990. In addition, missing restaurant sector data for Alaska, Delaware, and Rhode Island in the 1990s has been imputed from sub-sector data using a scaling ratio of sector to sub-sector numbers. (details are in Appendix 11).

markets in these two groups of states greatly diverged. We can see this in Panels B and D of Fig. 2 which plot the overall average weekly earnings (in logs) and per-capita employment for the two groups of states.¹⁴ As compared to the “never-treated” states, the “ever-treated” states’ employment rate

¹⁴ This is based on annual data from Quarterly Census of Employment and Wages (QCEW).

rose sharply in the 1980s, and then fell back in the 1990s. What is striking is that this happened even though the minimum wage barely differed between these two groups over this period (see Panel A). This pattern strongly suggests a violation of parallel trends in the potential outcomes for employment, because: (1) in general we do not expect *overall* outcomes to be move very much from minimum wages given the small share of workers affected by the policy; (2) there was not much change in the relative minimum wage between these two groups; (3) but even so, average earnings rose more strongly in “ever treated” states in the late 1980s.¹⁵ In contrast, after the mid-1990s, overall employment and earnings trends were much more alike between the two groups of states. As the figure shows, this is also precisely the period in which we see major differences in the minimum wage between the “ever treated” and “control” states.

Of course, most minimum wage analysis is not based on overall employment (with a few exceptions like ([Meer and West, 2016](#))). Turning to the low-wage labor market, Panels C and E of [Fig. 2](#) plot log earnings and employment in the restaurant sector, a much studied group.¹⁶ Here we see that restaurant employment fell greatly in the “ever treated” states between the early 1980s and the late 1990s, even though the relative minimum wages were virtually identical in 1980 and 1995, consistent with a violation of parallel trends. More generally, there is very little relationship between the timing of the minimum wage changes and the changes in employment.¹⁷

In contrast, over the post-1998 period, we see that the gap in restaurant employment as well as overall employment to be highly stable between the two groups of states—even as minimum wages and average restaurant earnings varied greatly. The post-2013 period, in particular, led to a very durable gap in wages between the “ever-treated” and “never-treated” states, but without any visible gap in restaurant employment. Concretely, between 2000 and 2019, average restaurant earnings increased by 0.07 log points more in the “ever-treated” states, while log employment increased by 0.02 log points more.

We can use these state-by-year data to shed light on how estimates based on the popular TWFE-log(MW) regression with long panels are sensitive to

¹⁵The regional labor market differences were recognized at the time. For example, ([Freeman, 1990](#)) analyzed the effects of this boom in states such as California, New York, and Massachusetts (dubbed the “Massachusetts Miracle”) on young Black workers.

¹⁶The data from the QCEW for restaurants is available based on NAICS codes only after 1990; SIC-based data is available through 2000. For restaurants, the differences in employment levels between SIC and NAICS is modest. We construct a state-level NAICS prediction for the pre-1990 period as described in [Appendix D](#).

¹⁷While the federal minimum wage fell in real terms over the 1980s, note that the average restaurant earnings were similar in the “ever-treated” and “never-treated” states in the early 1980s; this means that the “bite” of the federal minimum decline did not differ very much across these two groups; and hence the relative reduction in restaurant employment is unlikely to reflect federal changes.

violations of parallel trends in the 1980s and 1990s. To start, we present event study estimates for restaurant earnings and employment from Eq. (5) in Table 1, Panel A, column 1.¹⁸ These are based on 60 combined events (which combines successive minimum wage increases into a single event). The details of the event construction are provided in Appendix E; Appendix Fig. A1 plots the events by state and time; the number of control units, the duration of the post-treatment windows, and the size of the increase by the year of the event are reported in Appendix Table A1. Averaged over a maximum of 6 years following the initial increase, treated states saw a rise in minimum wages of 0.26 log points, a rise in restaurant earnings of 0.03 log points, and a fall in employment of -0.001 log points (statistically indistinguishable from zero). This yields a small OWE of -0.03 (s.e.=0.15).¹⁹ Focusing on the 46 events after 1998 yields an OWE of 0.15 (s.e.=0.14). (See Appendix Fig. A4 for the event study plots.) The other columns present outcomes for other groups of workers than in restaurants, including for teens, as well as a demographic prediction-based group (“high recall”) that captures most minimum wage workers. As we can see, the event-based OWE estimates for these groups are small or positive, falling between -0.07 and 0.34.

In contrast, when we estimate static TWFE-log(MW) regressions in the 1980–2019 sample, we obtain estimates of 0.28 for restaurant wage and -0.15 for restaurant employment, leading to an OWE of -0.52 (s.e.=0.35). Importantly, when we estimate distributed lag versions of this regression (with 5 annual lags and 2 leads), the implied dynamic responses from the distributed lags regression also look very different from the event study estimates, as shown in Appendix Fig. A2. In contrast, as Table 1 also shows, the TWFE estimates in the post-1998 period are more similar to the event study estimates, with an OWE of 0.5 (s.e.=0.41). For the 1980–2019 sample, the “long run” effect on restaurant employment is large, negative, and statistically significant (panel B of Appendix Fig. A2); but for the post-1998 sample, it’s very close to zero (Panel D). Other low-wage groups such as teens (Table 1, column 4) follow a similar pattern, with the TWFE estimates from the 1980–2019 sample indicating larger dis-employment than either event study estimated or TWFE estimates from 1998 and later.

Why do the TWFE estimates differ so much from event study estimates when including data from 1980s and 1990s? After all, there is not much geographic variation in the minimum wage during this earlier period, as we have already seen. The reason is because inclusion of the early period affects

¹⁸ The one difference from Eq. (5) is that we have 6 years as a *maximum* post-period, which for certain events may be truncated due to either federal minimum wage increases or end-of-sample considerations. See Appendix E for details.

¹⁹ Recall that these estimates based on Eq. (5) represent a variance-weighted ATT. Instead, if we equally weight each event by re-weighting, the OWE is 0.02 (s.e.=0.16) as shown in Appendix Table A3, panel A, column 5.

TABLE 1 Effects of increased MW - Event study and TWFE estimates.

	(1) Restaurant sample	(2) Overall sample	(3) High prob sample	(4) Teen sample	(5) High recall sample	(6) Low prob sample
A. 1980–2019 event study (Total events: 61); $\Delta \log MW = 0.197$ (s.e. = 0.008)						
Log wages	0.031 ^b (0.005)	0.015 ^b (0.004)	0.037 ^b (0.005)	0.063 ^b (0.014)	0.020 ^b (0.007)	0.007 (0.005)
Log employment	-0.002 (0.004)	0.005 (0.004)	-0.006 (0.010)	0.004 (0.017)	0.007 (0.007)	0.003 (0.002)
OWE	-0.079 (0.135)	0.327 (0.202)	-0.158 (0.265)	0.070 (0.273)	0.319 (0.308)	- (-
B. 1998–2019 event study (Total events: 47); $\Delta \log MW = 0.206$ (s.e. = 0.007)						
Log wages	0.035 ^b (0.005)	0.014 ^a (0.006)	0.038 ^b (0.006)	0.067 ^b (0.019)	0.020 ^b (0.007)	0.009 (0.005)
Log employment	0.003 (0.005)	0.007 (0.004)	-0.002 (0.012)	0.007 (0.018)	0.009 (0.008)	0.003 (0.002)
OWE	0.087 (0.125)	0.487 ^b (0.169)	-0.048 (0.309)	0.098 (0.268)	0.476 (0.373)	- (-)

C. 1980–2019 TWFE: implied $\Delta \log MW = 1$

Log wages	0.284 ^b (0.035)	0.160 ^b (0.044)	0.191 ^b (0.021)	0.413 ^b (0.038)	0.041 ^a (0.020)	0.044 (0.028)
Log employment	-0.184 ^a (0.081)	-0.028 (0.033)	-0.217 ^b (0.048)	-0.237 ^b (0.059)	0.039 (0.037)	0.040 ^a (0.015)
OWE	-0.647 ^a (0.272)	-0.177 (0.201)	-1.136 ^b (0.277)	-0.574 ^b (0.158)	0.951 (1.125)	- -

D. 1998–2019 TWFE: implied $\Delta \log MW = 1$

Log wages	0.218 ^b (0.025)	0.033 (0.029)	0.194 ^b (0.036)	0.407 ^b (0.056)	0.081 ^a (0.035)	-0.022 (0.020)
Log employment	0.074 (0.072)	0.056 (0.037)	-0.109 ^a (0.045)	-0.032 (0.062)	0.023 (0.028)	0.023 ^a (0.011)
OWE	0.341 (0.320)	1.687 (1.281)	-0.561 ^a (0.227)	-0.079 (0.152)	0.283 (0.354)	- -
Data source	QCEW	QCEW	CPS	CPS	CPS	CPS

Each cell represents a coefficient from a separate regression. The columns are sub-samples, while the rows are outcomes. The first 2 rows in panels A and B report event study estimates from Eq. 5. Events are starting years of prominent MW increases as explained in Appendix 12. The estimated $\Delta \log MW$ reported in Panels A and B is the coefficient from the same regression with $\log MW$ as dependent variable. This coefficient does not depend on sub-samples. The restaurant sample has 58 out of 61 events in panel A, and 45 out of 47 events in panel B, as we don't have NAICS restaurant data in Rhode Island and Delaware before 2000. Thus, MW increase events in 1986 and 1999 in Rhode Island, and in 1999 in Delaware get dropped for that sample. The first two rows in panels C and D are TWFE estimates from regressions of the outcome on $\log(MW)$ minimum wage (as per Eq. 3). As the independent variable itself is $\log MW$ for this regression, the implied effect on $\log MW$ is 1. The third row in each panel shows own wage elasticity which is the ratio of the employment estimate to the earnings estimate. For wages, the restaurant and overall samples use real weekly log earnings, while all other samples use real log hourly wage. Standard errors in parentheses are clustered by state, and all regressions use state population weights.

^a $p < 0.05$.

^b $p < 0.01$.

the estimation of the state fixed effect—and hence the “baseline” employment—in somewhat non-transparent ways. As we add older pre-treatment observations, the regression adds comparisons between the post-treatment period and the newly-added older observations (e.g., from the 1980s).

As a demonstration, we can show how adding older data affects the estimates in an otherwise clean event study-type design. To begin, we consider the 2010–2019 period with a sharp increase in the minimum wage. Here we take the same 35 “ever-treated” states ($D_s = 1$) and 15 “never-treated” states ($D_s = 0$) and consider a regression:

$$y_{st} = \beta_\tau \times Post_t \times D_s + \gamma_s + \tau_t + \nu_{st} \quad (8)$$

Here $Post_t = 0$ if the year is 2010–2013, while $Post_t = 1$ if year is 2014–2019. 20 of the 24 states experiencing a prominent increase in minimum wage in the 2010–2019 period first did so in 2014 or later, which is why we use 2014 as the cutoff point for $Post$. This formulation is very close to the event study estimates we saw from Panel A, column 1 of [Table 1](#), except that to keep things even simpler, we are using an early event-start date as the common event-start date for all 35 states to avoid any staggered adoption issue. For this reason, we call this a “quasi event study”. The wage and employment estimates can be thought of as being for the “intent to treat” since some of the post periods are not treated; but the OWE estimates can be interpreted just as before. As first shown in Column 1 of the Appendix [Table A2](#), the log minimum wage gap between the two groups rises by 0.15 points (goes up from 0.05 to 0.20) in the $Post$ period, a large increase. Consequently, log restaurant earnings increase by 0.04, while log employment increases by 0.005, with an OWE of 0.13 (s.e.=0.27), close to the post-1998 event study estimate of 0.15 (s.e.=0.14) in [Table 1](#).

Now consider the same regression, but expand the Pre period to include all years between 1980–2013 when the gap in log minimum wage between the “ever treated” and “never treated” groups was smaller than the original pre-treatment (2010–2013) gap of 0.05. This adds in years 1980–1988, 1990–2000, and 2009 to the Pre period. Doing so does not substantially change the treatment effect on log minimum wages (estimate changes from 0.15 to 0.17). However, now restaurant wage effects are somewhat larger and employment effects are much more negative, leading to an OWE of −0.76 (s.e.=0.36). The reason for this is obvious from the inspection of [Fig. 3](#): the “ever-treated”–“never-treated” gap in restaurant employment was much larger in the 1980s and early 1990s. Simply expanding the Pre period in the event study by adding data from the 1980s and 1990s leads to an OWE much closer to the TWFE estimate of −0.52 from the 1980–2019 period ([Table 1](#)). This exercise also offers intuition behind why the TWFE regression elasticity becomes much more negative from adding data from 1980 to 1999—a period with relatively few minimum wage events. Namely, it affects the baseline employment through the estimation of the state fixed effects.

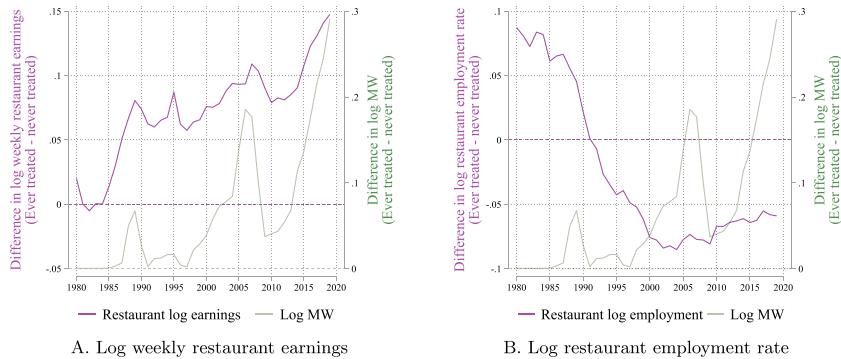


FIG. 3 Difference in minimum wage, restaurant employment and restaurant earnings between ever-treated and never-treated states. Notes: This figure plots the population-weighted difference in restaurant log earnings/log employment between 35 ever treated and 15 never treated states on one axis, and the difference in log MW between these states on the other. Log average weekly earnings (in 2023 dollars) are shown in Panel A, and log employment rate in Panel B. Ever treated states had at least one state MW increase, over and above any federal increases from 1980–2019. Never treated states had no such increase.

Although in this case Fig. 3 transparently indicates a likely failure in the parallel trends assumption during the 1980s and 1990s, a distributed lag version of the TWFE regression does not necessarily clarify that point. This is related to the fact that the weights put on different events by the TWFE model can be possibly negative; this, combined with some violations in parallel trends, can lead to potentially spurious impulse response estimates in the distributed lag TWFE model. For example, the implied dynamics of restaurant employment around the event date according to the TWFE model (see Appendix Fig. A2) look very different from those in the event study (see Appendix Fig. A4).

Importantly, the problem with the TWFE model is not primarily a matter of using continuous versus discrete treatment measures. For example, take exactly the same 60 events D_{st} used in the event study analysis, and cumulate those over time within the state as follows: $\tilde{D}_{st} = \sum_{k \leq t} D_{st}$. Now estimate a TWFE regression (static or distributed lag) using \tilde{D}_{st} using the 1980–2019 sample. This TWFE-binary regression generates similarly spurious effects as the TWFE-log(MW) regression, and the associated impulse response looks very different from the event study estimates we saw before, with an OWE of -1.43 (see Column 7 of Appendix Table A3, and Appendix Fig. A3).²⁰ This is because both regressions use all possible long difference comparisons over the full 40 year window, as opposed to the event study estimates from Eqs. (4) or (5) which make comparisons only within the actual event window.

²⁰ For an example of a recent paper which implements such a TWFE-binary regression with distributed lags over a long horizon, see (Jha et al., 2024).

To further clarify the nature of the bias in the TWFE estimates, in Appendix B, we conduct a Monte Carlo simulation with two treatment cohorts and staggered adoption. The simulation shows that when there are violations of the parallel trends assumption early on in the sample, TWFE distributed lag models can produce highly misleading impulse responses, even when the true treatment effects are constant and equal to zero. Importantly, in our simulations, these violations occur well before the treatment events and outside of the event window. As a result, while the TWFE model's stricter parallel trends assumption (that it holds across the full sample) leads to biased conclusions, modern difference-in-differences event study approaches—such as ([Callaway and Sant'Anna, 2021](#); [Cengiz et al., 2019](#)) or ([Dube et al., 2023](#))—yield unbiased estimates and correctly show a null effect.

Overall, our take-away from the recent literature is that an event-based difference-in-differences design (which has become increasingly popular) is a much more promising approach and should be the default among researchers using a DiD approach. Estimates from a TWFE panel regression with log minimum wage have proved opaque and fragile, and we think they are unlikely to yield convincing evidence going forward. This should also be kept in mind when reviewing the older employment evidence presented in [Section 3.3](#), much of which has been based on the TWFE model. It is also good to recognize that much of the debate around the sensitivity of results to specifications (as recounted in ([Manning, 2021](#)) took place because the early sample with substantial state-level variation (i.e., 1980s and 1990s) had some serious violations in parallel trends, making identification genuinely more difficult. Thankfully, these issues have become much less pronounced over the past 25 years.

Finally, this discussion has focused on the U.S., which is unique in having about half the country with no change in minimum wages for over 15 years, while some parts have enacted very large increases. At the same time, the issues around event-based estimates and the TWFE are broadly relevant also for other contexts with sub-national minimum wage variation, including Canada, China, and India.

3.2.2 Other considerations for the comparability of treatment and control groups

In this section, we discuss further considerations about adding controls to the regression and choosing the treated and control units in the regression.

Controlling for covariates. It is common to include covariates such as state-level unemployment rate ([Neumark et al., 2014](#); [Allegretto et al., 2011](#)), housing price index ([Clemens and Wither, 2019](#)), state-level income per capita ([Clemens and Michael, 2021](#)), or pre-treatment sectoral and demographic characteristics of areas ([Dube and Lindner, 2021](#)) in the estimating equations. Some of these are contemporaneous (time-varying) controls, while others are pre-determined.

In general, as common as it is, the inclusion of variables that could be directly affected by the policy (e.g. state unemployment rate) should be avoided. It is better to construct covariates of state-level business conditions that exclude low-wage workers (for example, (Cengiz et al., 2022) use the employment rate of those who are unlikely to be minimum wage workers as a control in one specification).

However, even when a variable is not directly affected by minimum wage treatment, there are subtleties about how best to control for it in estimation (see the discussion in (Roth et al., 2023)). A common approach is to add time-invariant (e.g., pre-determined) covariates (X_j) interacted with time effects. However, this can bias estimates when the effect of the minimum wage varies by the level of X_j , because the $X - by - t$ interactions would constitute “bad controls,” as they are also mediators. However, this problem can be addressed by “regression adjustment,” where the coefficients on the $X - by - t$ interactions are estimated using the clean control sample only.²¹ Alternatively, researchers can use propensity score reweighting of control units to match the averages of X in treatment units.

Time-varying covariates can be handled in an analogous fashion. If the effect of the minimum wage treatment is heterogeneous by a time-varying control X_{jt} , simply including X_{jt} as a covariate in the regression poses a problem for similar reasons as described above. For example, if the minimum wage effect on employment differs during periods of high aggregate unemployment, then the inclusion of aggregate unemployment (or unemployment rate of individuals unlikely to be minimum wage workers) as a covariate can lead to a biased estimate, since the unemployment variable is also a mediator. Regression adjustment is one option. As an illustration, column 6 of Table A3 reports estimates from Eq. (5), but uses regression adjustment to additionally control for the change in employment rate of a “low-probability” group (whose construction using demographic predictors is explained later in this section). One can also use a propensity score re-weighting approach (see (Caetano et al., 2022)).

Border discontinuity. Since (Card and Alan, 1994b)'s comparison of the border region of Pennsylvania to New Jersey, the minimum wage literature has sometimes considered nearby areas to be more reliable control groups (Dube, Lester et al., 2010). generalized this approach by considering all contiguous border county pairs (BCP) and estimating a regression with pair-specific time effects. This essentially only considers variation within each BCP, washing out variation between pairs. The analysis in (Dube, Lester et al., 2010) was based on using log minimum wage and a modified TWFE model, TWFE-BCP (where the time effects are interacted with BCPs). However, the findings—strong

²¹ In STATA this can be implemented using teffects,ra command. For staggered DiD designs, something analogous is done within new DiD packages such as by (Callaway and Sant'Anna, 2021), (Borusyak et al., 2024), or (Dube et al., 2023).

earnings effects and small employment effects in the restaurant sector—are similar when we implement the BCP design using a stacked event study regression (see Appendix G of (Cengiz et al., 2019)). Appendix G of (Cengiz et al., 2019) also shows that the TWFE-BCP estimates are not sensitive to the inclusion of the earlier data from 1990–1997, in sharp contrast to the classic TWFE. Taken together, these results are consistent with the idea that the looking across the border can help mitigate the bias from violation of parallel trends in the longer panel with data from the 1990s.²² More recent work using the border discontinuity design has found broadly similar results as using a standard DiD (Coviello et al., 2022). This is not surprising given the lack of pronounced differences across states raising or not raising minimum wages after 1998, as shown above.²³

However, there are disagreements about whether the border discontinuity design is superior to other strategies, including comparing all counties within the U.S (Neumark et al., 2014; Jha et al., 2024). On many observable dimensions, BCP's appear to be more similar to each other than a randomly chosen pair of two counties from the United States (see, e.g., (Dube et al., 2016)). There is also indication from other, non-minimum wage settings that BCPs can help mitigate endogeneity problems in policies (for example, (Boone et al., 2021) show this in the context of unemployment benefit extensions during the Great Recession.).

At the same time, as we see, there are several possible limitations to this design. In some cases, neighboring areas are particularly different from areas farther away. One example is estimating the effects of city-wide minimum wages. Most cities are surrounded by areas that are less urban and possibly less comparable than other cities farther away. This suggests that the border design is better suited for studying state-level policies than city-level ones. The second concern relates to the external validity. Border counties experiencing minimum wage changes in some cases may be quite different from interior counties experiencing the same event, so the causal effects found there may not generalize (in the western United States, most border areas tend to be more sparsely populated).

²² (Jha et al., 2024) find that using a TWFE-CZ design and cross-state commuting zones when the data stretches back to 1990 yields more negative estimates for restaurant employment than from a TWFE-BCP design. However, (Dube et al., 2010) show that these TWFE-CZ results are driven by the same violation in parallel trends in the 1990s that we have discussed above; and use of clean event studies suggests small employment effects for either BCP or CZ design. The substantive stake from whether to use TWFE-BCP versus TWFE-CZ is minor, since we have already seen the issues with TWFE models and data from earlier years. Nonetheless, this does point to the practical issue that results can sometimes be sensitive to “bandwidth choice” or the definition of local labor markets.

²³ The border discontinuity design has also been used to study minimum wage policies outside of the U.S., like the study of Indian minimum wages by (Soundararajan, 2019).

The third—and possibly the most important—issue is that there may be cross-border spillovers from the treated area that can contaminate the control area, violating the stable unit treatment value assumption (SUTVA). One possible mechanism is the flows of workers across the border. There is some evidence that this might play a role, at least when looking very close to the border, from (Jardim et al., 2017). At the same time, this is a case study of a city minimum wage, which is probably not a good setting for this design to begin with. In contrast, (Coviello et al., 2022) do not find that cross-border worker flows respond to minimum wage increases in the retail sector. Given the uncertainty, empirical spillover tests are therefore important to implement when using the border discontinuity design. Relatedly, estimates that are more robust to cross-border spillovers can be used, such as using “control rings,” motivated by the “donut holes” in regression discontinuity designs. That is, one can exclude areas that are very close to the border where cross-border spillover may be more likely. (Most spillover estimates die out farther away from border (e.g., (Jardim et al., 2017))). This could retain the benefits of proximity while mitigating some of the possible bias arising from spillovers (for examples of this, see (Boone et al., 2021) and (Jha et al., 2024)).

3.2.3 Methods to estimate the overall effect of the policy

So far in this section, we have focused on constructing the counterfactual for minimum wage treatment. However, there is also the question of how to define the treatment group. The key choices here revolve around: (1) whether to define the group narrowly or broadly, and (2) what measures to use in defining the group.

As we will see in our review of the evidence on employment effects in Section 3.3.2, most studies in the existing literature consider relatively narrow groups, driven largely by the need to find a group where the policy is clearly binding. For example, a large part of the literature focuses on teens because it is easier to detect a wage effect for this group than for broader demographic groups, such as all workers, or all workers without a college degree. This is especially true during periods when the bite of the minimum wage has been low, making it binding for a relatively small share of the workforce. However, the external validity of estimates based on teens is questionable, particularly since teens constitute a steadily declining share of the low-wage workforce.

Another approach has been to study narrow industries like restaurants, which employ a sizable share of the minimum wage workforce. As we saw in Section 3.2.1, (Table 1, Panel A, column 1) there is a clear increase in restaurant earnings after minimum wage events. However, many minimum wage workers are not in the lowest-wage sectors. To assess the impact on those workers, a different approach is needed.

One option is to look at the entire population or workforce. This may be feasible when studying large, highly binding, changes. For example, the 1967

increase in coverage and level of the U.S. minimum wage was substantial enough to have a detectable effect on the average wage. This allowed researchers to estimate an OWE for all working age (16–64) or prime age (25–54) individuals (Bailey et al., 2021; Derenoncourt and Montialoux, 2021). Studying recent U.S. minimum wages in low-wage areas, (Godey and Michael Reich, 2021) is also able to obtain a clear earnings estimate for workers with high school or less education.

However, studying such broad groups of workers is not feasible for all minimum wage increases studied in the literature, especially when studying changes at a relatively low level of the minimum wage. For example, while the event study estimates for overall earnings are statistically significant for the full period (Table 1, Panel A, column 2), they are not for events before 1998 (wage estimate of 0.017 (s.e.=0.10) not reported in the table). The existing literature has addressed this issue in three different ways: frequency distribution and bunching; demographic prediction and machine learning; and incumbent workers.

Bunching in the frequency distribution. When the minimum wage is raised to, say, \$12 an hour, and there is full compliance, jobs that would pay below \$12 absent the policy would now “disappear” from that part of the frequency distribution. The number of these “missing jobs” (missing from the part of the distribution below \$12) as a share of the overall workforce gives us an indication of the “bite” of the policy change, or the fraction of workers directly affected by the minimum wage increase. Some of these missing jobs are in fact upgraded—now paying at or slightly above \$12—creating a “bunching” or “excess jobs” in the frequency distribution (here excess is relative to the counterfactual frequency). In contrast, some of the missing jobs may be destroyed if they are no longer profitable, and do not show up in the frequency distribution at all. Yet other jobs paying at or slightly above \$12 may come into existence if, for example, an employer forced to pay a higher wage is better able to retain or recruit workers—thereby adding to the excess job count. This implies that by subtracting the missing jobs count from the excess jobs count provides us with an estimate for the overall effect of minimum wage policy on the number of low-wage jobs.

The key to this exercise is estimating the counterfactual distribution absent the minimum wage increase. An early approach by (Meyer and David, 1983) used parametric assumptions about the wage distribution to infer the employment effect; but these proved fragile (Dickens et al., 1998). Instead, (Cengiz et al., 2019) proposed estimating how the entire wage frequency distribution (that is, jobs per capita by wage bin) changes in response to a minimum wage increase using a DiD design. Fig. 4 shows the bin-by-bin estimates using this approach applied to 138 increases in state-level minimum wages during the 1979–2016 period. There is clearly a reduction in jobs that pay less than the new minimum wage (bin labeled “\$0”); at the same time, there is an increase in jobs that pay at or within a few dollars of the new

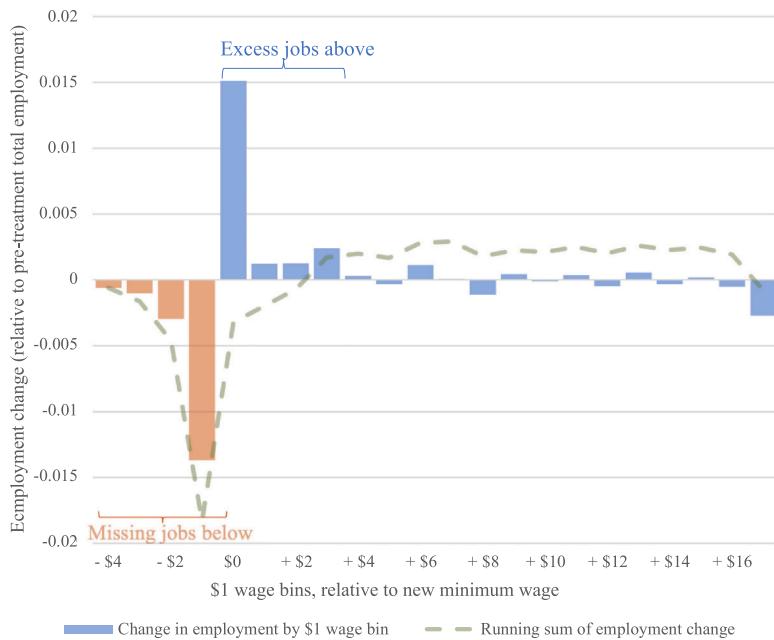


FIG. 4 Impact of minimum wages on the frequency distribution of wages. Source: Based on estimates underlying Fig. 2 in (Cengiz et al., 2019). Notes: The figure shows bin-by-bin employment changes resulting from 138 state-level minimum wage changes between 1979 and 2016. For each dollar bin (relative to the new minimum wage), the bars show the estimated average employment changes in that bin during the five-year post-treatment relative to the total employment in the state one year before the treatment. The dashed line plots the running sum of employment changes up to the respective wage bin.

minimum wage. When these are added up (as in the running sum plotted as a dashed line), the total effect on low-wage jobs is found to be close to zero (e.g., at +\$2 or higher).

There are a number of advantages to this approach. First, by locally evaluating excess and missing jobs around the minimum wage, we can study the impact for a wide variety of workers. Subject to sample size issues, it is in principle possible to get a “first stage” wage effect for low-wage workers even from groups with high average wages (such as college graduates). Second, beyond a narrow range above the new minimum, it is highly unlikely to have a major causal impact on the upper tail of the frequency distribution. Therefore, the lack of such upper-tail effects serves as a useful check on the research design. For example, Fig. 4 shows no such upper tail effects when studying state minimum wage increases. In contrast, the presence of such upper tail effects provides a warning against mistaken interpretation; an example of this comes from evaluating city-side minimum wages (e.g., Seattle), as discussed in (Dube and Lindner, 2021). Third, this approach allows joint estimation of the impact on low-wage jobs as well as on wages earned by those workers, including possible

spillover effects. Importantly, the spillover effects estimated using this method do not assume away possible disemployment effect, in contrast to approaches using the wage density only. Finally, excluding changes in the upper tail can also improve the precision of the employment estimates. A growing number of papers have implemented a frequency-distribution-based approach to estimate the overall impact on jobs (e.g., (Derenoncourt and Montialoux, 2021; Jardim et al., 2022; Wursten and Reich, 2023; Azar et al., 2023).

Demographic prediction-based approach. While the frequency distribution approach estimates the overall impact on low-wage employment, it is less useful for studying other outcomes, including labor force participation, job finding, or non-labor market outcomes such as health. An alternative approach uses demographic predictors to construct a likely low-wage group. An early example was (Card and Alan, 1995a), who used a probit model, CPS-based demographic variables, and a judiciously chosen set of higher order interaction terms to predict low-wage workers. Their “high probability” group successfully captured a somewhat broader set of workers than just using teens or other age/education based groupings.

More recent work by (Cengiz et al., 2022) builds on this in two ways. First, they use modern machine learning methods to optimally choose predictors, partitions, and interactions to classify workers. Second, they show how one can construct a broader “high recall” set of workers that in the U.S. can capture the vast majority of likely minimum wage workers while still obtaining a clear wage effect. They also show that optimally constructed groups are able to capture much larger shares of true minimum wage workers (a high “recall rate”) while maintaining the same share of workers classified as minimum wage workers who truly are so (the “precision rate”) than ad hoc groupings often used in the literature such as teens, individuals under age 30 without a high school degree, etc.

We can illustrate this approach using the event study design described previously. Appendix F describes the construction of a “high recall” sample covering 75 % of minimum wage workers using CPS-based demographic predictors and a gradient boosting algorithm. For this “high recall” group, in Table 1, Panel A, column 5, we see that there is a clear effect of the minimum wage events on log wages (an increase of 0.02), allowing for a meaningful assessment of the employment effect. The “high probability” sample (Column 3) shows a larger wage effect (0.03), hence focusing on lower-skilled workers; this group allows better assessment of possible labor-labor substitution. Neither group shows a statistically significant or sizable employment effect; the OWE’s are -0.07 for the “high probability” groups of 0.34 for the “high recall” group. Finally, the low-probability group (column 6) offers a useful falsification test: reassuringly, there is no statistically significant wage or employment effect for this group. Alternatively, the employment rate of the “low-probability” group can be used as a control in the event study design through regression adjustment. This is implemented in the Appendix Table A3, column 6; comparing columns 5 and 6,

we find that in this case, the additional control has very little impact on the wage or employment estimates.

As a practical matter, when implementing this approach, it is useful to use a model to directly predict who is a low-wage worker—as opposed to predict an individual’s wage, and then classify workers based on their predicted wage. With the latter approach, the model predicts the mean wage conditional on covariates and misses out on who might be making low wages conditional on covariates.

Incumbent worker based approach. Finally, a number of papers have constructed treatment groups by considering the wages that workers were earning before the minimum wage. This approach focuses on “incumbent workers” who were earning low-wages. For example, if the minimum wage is rising to \$15 per hour, we can select workers who were initially earning at or below that level (or perhaps below a slightly higher threshold, such as \$17). We could then compare outcomes for similarly earning workers in jurisdictions that raised the minimum wage versus jurisdictions that did not. This approach requires panel data that allows us to track the outcomes of workers over time; examples of papers using this approach include ([Clemens and Wither, 2019](#); [Jardim et al., 2022](#); [Hampton and Evan Totty, 2023](#)). We consider a related design of comparing these workers to others earning slightly higher wages in [Section 3.2.4](#). We also discuss some issues with this empirical design there.

3.2.4 Exploiting nation-wide variation in the level of minimum wages

Many countries set minimum wages at the national level only, meaning that there is no local variation. An extensive literature leverages major changes in such nation-wide minimums to study the effect of the policy, most notably in Germany, Portugal, the U.K., as well as the US. Here, we review the key methodologies used in such a setting.

Most studies of these changes in national minimum wage exploit the variation in exposure to minimum wage between groups. These groups could be defined on the basis of locations, individuals, firms, or industries. The key idea is to compare the evolution of employment and wages for the more exposed (“treated”) groups with the less exposed (“control”) groups. The definition of a group varies across applications, so we begin with a general discussion followed by group-specific aspects.

Studies of nation-wide minimum wages typically estimate the following class of difference-in-differences regression:

$$y_{gt} = \gamma_g + \beta BITE_g \times Post_t + \delta_t + \varepsilon_{gt} \quad (9)$$

where y_{gt} is the outcome variable (usually wages—to assess the first stage—and employment) in group g at time t , where δ_t and γ_g are group and time effects, respectively, and $Post_t$ is equal to one after the reform. The key variable, $BITE_g$, corresponds

to the exposure based on the pre-reform period. For example, $BITE_g$ may represent the average wage or the fraction of minimum wage workers in group g before the policy change. Alternatively, it could represent the expected change in average wages within group g if there were full compliance with the policy, sometimes referred to as the “gap” that must be closed to comply with the policy. In the U.S. context, some studies have leveraged pre-policy variation in statutory minimums across jurisdictions to study the effect of federal minimum wage changes (e.g., (Clemens and Wither, 2019; Derenoncourt and Montialoux, 2021)). In many applications, $BITE_g$ is calculated based on multiple pre-policy years to avoid issues related to mean reversion. Furthermore, some applications explore non-linear relationships between exposure ($BITE_g$) and the outcome (y_{gt}), or estimate the relationship non-parametrically.

The standard assumption in these types of DiD estimators is that of parallel trends: that changes in potential outcomes for highly exposed and less exposed groups would be the same in the absence of the policy. While this assumption is not directly testable, it is possible to check whether it holds during the years prior to the policy by studying pre-existing trends.

It is also important to clarify what this exposure design identifies and under what conditions. For example, one must assess whether the spillover effects from more-treated to less-treated units could create problems. In general, such spillover effects would imply a violation of the SUTVA, thereby biasing the estimates. As a practical matter, a serious violation of the SUTVA is unlikely in most applications given the low share of minimum wage workers in the economy. Even in the presence of significant spillovers, any effect on the outcomes of untreated groups will be limited if, for example, that group represents a much larger share of the population than the treated group. Nonetheless, in some situations where the analysis sample of both the treated and control groups is restricted to highly exposed industries or locations, a violation of the SUTVA could become a more important issue.

Another (related) concern about identification arises if workers are able to move across groups—for example, from highly exposed locations to less exposed ones, or from highly exposed firms to less exposed firms. In such cases, again, the resulting violation of the SUTVA means that the relative change in group-level employment would not represent the true causal effects of the policy.²⁴

Given these considerations, it is also useful to analyze the time series evolution of the outcomes for the least exposed or untreated groups around minimum wage hikes. A significant change in their outcomes could reflect some spillover effects and the violation of the SUTVA. Of course, such changes might simply reflect the presence of aggregate time effects, which the

²⁴ The presence of such worker mobility indicates that the untreated groups are also affected by the minimum wage, implying a violation of SUTVA. However, estimates can be adjusted to account for the effect on untreated firms. If the treated population represents a small share of the total population, this adjustment will have only a minor effect on the estimates.

DiD empirical strategy addresses under the parallel trends assumption. However, being transparent about the outcomes for the untreated groups helps us better understand the economic environment in which the minimum wage is instituted. It also allows us to assess whether the shocks affecting the untreated groups are likely to have a similar impact on the treated groups (i.e., the parallel trend assumption holds).

In addition to parallel trends, the identification of national minimum wage effects also relies on assumptions about the correlation between heterogeneous treatment effects and the exposure measure $BITE_g$. Consider a scenario in which the responsiveness of low-exposure groups to the minimum wage differs from that of high-exposure groups. For instance, when exploiting regional variation, it is possible that low-wage locations have higher labor market concentration, resulting in less negative employment effects in these regions—say exactly zero. In contrast, high-wage locations might be less concentrated, leading to strongly negative employment effects there. In this case, a regression estimate from the comparison of low- and high-wage regions would suggest a strong *positive* relationship between employment change and exposure, even though the *average* causal impact of the policy is negative.

A similar issue arises when exploiting variation across industries (see page 20 of ([Krueger, 1994](#)) for a discussion of this issue). Imagine an industry A (e.g., local services) that is highly exposed to the minimum wage hike but faces an inelastic output demand—so there is no change in employment. In contrast, industry B (e.g., manufacturing) is higher-paying with lower exposure but faces a very elastic output demand; therefore, the employment effects there are negative. In this hypothetical example, the average causal effect of the policy on employment is negative. However, when employment changes are compared across industries, there is a positive association between exposure and employment change.

These examples highlight the key issue with the exposure design in the presence of treatment effect heterogeneity. To see the problem more formally, suppose that the true data-generating process has the following form:

$$y_{gt} = \gamma_g + \beta_s BITE_g \times Post_t + \delta_t + \varepsilon_{gt} \quad (10)$$

where g reflects all the groups (e.g. 4-digit industries), while s reflects a higher group level aggregation (e.g., s could be 1-digit industries). The above equation can be rewritten as:

$$y_{gt} = \gamma_g + \beta BITE_g \times Post_t + \delta_t + (\beta_s - \beta) BITE_g \times Post_t + \varepsilon_{gt} \quad (11)$$

If we estimate the above equation by simply regressing $BITE_g$ on y_{gt} , then the error term is going to be $u_{gt} = (\beta - \beta_s) BITE_g \times Post_t + \varepsilon_{gt}$. We would obtain a consistent estimate of β only if u_{gt} is uncorrelated with $BITE_g$. Note

that this can hold only if the bite ($BITE_g$) is independent of the treatment effect ($\beta - \beta_s$) otherwise, we have a classic endogeneity problem. This issue has also received attention in other contexts (e.g., impact of medicare on healthcare expenditure). Consistent with the previous two stylized examples, (Sun and Jesse, 2022) formally show that when effects are heterogeneous across groups, the above estimator can perform very poorly and may produce an estimate that falls outside the range of individual treatment effects.

The above derivation also provides a hint on how to deal with heterogeneity. One solution proposed by (Sun and Jesse, 2022) is to parameterize this heterogeneity. In our case, this would mean estimating Eq. (10) and allowing β to vary across the broader group s . A different approach would include time-varying s -specific fixed effects, ψ_{st} , in the regression. Given that $(\beta_s - \beta)BITE_g \times Post_t$ is st -specific, this would address the endogeneity issue. The inclusion of such fixed effects also allows us to control for s -specific trends over time.

This second approach has a close connection to the “grouping” estimator of (Blundell et al., 1998). In many applications, groups are created by splitting the data based on two or more dimensions. For example, groups can be created based on location and age. The following regression estimates the effect of the policy

$$y_{rat} = \gamma_{ra} + \beta BITE_{ra} \times Post_t + \theta_{rt} + \psi_{at} + \varepsilon_{rat} \quad (12)$$

where y_{rat} is the outcome in location r of a worker with age a at time t , γ_{ra} are age-by-location group fixed effects, θ_{rt} are location-time effects, and ψ_{at} are age-time effects. This grouping regression design only exploits changes within location and within education; therefore, any heterogeneity along these dimensions will not lead to a bias in the estimates. This design also controls for location or age-specific trends.²⁵

A key limitation of these approaches is that they do not allow heterogeneity within s (e.g., at the g -level). As a result, researchers need some prior knowledge about the potential dimensions of heterogeneity in order to implement these methods effectively. In the absence of such prior knowledge, one can estimate the effect of the policy while allowing for arbitrary heterogeneity if there are some groups at the g -level that are not exposed to the policy. In this case, it is recommended to implement a fuzzy difference-in-differences design that conducts pairwise comparisons between unexposed units and exposed units (De Chaisemartin and d'Haultfoeuille, 2018). In case of the grouping estimator described above, this approach will only identify the effect of the policy for those ages and locations where unexposed units exist.

Regional bite. In practice, a popular approach in the literature is to exploit local variation in exposure to the minimum wage (Card, 1992b). ’s seminal

²⁵ Sometimes this grouping estimator is implemented in a first-differenced form (see (Manning, 2021)): $\Delta y_{rat} = \beta BITE_{ra} + \theta_r + \psi_a + \varepsilon_{ra}$.

paper studied the impact of the 1990 increase in the federal minimum wage on teenage workers by comparing low-, medium-, and high-wage states. The basic idea behind this approach is that the same nation-wide minimum wage will be more binding in low-wage locations like Louisiana than in high-wage locations like California.

Follow-up work has applied this approach to various contexts such as the introduction of the U.K. national minimum wage (e.g. (Stewart, 2002; Dolton et al., 2012).), the German national minimum wage (Caliendo et al., 2018), the 2007–2009 U.S. minimum wage increase (see (Clemens and Wither, 2019)), and the 1966 expansion of the U.S. minimum wage coverage (Bailey et al., 2021). All of these studies estimate a regression similar to Eq. (9), where the group, g , is defined by the region.

In many applications, the variation across locations is not large enough to produce first-stage wage effects when the outcomes y_{gt} are calculated for the total workforce. One solution to this problem is to restrict the sample to specific groups, such as lower-skilled workers or teens. In this case, it is important that $BITE_g$ is calculated for the same sub-sample of workers.

Another way to address the lack of first-stage effects is to take advantage of additional exposure variation by creating subgroups within locations based on worker characteristics, such as age or education (see, e.g., Kertesi and Köllö, 2003; Dickens et al., 1999; Dube, 2019a). However, most applications do not often include education-time or age-time fixed effects in the regression. Notable exceptions are (Manning, 2021) and (Dube, 2019a). Given the issues regarding heterogeneous treatment effects, when exploiting sub-regional variation in the bite, it is advisable to use (or at least report) estimates with the inclusion of time-varying age, education, and location effects.

A recent paper by (Giupponi et al., 2024) further develops the idea of using low-wage locations as a control group for high-wage locations. Their method combines the regional approach with the frequency distribution approach discussed in Section 3.2.3. They estimate the effect of the minimum wage throughout the entire skill distribution (instead of using the change in high-skilled workers employment as a control for low-skilled ones). They compare the change in the number of workers employed in a given wage bin at low-wage (treated) locations to the change in the number of jobs with similarly skill requirements at high-wage (control) locations. Once the effect—corresponding to the difference between high- and low-wage locations—is estimated for all job types (i.e., the full frequency distribution), they can zoom in and study the employment and wage changes of lower-skilled workers. Therefore, instead of using higher-skilled workers as a control for lower-skilled workers, they compare change in low-skill employment across high- and low-wage locations; this still allows them to use employment changes of higher-skilled workers for falsification.

A subtle but important aspect that is often overlooked is understanding the characteristics of “compliers” whose wages are shifted by treatment $BITE_g$.

These are the workers who would be below the minimum wage at certain low-wage locations but above the minimum wage at high-wage locations. Any identification strategy using regional variation necessarily estimates the average treatment effect only for this “complier” population, and not for those who are always treated—since the latter group earns below the minimum wage at every location. Estimating an overall average treatment effect on the treated—including the always-treated—would require some extrapolation from the complier groups.

It is also important to keep in mind that the estimated differences between high- and low- wage locations could reflect changes in worker mobility following the minimum wage. Most studies that directly test migration responses do not find any change in internal migration in response to the minimum wage (see [Section 4.1.5](#)).

Firm-level exposure. Another approach in the literature exploits firm- or establishment-level variation in the exposure to the minimum wages (e.g ([Card and Alan, 1994b](#); [Machin et al., 2003](#); [Draca et al., 2011](#); [Harasztsosi and Lindner, 2019](#)).). To apply this approach, panel data on firms are needed so that pre-policy exposure of firms can be calculated.

A key advantage of this design is that it allows us to examine the margins of adjustment of businesses that are the most strongly impacted by the minimum wage. Since these firms play a key role in absorbing the minimum wage shocks, this design can provide valuable insights into understanding the mechanisms through which the policy influences low-wage labor markets. Here, we make two points about interpreting the evidence from this cross-firm design.

First, it is important to emphasize that the estimated effects reflect the relative change in employment in highly exposed firms, which is not the same as the employment effects on workers. If workers in highly exposed firms are laid off but find jobs at less exposed firms, then the firm-level exposure design will overestimate the negative consequences of the policy on workers. Recent findings by ([Dustmann et al., 2022](#)) and ([Rao and Risch, 2024](#)) provide evidence for such a reallocation mechanism in the German and U.S. contexts.²⁶ Therefore, we recommend that firm-level estimates of employment be supplemented with worker- or local-level estimates that guard against such biases.

Second, the firm-level analysis inherently focuses on incumbent firms that existed before the policy change, as minimum wage exposure can be only calculated for these firms. This limits the “representativeness” of the analysis, although newly entering firms typically represent a small share of the workforce, at least for a few years after the policy change. Still, when implementing

²⁶The presence of reallocation leads to a violation of the SUTVA. However, in a context where a small share of the total workforce is in highly exposed firms, the change in outcomes for untreated firms will be limited.

this design, it is advisable to complement the analysis with evidence on the impact on firm-level entry rates to assess the importance of this concern. In Section 4.1.4, we review the evidence on firm entry, and show that although the effect on entry rate is inconclusive, most studies suggest only a limited change in firm entry following minimum wage hikes.

Worker-level exposure. Some studies exploit exposure to the minimum wage by pre-policy wages. The key idea is to compare the employment trajectories of workers earning below the new minimum wage to those of workers earning slightly above it (see, e.g., (Currie and Bruce, 1996); Abowd et al., 2000). Implementing this approach requires panel data of workers. This method is similar to the “incumbent workers” design (described above), which compares initially low-wage workers across treated and untreated areas. But different from that design, the worker-level exposure approach can be used in setting where there is only a national level minimum wage change.

This approach focuses on the wage and employment implications of individuals who are most strongly affected by the minimum wage. The estimates provide worker-level responses to the policy (instead of firm- or local-level ones) and speak more directly to the welfare consequences. However, it is important to remember that simple comparisons of high- and low-wage workers’ trajectories can be misleading. Workers at the bottom of the wage distribution typically experience greater wage growth and less stable employment than those higher in the wage distribution, even in the absence of any policy change (Ashenfelter and Card, 1981).²⁷ To address this issue, pre-policy years are often used to estimate and control for “placebo effects” in the employment trajectories of low- and high-income individuals (see (Clemens and Wither, 2019); Dickens et al., 1999; Dustmann et al., 2022)).

Finally, the standard design inherently focuses on incumbent workers with pre-policy wages, often missing the responses among those who are unemployed or out of labor force. To study the unemployed, researchers need to “predict” what their wage would be if they were employed. This prediction can be made in various ways (see our discussion Section 3.2.3). In practice, predictions are typically based on demographics (see (Currie and Bruce, 1996)) or on wages from pre-unemployment spells (see (Clemens and Wither, 2019)).

3.2.5 Exploiting minimum wage exemptions

In many contexts, there are minimum wage exemptions for small firms, long-term unemployed, or younger workers. A strand of the literature leverages this variation, which inherently creates “treated” (non-exempt) and control (exempt) workers. The most common approach exploits age cut-offs in the

²⁷ The higher wage growth at the bottom could reflect mean-reversion: many workers at the bottom might have had a temporary negative shock and then revert back the following year. Additionally, it could reflect that wages at the bottom can only go up (not down). In terms of employment, the lower stability reflects that minimum wage jobs tend to have high turnover.

level of the minimum wage (see e.g., ([Kabátek, 2021](#))), but cut-offs in the duration of unemployment that define who is considered long-term unemployed are also used (([Umkehrer and Berge, 2020](#))).

These papers effectively demonstrate the employment and wage differences around the cut-off by applying a regression discontinuity design (see, e.g ([Kreiner et al., 2020](#)).). However, it is unclear what is identified by comparing the outcomes of very similar workers who are subject to different wage-floors. The fact that a firm might choose the cheaper worker when faced with two similar workers earning different wages tells us more about the substitutability of these two workers and labor market frictions that might dampen this substitution. It is difficult to extrapolate these findings and derive implications for minimum wage policies that apply to all individuals. This discussion points more broadly to the importance of clarifying what theoretical objects a particular research design allows us to identify.

A second related issue is that, in practice, firms in many countries do not seem to utilize these sub-minimum wages, leading to no first-stage wage differences around the cut-off. Although it remains an open question when and why firms do not utilize sub-minimum wages, recent evidence suggests that it could be related to fairness norms (see ([Giupponi et al., 2024](#))). The presence of such norms suggests that variation by age group might reveal the influence of such norms, rather than the impact of increasing wage floors across the board per se.

3.3 Review of the evidence on employment effects

Having laid out various methodological considerations when estimating the effects of minimum wage on employment, we now turn to reviewing the body of empirical evidence. Typically, minimum wage reviews focus on the elasticity of employment with respect to the minimum wage; even worse, sometimes they mix that elasticity with the own-wage elasticity (OWE) ([Brown et al., 1982](#)). found that the employment elasticity with respect to the minimum wage ranges between -0.1 and -0.3 for teens. The focus on teens allows for a more apples-to-apples comparison of the elasticities, but at the cost of reduced external validity ([Neumark and Wascher, 2008](#)). and ([Neumark and Peter Shirley, 2022](#)), mixing various elasticities, concluded that the average elasticity is negative. On the other hand, ([Doucouliagos and Tom, 2009](#)) and ([Belman and Paul, 2014](#)) focus on the elasticity of employment with respect to the minimum wage, suggesting that the overall impact of minimum wages on employment is minimal. While these reviews are somewhat informative, we discussed the problems with comparing estimated elasticities of employment with respect to the minimum wage across studies in [Section 3.1](#).

Recent articles, such as ([Harasztosi and Lindner, 2019](#)) and ([Dube, 2019a](#)), review the evidence on OWE by considering studies that report both first-stage wage effects as well as employment effects. A more comprehensive list of papers has been compiled by ([Dube and Zipperer, 2024](#)), who have created an

online repository of minimum wage studies with both employment and wage estimates.²⁸ These are studies published starting in 1992 that use either “quasi-experimental” or “experimental” variation, and study statutory minimum wages (as opposed to floors set by collective bargaining, private platforms, etc.). For studies that meet these criteria, (Dube and Zipperer, 2024) consider the database to be close to comprehensive for the U.S., U.K., Germany, and Canada. The database contains a single OWE estimate from each study, based on an assessment of the authors’ preferred empirical specification; when this was unclear, (Dube and Zipperer, 2024) reached out to the authors for guidance on model or sample selection. When there were estimates for a range of groups, the database includes the estimate for the broadest group of low-wage workers presented in the paper. In some instances, the database includes an average of multiple estimates; this is driven by either the presence of multiple low-wage groups or specifications, or because of the authors’ preferences, as communicated directly. As of the time of writing, the database (version 1.0) contains 88 studies, 72 of which were published in academic journals. In this chapter, we will focus on these 72 studies only.

When the authors report the OWE estimate and standard errors (as was the case for 28 of the studies), these are the estimates included in the database. When the authors do not report the OWE, the estimates in the database are obtained by dividing the employment effect by the wage effect, as described above. Following (Harasztsosi and Lindner, 2019), confidence intervals are constructed using the delta method and assuming the independence of employment and wage estimates.

When considering the magnitudes of the OWE, it is useful to think about the welfare consequences. In this respect, a key ingredient in the welfare analysis of minimum wages is the elasticity of total pre-tax earnings of all low-wage workers (including those without a job) with respect to the policy. When the OWE is -0.1 , it means that around 10 % of the potential increase in earnings is forfeited due to job losses. Conversely, when the OWE is -0.9 , it indicates a loss of 90 % of that potential earnings growth.

Taking these considerations into account, we follow (Dube and Zipperer, 2024) and categorize the OWE’s as follows. We consider an OWE that is less negative than -0.4 to represent either a “small negative” or “positive” effect on jobs—as they imply that the total wage bill collected by workers increased by at least 60 % from the policy. We characterize an OWE between -0.4 and -0.8 as having a “medium negative” impact on employment. Finally, we take an OWE more negative than -0.8 to signify a “large negative” impact on jobs—as the disemployment erases more than 80 % of the potential earnings gains. Like all such categorizations, ours inherently reflects some degree of subjectivity. Others may choose to make their own assessment using the underlying OWE data.

²⁸ The repository is available at: <https://economic.github.io/owe/>.

3.3.1 Estimates across all studies

We begin by describing the central tendencies in the OWE estimates, drawing from the discussion in (Dube and Zipperer, 2024). Spanning more than thirty years of research and a wide range of affected workers, the estimates from the OWE repository suggest that the overall impact of minimum wages on employment is small. The median OWE estimate from the analysis of 72 published studies is -0.13 , suggesting a minor impact of the minimum wage on jobs. This estimate implies that the total earnings of low-wage workers rise by 87 % of what one would expect if there were no job losses due to the policy. Put differently, employment reductions offset only about 13 % of the potential earnings gains. While the mean OWE estimate of -0.25 is somewhat larger in magnitude, it is still consistent with fairly modest job losses and relatively large earnings gains. This conclusion remains when we restrict attention to the 57 US studies: for that group, the median and mean OWE are -0.11 and -0.22 , respectively.

In addition to the central tendencies, we can also look at the size distribution of estimates using our rubric: more negative than -0.8 as “large negative”, those between -0.4 and -0.8 as “medium negative”, and those more positive than -0.4 as “small negative or positive”. Fig. 5 shows the distribution of OWE for the 70 published studies. 51 (or around 71 %) of the published studies have positive or small negative estimates; in contrast, 21

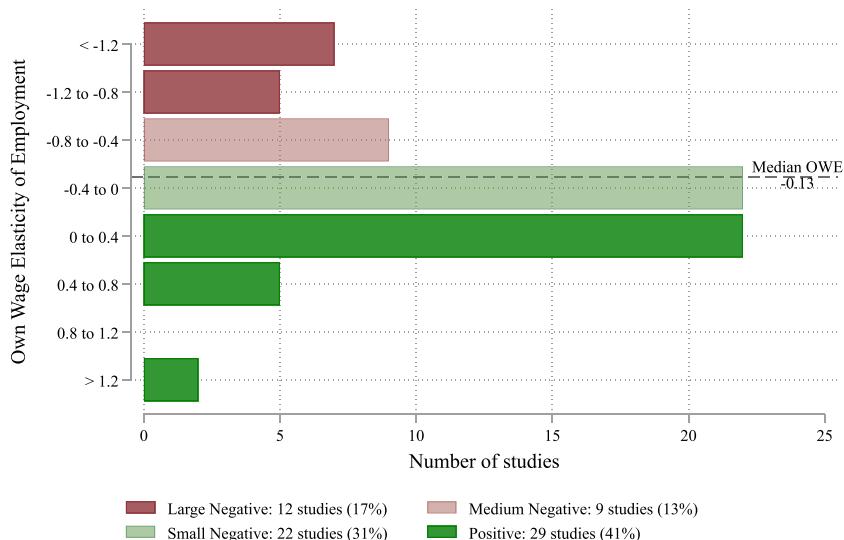


FIG. 5 Distribution of published studies, by own-wage elasticity estimate range. Notes: The figure shows the frequency distribution of own-wage elasticities from the (Dube and Zipperer, 2024) database. The bars are additionally colored differently for four broad groups: Positive ($OWE \geq 0$); Small Negative ($-0.4 \leq OWE < 0$); Medium Negative ($-0.8 \leq OWE < -0.4$); and Large Negative ($OWE < -0.8$).

studies (or around 29 %) have estimates that would imply a large or medium negative effect. The overall evidence base suggests that to date, minimum wage policies have raised wages of low-wage groups much more than they have reduced their employment—thereby effectively increasing total earnings for low-wage workers.

3.3.2 Estimates for broad groups

Most (51 out of 72) studies considered so far are from narrow subgroups. These include teens, restaurant workers, nurses, and grocery workers, to name a few common groups. However, estimates from these groups may not be very informative about the overall impact of the policy on all affected workers (the average effect of treatment on the treated, in the language of program evaluation).²⁹

Given the concerns about external validity of such narrow groups, it is useful to separately consider papers that provide estimates for broad groups of low-wage workers that better represent the overall impact of the policy (Dube and Zipperer, 2024). also classify the papers in the OWE database as “broad” when they are likely to capture the majority of workers affected by the policy. This includes, for example, (Cengiz et al., 2019), which provides an estimate of the impact of US minimum wages on the total number of low-wage jobs as described above (Derenoncourt and Montialoux, 2021), which captures all adults aged 25 to 55 years, and (Neumark et al., 2004), which captures all incumbent low-wage workers. Finally, OWE estimates which are focused on all workers with at most a high school degree, as in (Monras, 2019), are considered broad as well.

Based on this classification, just under one third of published studies (or 21 of 72) can be classified as “broad.” (Interestingly, the share of “broad” studies has risen considerably over the years, from zero in the 1990s to 48 % in the 2020s, as shown in Appendix Fig. A8). Fig. 6 plots the distribution of the OWE estimates from these studies and also compares it to the distribution of the other 47 studies of narrower groups. Broad-group studies tend to have estimates closer to zero, with a median OWE estimate of 0.02. 19 of the studies (or 90 %) have OWE estimates that are either positive or small negative. The contrast with narrower groups is also illuminating. The narrow-group studies are also consistent with small employment effects, but the median OWE estimate for those 51 studies is somewhat more negative at −0.17. One possible explanation for a slightly more negative effect found among narrower groups is that these studies are more likely to reflect labor-labor substitution or possibly reallocation across industries, although many other explanations exist.

²⁹ For example, Congressional Budget Office has assumed different OWE estimates for teens and adults in its analysis of prospective minimum wage increases (Congressional Budget Office, CBO, 2019).

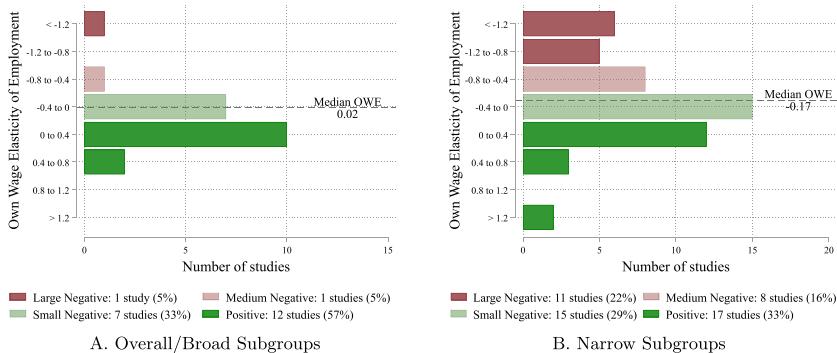


FIG. 6 Distribution of published studies, overall/broad and narrow subgroups. Notes: The figure shows the frequency distribution of own-wage elasticities from (Dube and Zipperer, 2024), separately for overall/broad and narrow subgroups. Common examples of narrow subgroups are teens or restaurant workers. Positive ($OWE \geq 0$); Small Negative ($-0.4 \leq OWE < 0$); Medium Negative ($-0.8 \leq OWE < -0.4$); and Large Negative ($OWE < -0.8$).

3.3.3 Heterogeneity of employment effects

We can further disaggregate the 51 estimates from narrow group studies. Of these, 21 were about restaurants or retail sectors, while 15 were for teens. There has also been a notable decline in teen share of studies, falling from around 50 % in the 1990s to 4 % in the 2020s, which is consistent with declining importance of teens among minimum wage workers. In contrast, the share of restaurant / retail studies has remained fairly stable, going from 33 % in the 1990s to 26 % in the 2020s (see Appendix Fig. A8).

When we focus on restaurants or retail sectors, we obtain a median OWE of -0.09 and a mean of -0.17 . So, the low-wage sector studies yield fairly similar estimates as the overall evidence base, though a bit less negative. In contrast, the 15 teen studies had a median OWE of -0.17 and a mean of -0.25 . In other words, the estimates for teens appear to be a bit more negative, potentially indicating more labor-labor substitution, even as the mean OWE of -0.25 still suggests any teen disemployment effect has been small.

Of course, there are other confounders when comparing these studies, including the methods and quality of the study. Ideally, when assessing heterogeneity, we would account for such differences. As a simple gauge, it turns out that the OWE estimates have tended to become less negative over time, possibly due to falling publication bias or due to improved data and methods. For example, studies published before 2010 had a median estimate of -0.40 , while studies published since then had a median of -0.04 . As it turns out, slightly fewer teen studies have been published since 2010 than before (8/15); in contrast, the vast majority (41/57) of the non-teen studies were published after 2010. When we look within these two time periods, there is no systematic indication that teen studies are more negative; on average, the gap in the median OWE between teen and non-teen studies was positive prior to 2010, and negative afterwards, with an overall

(within-period) gap in the median OWE close to zero (based on a median regression of the OWE on teen and period indicators).

Age. Since so many aspects may differ across studies of teens and non-teens, a better assessment of heterogeneity can come from studies that have explicitly provided estimates for teens or young workers along with an estimate for overall low-wage workers (or a broader demographic group). Here we consider papers that provide wage and employment estimates for both groups. Using event-based approaches, both (Cengiz et al., 2019) and (Cengiz et al., 2022) find similarly small/positive and statistically insignificant OWE's for teens (0.36 and 0.40, respectively), as well as overall low-wage workforce (0.41 and 0.11, respectively). These mirror our event-based findings in Table 1: teen and high recall OWEs are 0.12 and 0.34, respectively.³⁰ (Hampton and Evan Totty, 2023) provides employment elasticities using an event-based approach for incumbent low-wage workers by age bins: 16–21, 25–34, 35–44, 45–54, 55–61, and 62–70. The average OWE is 0.07, and for all but the last group, OWEs fall between −0.02 and 0.40, and employment estimates are indistinguishable from zero. For the oldest (retiring-age) group, they find a positive and statistically significant employment effect with OWE of 0.82 (which is consistent with delayed retirement, a result also found in (Borgschulte and Cho, 2020; Manning, 2021) provides estimates for teens as well as individuals aged 20–24 year using various TWFE-logMW specifications. Without controls for state-specific trends or regional controls, the OWEs for teens and older workers are both around −1.2; and both are highly sensitive to inclusion of controls for heterogeneity, for reasons we have already discussed. For the UK, (Giupponi et al., 2024) report an overall OWE of −0.20 for 25–64 year olds; adding younger workers and considering 16–64 year olds produces an estimate of −0.10, suggesting broadly neutral/positive effects for younger workers. Our general take-away is that there is little difference in teen and non-teen OWEs *within studies*; so the between-study differences in estimates uncovered above likely reflects methodological and/or sample differences.

Most studies examine the heterogeneous impact of the policy across broad age groups. One notable exception is (Giuliano, 2013), who finds that in response to a federal minimum wage increase, a large retailer substitutes *towards* teens overall, but substitutes *away from* teens from low-income households toward teens from higher-income households. This is consistent with the presence of substitution when we look at narrow groups, which we will return to in Section 4.1.3, but it could also reflect an increased labor supply of teens from higher-income families.

Race. Besides age, another notable within-study heterogeneity to consider is race. Using the 1960s expansion in the US minimum wage, two recent

³⁰ Our estimates in this chapter are slightly different from (Cengiz et al., 2022) due to our combining multiple phases of events, and using slightly different event definitions.

studies arrived at slightly different conclusions ([Derenoncourt and Montialoux, 2021](#)). find similar null effects on employment overall and for Black workers in particular, with OWEs of 0.06 and 0.15, respectively ([Bailey et al., 2021](#)). also finds a small OWE of -0.14 for workers overall, but a somewhat larger magnitude -0.29 for Black men. At the same time, all these estimates would be considered “small” using our heuristic. In more recent studies that have specifically report wage and employment effects by race, ([Cengiz et al., 2019](#)) finds an an overall OWE of 0.41, as compared to -0.09 for Black/Hispanic workers; however, the latter is fairly imprecise. In ([Cengiz et al., 2022](#)) for the high recall group overall the OWE is 0.11, while for Black/Hispanic workers it is found to be -0.52 ; again, however, the latter estimate is imprecise, making the contrast difficult to ascertain. Finally, ([Wursten and Reich, 2023](#)) implement a stacked event study analysis with prominent minimum wage changes (more than 5 %) between 1979–2019, considering low-wage workers with a high school or less education (as well as workers in the restaurant sector). They find OWEs between -0.01 and 0.04 for white, Black and Hispanic workers.

Overall, there is no clear evidence on systematic differences by race. One obvious challenge is sample size: the samples for Black or Hispanic workers are much smaller, which makes detecting (true) heterogeneity by racial sub-groups difficult.

Composition of jobs. The minimum wage can also affect the composition of jobs in the economy. Here we review the evidence on industry composition, while we discuss the impact on occupation/job tasks in [Section 4.1.3](#), since those evidence are often interpreted in the context of capital-labor substitution and automation. We find a strong indication of heterogeneous impact of the policy by sector ([Cengiz et al., 2019](#)). and ([Gopalan et al., 2021](#)) find negative effect in the tradable/manufacturing sector, while there is no discernible employment effect in the local service/non-tradable sector. It is worth keeping in mind that in the U.S. (and other high-income countries), most low-wage workers are in local service jobs, and so the overall policy impact will be dominated by the impact in that sector. In line with that, the across studies median OWE from the retail and restaurant sectors (-0.09) is very similar to the overall median OWE of -0.13. Similarly, ([Harasztsosi and Lindner, 2019](#)) find a significant reduction in employment in the tradable sector, but not in the local service sector in Hungary. A larger share of low-wage workers in Hungary are employed in the tradable sector, so those effects have more important overall implications there. Still, the significant reduction in employment in one sector does necessarily lead to a reduction of overall employment in the presence of reallocation of workers to other sectors ([Harasztsosi and Lindner, 2019](#)). find a more muted disemployment effect in their worker level (as opposed to firm-level) analysis—consistent with the presence of such reallocation.

Short vs. medium run estimates. One concern often raised about the relatively small employment estimates is that they reflect short-run effects that do not capture the dynamics (e.g., firm exit) that take longer to take effect.

Given the nature of minimum wage variation in the U.S., very long-run estimates are difficult to obtain. At least until recently, most state-level variations did not last for more than a decade. Some researchers have used filtering of the minimum wage measure to concentrate on “low frequency” variation by decomposing minimum wage variation to a permanent and a transitory components (e.g., (Baker et al., 1999)). In practice, the (Baker et al., 1999) approach is similar to adding lagged minimum wages as explanatory variables (distributed lags models), but it is not a particularly transparent way of understanding long run impacts.

However, what is true is that quite a few studies estimate what we would call “medium run” estimates (e.g., 4–7 years out effects). Historically, these were estimated based on distributed lags TWFE-type models (e.g., (Allegretto et al., 2011; Dube, Lester et al., 2010; Neumark et al., 2014; Jha et al., 2024)). Some of these TWFE-type estimates show larger longer-run estimates, but are highly sensitive to specification, as demonstrated already.

Numerous recent work tends to report event-based estimates up to at least 4 years (sometimes up to 7 years) after a policy change (Azar et al., 2023; Bailey et al., 2021; Cengiz et al., 2019; Harasztsosi and Lindner, 2019; Clemens and Michael, 2021; Derenoncourt and Montialoux, 2021; Godey and Michael Reich, 2021; Godøy et al., 2024; Monras, 2019; Ruffini, 2022; Wursten and Reich, 2023; Monras, 2019). finds that the employment effect grows steadily more negative between event years 0 and 3.³¹ (Clemens and Michael, 2021) also find an indication of a more negative medium run effects, but only for the 6 “large increase” states (more on this below). However, the rest of the papers do not suggest that medium run employment effects are substantially more negative than shorter run effects. This comes out of the 1967 expansions (studied by (Derenoncourt et al., 2021) and (Bailey et al., 2021)); state-level changes after 1979 (studied by (Cengiz et al., 2019; Cengiz et al., 2022; Godey and Michael Reich, 2021; Rao and Risch, 2024; Vergara, 2023)) or studying the impact of doubling the minimum wage from Hungary (Harasztsosi and Lindner, 2019).

An added piece of evidence comes from the 60 events we consider in this chapter. For at least 35 of these events, we can calculate a 4 year out effect, and for 17 we can calculate a 6 year out effect, as shown in Appendix Fig. A4 for restaurants, and Appendix Fig. A5 for the high-recall group. In both cases the 6-year out impact on wages is larger than at other time horizons. However, the 6-year out employment estimates are small, positive in sign, and do not indicate a more negative effect on employment than at shorter horizons.

Finally, using region-by-demographic variation, the 20-year effect of the 1998 introduction of the UK National Minimum Wage was also found to be

³¹This finding is critically dependent on the use of a linear pre-treatment trend adjustment, which is different from the other event-study approaches discussed here.

close to zero (see the working paper version of (Manning, 2021), and (Dube, 2019a) which reports an OWE of -0.04 (s.e.=0.21)).

High versus low minimum wages. Naturally, there is a lot of interest in understanding possible non-linearities in the employment effect at higher levels of minimum wages. This is especially so given the increasing experimentation by governments on more ambitious wage standards.

Cengiz et al., (2019) use event-by-event estimate to assess heterogeneity by the bite of minimum wage (as measured by the Kaitz index) at the state level, and finds no heterogeneity for events up to 2016 (Godey and Michael Reich, 2021). estimate impact for broad group of low-wage workers at the sub-state level and does not find larger job losses in more binding areas (Clemens and Michael, 2021). consider heterogeneity by size of increases in the post-2013 period using an event study design. Considering all 27 events, their preferred specification find a small overall OWE of -0.26 for their “low-skill” group (individuals ages 25 and under without a high school degree); for the six states (and D.C.) raising their minimums by \$2.50/hour or more, they find an OWE of -1.01 , while for the states raising less the OWE estimate is 0.46 .³² (For a different group—those under 21—the two OWEs are much closer at -0.41 and -0.03 .) However, it is unclear how sensitive their findings are to the particular cutoff, or use of specific set of time-varying covariates (such as the state GDP).

Additionally, we consider estimates from our 60 events by the size of the minimum wage change. In Appendix Table A3, column 4, we report the estimates from the top tercile (third) of the events in terms of the size of minimum wage increase. We find a larger wage effect for the restaurant sample, but the OWE of 0.23 (restaurant) and 0.41 (high recall) are not meaningfully different from their counterparts based on all events in column 1 (-0.03 and 0.34 , respectively). But as we mentioned, it is quite useful to plot the full set of event-by-event estimates to transparently show the patterns. This is what we do in Fig. 7 for restaurant sample, and Appendix Fig. A7 for the high-recall sample; where the estimates are all sorted by the size of the minimum wage increase associated with each event. There is an overall positive association between the wage estimates and the log minimum wage change, with correlation coefficients of 0.43 for restaurants, and 0.27 for high recall. However, there is no evidence of a negative association between employment estimates and the minimum wage change; the correlation coefficients are 0.17 and 0.12 for restaurants and high recall, respectively.

Outside of the U.S. context, (Harasztsi and Lindner, 2019) study the impact of doubling the minimum wage in Hungary—raising from the current U.S. median to minimum wage ratio to the level of France—in two years. They estimate an OWE

³² Hourly wages increased 18 % in the six states, raising their minimums by \$2.50/hour following the minimum wage, while 11 % in other treated states. Therefore, these OWE estimates imply a very high degree of non-linearity in employment responses.

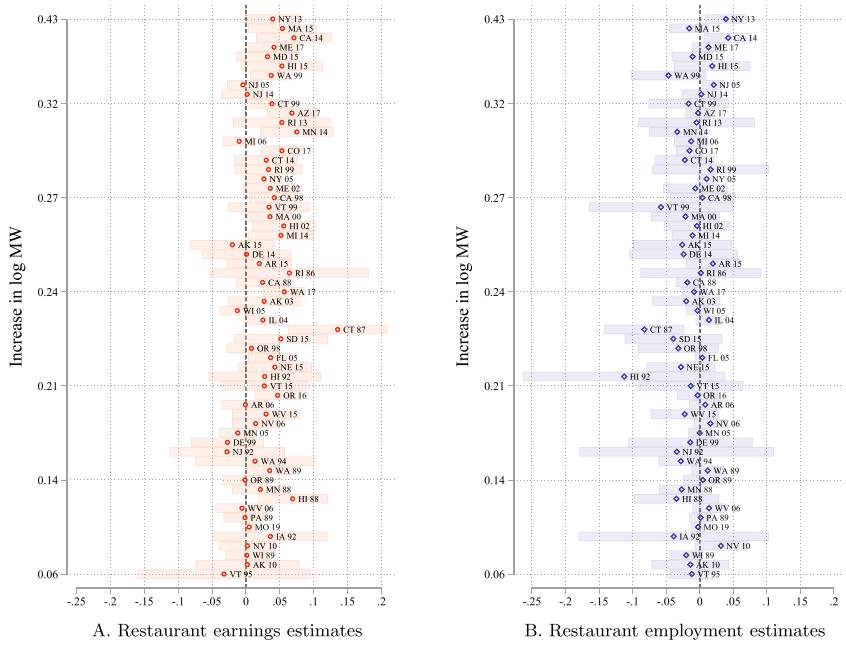


FIG. 7 Event-by-event employment and earnings estimates for the restaurant sector. Notes: These figures plot the event-by-event earnings and employment estimates for the restaurant industry. Thus, each point estimate is from an event-study regression (Eq. 5) with a particular combined event and its corresponding clean controls. The dependent variable is log average weekly earnings (2023 dollars) for Panel A, and log employment rate for Panel B. All regressions use state population weights. The 95 % confidence intervals use Ferman-Pinto standard errors. We identify 60 events, where successive minimum wage increases are combined into a single event (see Appendix 12). The events are sorted by total increase in log minimum wage.

of -0.22 in the medium term, as we discussed above, mainly coming from the manufacturing sector. The sharp increase in the UK National Living Wage wage between 2016 and 2019 also offers another example, and the OWE from the UK in this period were all small (e.g., Giupponi et al., 2024).

Going forward, we think further research is needed to carefully assess possible “non-linearities” from minimum wage increases that are both large and are at high levels. The challenge is to consider the different dimensions of what a “bigger” minimum wage means (size of change, pace of change, ending level), while at the same time recognizing how results on thresholds may be fragile, especially when considering different dimensions of the policy. For this reason, we think it is valuable for researchers to transparently show event-specific estimates along various relevant dimensions, such as the minimum wage level, size of the increase, bite of the policy, etc. This can mitigate some of the challenges arising from false discovery when searching for heterogeneity by thresholds.

Market concentration. Models of imperfect competition in the labor market imply more positive employment effects when the labor market is more monopsonistic (Azar et al., 2023). assess this by considering heterogeneous effects of minimum wage hikes during the 2010–2016 period, focusing on the retail sector. Using a similar bunching approach as (Cengiz et al., 2019), they show that in non-concentrated markets ($HHI < 0.25$), there is a negative (but statistically insignificant) overall OWE of around -0.50 . In contrast, for concentrated markets ($HHI > 0.25$) there is a sizable positive and statistically significant OWE of 1.8. Interestingly, this heterogeneity appears to be the clearest in the retail labor market; for restaurants, the differences are much more muted, though that may in part reflect less accurate concentration measurement. In a similar vein, (Wiltshire, 2022) finds a more positive employment effect from the 1996–1997 federal minimum wage increase in local labor markets that had become more concentrated from prior Walmart Supercenter entry. Finally, and broadly along the same line, (Okudaira et al., 2019) find less negative employment effects from minimum wage increases in Japan on plants with greater estimated monopsonistic markdown. There is similar evidence from developing country contexts as well: employment effects (OWEs) are found to be more positive in more concentrated local labor markets with greater concentration in India (Soundararajan, 2019).

Interactions with earnings subsidies. Minimum wages are often presented as a substitute for earnings subsidies such as the Earned Income Tax Credit (EITC). However, there may also be important interactions between these policies. From a theoretical perspective, while the EITC is thought to encourage labor supply—a desired outcome—this in turn also push down pre-tax wages, an unintended consequence (Rothstein, 2010). In principle, the minimum wage can mitigate such a downward push in pay (e.g., (Lee and Saez, 2012)). The empirical evidence on the EITC and minimum wage interaction is quite limited (Neumark and Wascher, 2011). use a TWFE-logMW model and interactions with state minimum wage and EITC policy over the 1997–2006 period. They find that the earnings effects from EITC for single mothers are more positive in the presence of a higher minimum wage, though the earnings and employment effects are more negative for some childless adults. Given the problems with the TWFE estimates (compounded when dealing with two separate policies), we think it would be worth re-visiting this topic using a clean event-based design.

Summary. So what can we learn from the heterogeneity of employment effects found in the literature so far? First, the literature suggests limited heterogeneity by age group, especially when we look within studies. The evidence on differences by race is somewhat under-powered due to small sample problems, but the differences found tend to be modest. There is more evidence on the changing composition of jobs—including by sector or occupation—but these are unlikely to be contributing to an overall decline in employment. At the same time, more work needs to be done to understand

heterogeneity by policy environment and market structure, especially in light of existing work suggesting relatively more positive employment effects in concentrated markets. Finally, most event study estimates do not show significantly larger medium-term effects for typical minimum wage increases observed to date. However, the variation across the size or level of minimum wages remains poorly understood. Further rigorous research on non-linearities in employment effects, and the identification of potential turning points (beyond which disemployment is more pronounced) would be highly valuable.

3.4 Effect on total hours

Much of the literature has focused on headcount employment. However, some papers directly study the adjustment in hours conditional on employment ([Zavodny, 2000](#)) considers teen employment for the 1979–1993 period and finds no meaningful effect on hours ([Couch and David, 2001](#)) find large negative hours elasticities, but these are based on specifications without any year fixed effects (i.e., pooled time-series regressions) ([Neumark et al., 2004](#)). finds some reductions in hours using a short panel of workers using household CPS data from 1979–1997. Using a TWFE model, ([Sabia, 2009](#)) finds no effect on hours conditional on employment for the retail sector during the 1979–2004 period ([Dube et al., 2007](#)). study the 2004 San Francisco minimum wage increase using a restaurant survey and found no impact on hours; both the FTE and the headcount estimates were close to zero ([Allegretto et al., 2011](#)). find that estimates for teen hours are sensitive to regional and trend controls during the 1990–2009 period, mirroring employment results.

More recently, ([Cengiz et al., 2019](#)) use an event-based approach and changes in the frequency distribution, considering the effect on headcount, as well as full-time equivalent (FTE) employment for all affected workers. The OWE for FTE employment is 0.60 while for headcount it is 0.41, both being statistically indistinguishable from zero and indicating little overall impact on hours ([Cengiz et al., 2022](#)). use a similar event-based design for demographic probability groups and find a small (but statistically significant) increase in full-time share for the narrower (lower wage) high-probability group, but a null effect on the broader high-recall group, with no change in overtime shares. Using private payroll data from a sample of businesses, ([Gopalan et al., 2021](#)) do not find any significant reduction in weekly hours conditional on employment either among incumbent workers, or affected establishments. Studying the increase in the 2014 Seattle minimum wage using administrative data on hours and wages, ([Jardim et al., 2022](#)) find a large reduction in total hours worked in jobs paying below \$19, although, as discussed previously, these results likely reflect more hours being paid above \$19. Following incumbent workers, they find a statistically significant reduction in hours. However, the implied magnitudes are small: the average overall FTE-based OWE across specifications is around –0.2 Using SIPP data to track incumbent workers and

events over the 1984–2014 period, ([Hampton and Evan Totty, 2023](#)) find a small own-wage elasticity of hours of around -0.2 for all affected incumbent workers. Based on event-based estimates from 2005–2017 and ACS data, ([Godoey and Michael Reich, 2021](#)) find no reduction in weekly hours, either overall or in local areas where the minimum wage was highly binding.

Overall, most U.S. evidence—especially more recent, higher quality estimates—do not suggest large adjustments to hours of work conditional on employment. This is particularly true for low-wage workers generally; for particular subsets of workers, there are some estimates in both negative and positive directions.

Outside the U.S., a handful of papers have examined the impact on hours adjustments. The evidence is mixed at best. In the UK, ([Stewart and Joanna, 2008](#)) finds a reduction in hours, while ([Connolly and Gregory, 2002](#)) find no effect. In Ireland, ([McGuinness and Redmond, 2018](#)) find a reduction in hours, especially for workers on temporary contracts. In Germany, ([Bossler and Hans-Dieter, 2020](#)) and ([Burauel et al., 2020](#)) find a reduction in contractual hours. However, ([Burauel et al., 2020](#)) shows that this does not translate into a significant reduction in actual hours worked. This suggests that some of the responses in hours may reflect reporting rather than a substantial change in behavior.

4 Margins of adjustment

The previous section reviewed the extensive empirical literature on how minimum wages affect employment and wages. In this section, we broaden the scope and consider the impact on low-wage labor markets more generally, with a focus on unpacking the various margins of adjustments. The outcomes considered in this section are those that can shed light on how firms actually respond to the cost shock resulting from higher minimum wages.

4.1 Review the evidence on various margins of adjustment

There is a long list of potential channels through which minimum wages affect the low-wage labor market. Here we list the most important ones:

1. Increased wages
2. Non-compliance of the policy (i.e., firms pay below the minimum wage)
3. Reduction in non-wage compensation or amenities (e.g., fringe benefits or working conditions)
4. Change in employment (firm exit/entry; substitution to other inputs)
5. Change in population through in/out-migration of workers, or change in labor force participation
6. Wage retrenchment of higher-skilled workers
7. Pass-through to consumers by raising output prices
8. Pass-through to suppliers by lowering input prices (e.g., rent, prices of intermediate goods and services)

9. Pass-through to firm-owners by lowering profits
10. Change in worker turnover and cost reduction coming with lower training costs
11. Improved productivity (increased effort, reorganization, greater allocative efficiency)

Among these mechanisms, the first five are closely related to the previous section. The evidence on non-compliance (Channel 2), change in non-wage amenities (Channel 3), capital-labor substitution and firm entry and exit (Channel 4), migration and participation responses (Channel 5) refine the measurement of the employment responses and the change in worker's compensation. For example, the presence of a substantial cut in non-wage amenities could imply that firm-level labor costs are less affected than suggested by the first-stage wage estimates. Furthermore, estimates on firm dynamics can shed light on sources of employment changes and inform modeling choices.

In contrast, channels 6 through 11 shed light on how the change in labor cost is absorbed—if not through curtailing low-wage employment. A rise in the costs associated with low-wage workers could be covered by: a reduction in the expenses for high-wage workers (Channel 6.); an increase in total revenue through higher output prices—which in turn is aided by insensitive consumer demand (Channel 7.); lower rents or prices of other inputs (Channel 8.); lower profits (Channel 9.); decreased turnover and the implied savings on training and recruitment costs (Channel 10.); and improved productivity (Channel 11.). Therefore, these channels provide us with critical evidence on the incidence of the policy.

Next, we review up-to-date evidence on all of these channels. For each adjustment margin, we list the relevant studies along with our summaries in [Tables 2–5](#). For some of these margins (like price, turnover, employment) we have relatively well-developed literature with many available studies. In these cases, we restrict our attention to published papers, as they have been peer-reviewed and are finalized. However, for some margins, we only have a few (or no) published papers to consider. Furthermore, sometimes this limited evidence comes from different national contexts (e.g., U.S., U.K., Germany, Hungary, China), making it difficult to compare the results. In such cases, we expand our discussion to include relevant working papers in order to paint as complete a picture as possible. We also apply our subjective assessments on the quality and credibility of papers to draw some (at times tentative) conclusions whenever possible. Our hope is that future research will fill in the gaps in these areas, allowing us to reach more definitive conclusions.

4.1.1 Non-compliance

As we discussed in [Section 2.2](#), to date, the core goal of the minimum wage literature has been to understand the link between increased labor costs and the change in employment. Indeed, the lack of a sizable overall employment effect—as documented in [Section 3.3](#)—is notable only if minimum wage

TABLE 2 Margins of adjustment: non-compliance and fringe benefits.

Adjustment	Evidence	Summary
Non-compliance	Significant increase in wages (both in administrative and survey data) Modest (Ashenfelter and Robert, 1979 ; Bernhardt et al., 2013 ; Caliendo et al., 2019) to low non-compliance (Goraus-Tańska and Lewandowski, 2019 ; Clemens and Michael, 2022)	Not clear; likely to be small
Cutting fringe benefits	Amenities: No change in fringe benefits and working conditions (Simon and Kaestner, 2004); No change in non-cash benefits (Harasztsosi and Lindner, 2019) Health insurance: Reduction in health insurance coverage (Meiselbach and Jean, 2023) offsetting 9–16 % of wage increase for low wage workers (Clemens et al., 2018); reduction or no effect depending on the presence of non-discriminatory laws (Marks, 2011) On-the-job training: Mixed evidence: reduction in training (Neumark and Wascher, 2001a ; Hara, 2017); No change in training (Grossberg and Sicilian, 1999); increase in training (Acemoglu and Pischke, 2003 ; Arulampalam et al., 2004); No change in training, some decrease in intensity (Bellmann et al., 2017)	At most a small offset of the wage increase

Notes: This table reviews the literature on the margins of adjustment associated with non-compliance and fringe benefits.

policies significantly raise labor costs. If firms simply do not comply with the policy, or if they are able to offset policy-induced wage changes by sharply reducing non-wage benefits, there would be no reason to necessarily expect a substantial change in employment.

What are some ways in which non-compliance could occur? These would, in general, depend on the institutional setting, and here we list some possibilities. First, and most simply, firms may pay their regular employees below the mandated floor. Second, higher minimum wages could create incentives for employers to shift some of their employment to the informal sector characterized by greater non-compliance. This is especially relevant in the context of low- and middle-income countries with a higher degree of informality; therefore, we discuss this form of non-compliance in [Section 6](#). Third, firms

TABLE 3 Margins of adjustment: employment and refinements.

Adjustment	Evidence	Summary
Cutting employment	See Section 3.3 on employment effects	Limited Employment change
Substitution (capital-labor)	<p>Firm-level on capital: evidence for some capital-labor substitution ((Harasztsosi and Lindner, 2019) in Hungary; (Hau et al., 2020); (Geng et al., 2022) in China)</p> <p>Firm-level on robots: evidence for increased robot adoption in some periods, but not in others in China ((Fan et al., 2021))</p> <p>Automation (inferred from occupation change): A raise in automation, disagreement on employment consequences ((Lordan and Neumark, 2018); (Aaronson and Brian, 2019)); Decrease in automation ((Downey, 2021))</p>	Some indication for substitution away from certain types of workers
Substitution (labor-labor)	<p>Lack of heterogeneous impact by demographic groups (see Section 3.3.3)</p> <p>Substitution to more skilled or productive workers ((Clemens et al., 2021); (Horton, Forthcoming))</p>	No substitution b/w broad demographic groups, some substitution from lower to higher productive ones
Exit & entry	<p>Exit rate: Increase in exit rate ((Draca et al., 2011); (Mayneris et al., 2018); (Aaronson et al., 2018); (Harasztsosi and Lindner, 2019); (Luca and Luca, 2019); (Dustmann et al., 2022); (Chava et al., 2023)); No increase in exit ((Rohlin, 2011))</p> <p>Entry rate: Increase in entry rate ((Aaronson et al., 2018)); No effect on entry rate ((Harasztsosi and Lindner, 2019)); Decrease on entry rate ((Rohlin, 2011); (Draca et al., 2011); (Luca and Luca, 2019))</p> <p>Number of firms: A reduction in number of firms ((Orszag and Peter Mattila, 2002); (Dustmann et al., 2022))</p>	Increase in exit rate, unclear effect on firm entry. Number of firms fall

Continued

Table 3 Margins of adjustment: employment and refinements.—Cont'd

Adjustment	Evidence	Summary
Migration	<p>Low-skilled population size. Evidence for outmigration (Castillo-Freeman and Richard, 1992) though results are fragile (Krueger, 1994). Decrease in share of low-skilled population (Monras, 2019) in U.S.; Increase in low-skilled population (Minton and Wheaton, 2023) in U.S.; (Giupponi et al., 2024) in UK, (Ahlfeldt et al., 2018) in Germany).</p> <p>Immigrant population. Mixed evidence: Reduction in immigration (Orrenius and Zavodny, 2008; Cadena, 2014). Increase in immigration (Boffy-Ramirez, 2013; Giulietti, 2014)</p>	In most cases no significant change in the local population, unclear evidence on immigrants
Participation	<p>Extensive margin: Increase in overall participation (Laws, Athene, 2018; Agan and Makowsky, 2021), for prime-age unskilled (Luna-Alpizar, 2019); close to retirement (Borgschulte and Cho, 2020; Hampton and Evan Totty, 2023) for parents with young children (Godøy et al., 2024); Reduction in participation overall (Wessels, 2005), for young (Lavecchia, 2020); No change in participation (Cengiz et al., 2022)</p> <p>Unemployed's job search: Modest negative effect (Laws, Athene, 2018); No effect (Adams et al., 2022); Positive effect (Piqueras, 2023)</p>	Mixed evidence. Probably a limited effect on the participation, and some increase in job search

Notes: This table reviews the literature on the margins of adjustment associated with employment and various refinements of the employment estimates including capital-labor and labor-labor substitution, firm dynamics, migration and participation.

TABLE 4 The effect of minimum wages on prices.

Study	Industry	Effect on Prices	Pass-through
(Lemos, 2006)	All consumers	Increase	Not reported (likely to be close to full-pass through)
(Aaronson, 2001)	Restaurant	Increase	Pass-through depends on specification
(Dube et al., 2007)	Restaurant	Increase for Fast-food restaurant	More than full pass-through
(Aaronson et al., 2008)	Restaurant	Increase for Fast-food restaurant	More than full pass-through
(Allegretto and Reich, 2018)	Restaurant	Increase	Nearly full pass-through
(Katz and Alan, 1992)	Fast-food restaurant	No effect (imprecisely estimated)	-
(Card and Alan, 1995a)	Fast-food restaurant	Mixed (imprecisely estimated)	-
(Basker and Khan, 2016)	Fast-food restaurant	Increase	Sizeable pass-through
(Ashenfelter and Štěpán Jurajda, 2022)	Fast-food restaurant	Increase	Full pass-through
(Leung, 2021)	Retail	Increase in grocery stores	More than full pass-through
(Renkin et al., 2022)	Retail	Increase of grocery prices	Full pass-through
(Agarwal et al., 2024)	Hotel	No effect on price	-
(Harasztsosi and Lindner, 2019)	Manufacturing	Increase	Full pass-through

Notes: The table provides an overview of the impact of minimum wages on prices. It also shows the implied pass-through rate: the size of the change in the unit price of the output relative to the change in unit labor costs.

TABLE 5 Margins of adjustment: incidence.

Adjustment	Evidence	Summary
Prices	Significant price pass-through. See Table 4 for details	Almost full pass-through
Revenue or Quantities	(Leung, 2021) and (Alonso, 2022) find an <i>increase</i> in quantities following the minimum wage changes, while (Renkin et al., 2022) find no effect (Agarwal et al., 2024). find reduction in hotel occupancy (Harasztsosi and Lindner, 2019). find negative revenue effect in the tradeable sector, and positive effect in the non-tradeable sector.	No reduction in local service, significant reduction in tradable
Rents or Suppliers	No evidence on rents (Harasztsosi and Lindner, 2019). finds short term increase in spending on intermediate goods and no effect in the medium term	Unclear
High-skilled workers' wage	Within-firm: Small effect (Hirsch et al., 2015); Modest positive wage spillovers (Gopalan et al., 2021; Dube et al., 2019); Positive spillovers on executive compensations in China (Yao et al., 2023). See Section 5 for details.	Modest positive wage increase for slightly higher skilled workers, no effect on others
Profits	Stock-market value: No effect (Card and Alan, 1994a); Large reduction (Bell and Machin, 2018) Profitability (Profit/Revenue): Reduction in the U.K (Draca et al., 2011); No change in China (Mayneris et al., 2018) Profit: Increased profit among surviving pass-through businesses (Rao and Risch, 2024), Small effect on profit for restaurants in Georgia and Alabama (Hirsch et al., 2015); Small positive or negative effect on profits depending on the specification for a large U.S. retailer (Coviello et al., 2022); Modest fall in Hungary (Harasztsosi and Lindner, 2019); Significant fall in Israel (Drucker et al., 2021); Significant fall in the U.S (Vergara, 2023).	Mixed but likely a modest fall

Turnover	No change in turnover (Hirsch et al., 2015); Mixed evidence (Gopalan et al., 2021); Significant reduction in turnover (or increase in tenure) (Portugal et al., 2006 ; Dube et al., 2007 ; Brochu and David, 2013 ; Dube et al., 2016 ; Gittings and Schmutte, 2016 ; Liu et al., 2016 ; Jardim et al., 2022 ; Coviello et al., 2022 ; Brochu et al., 2023 ; Wursten and Reich, 2023 ; Wiltshire et al., 2023)	Lower worker turnover
Productivity	Quantity increase: (Ku, 2022) in agriculture; (Coviello et al., 2022) in retail; (Hau et al., 2020) for exporting firms Quality increase: (Ruffini, 2022) in nursing home Indirect evidence on cost saving from turnover reduction: see above Change in efficiency through market allocation: (Dustmann et al., 2022)	Increase in productivity

Notes: This table reviews the literature on the margins of adjustment associated with the incidence of the policy. The increase in labor costs can be borne by consumers (higher output prices, no decrease in quantity), suppliers, higher wage workers, lower profits, and higher productivity (reduced turnover, increased operational efficiency).

could also shift their hiring toward undocumented immigrants, for whom minimum wage policies may be less likely to be enforced. Finally, another important margin is shifting work to the gig economy (independent contractors) to get around minimum wage and other employment regulations. Importantly, a growing number of jurisdictions (including California, Massachusetts, and New York) have used litigation and/or legislation to enact wage floors for some gig workers, such as ride-share or delivery drivers ([Jacobs et al., 2024](#)). Given the growing importance of such jobs in the economy, this channel could be more important in the future ([Boeri et al., 2020](#)).

The extent of non-compliance is primarily an empirical question. However, it is worth noting that compliance might not be driven solely by legal enforcement. In fact, a recent paper by ([Stansbury, 2024](#)) documents that employers' legal incentives to comply with the minimum wage laws in the U.S. and U.K. are rather limited. At the same time, compliance could still end up being widespread if the minimum wage also shapes fairness perceptions, raises reservation wages, or shifts wage-setting norms.

Studying non-compliance is inherently difficult since it is not directly observed in most data. A common approach pioneered by (Ashenfelter and Robert, 1979) proxy non-compliance with the share of workers who report earning wages below the statutory floor in surveys such as the CPS. This approach usually indicates a nontrivial level of non-compliance (Ashenfelter and Robert, 1979). find 20–40 % of workers earn below the statutory minimum wage; more recently, (Bernhardt et al., 2013) find that around 20 % of low-wage workers experience some form of wage-and-hour violation in the U.S. Notably, the incidence of minimum wage violations is lower than the violations related to other types of workplace regulations such as overtime, meal breaks, or off-the-clock working hours (Caliendo et al., 2019). find a similar extent of non-compliance (around 7 %–25 %) in Germany. Studies have also found that the share of workers below the minimum wage seems to increase with the level of minimum wages (Goraus-Tańska and Lewandowski, 2019; Clemens and Michael, 2022; Clemens and Michael, 2022). estimate that underpayment reflects around 16–20 % of the realized wage gains.

The key problem with these estimates is that separating actual underpayment from measurement error in reported wages is inherently difficult. For example, even if there is no true under-payment, but a constant share (γ) of workers mis-report their wage as being below or above the true level with equal probability, at least $\frac{\gamma}{2} \times P(MW)$ of workers would report earning below the minimum wage, where $P(MW)$ is the share of workers truly earning exactly the minimum. Moreover, one would find that this measured share rises with the level of the minimum wage, which naturally raises $P(MW)$.³³ This illustrates the challenges in reliably quantifying how non-compliance mediates the impact of minimum wage policy on earnings. Furthermore, some of the actual payments below the minimum wages could reflect exemptions (e.g., tipped workers in most states in the U.S.) and not real non-compliance with the policy. Most studies deal with this issue by focusing on non-exempt workers, but that itself is measured noisily in surveys. Therefore, this approach is likely to overestimate the extent of non-compliance.

Regardless of the precise *level* of non-compliance, it is unlikely that it very substantially erodes the *change* in earnings resulting from a minimum wage increase. One way to think about this is through an analogy with speed limits on highways. There is substantial non-compliance with speed limits, as many drivers speed. At the same time, when the speed limit drops from 65 mph to 55 mph, even most non-compliers adjust their speed downward to avoid being far out of compliance. As a result, the average driving speed can fall substantially even as there is a non-trivial amount of noncompliance at both 55 and 65 mph regimes. Applying this logic to the low-wage labor market would suggest that minimum wages can substantially raise wages even without strict compliance.

³³ Using this same setup, (Autor et al., 2016) show that measurement error can also inflate the measured spillovers in wages (in a manner analogous to overstating non-compliance).

This logic is consistent with the empirical fact that we observe clear wage increases among low-wage workers following minimum wage hikes in both household surveys and administrative data. This wage increase is also present for undocumented immigrants (see ([Orrenius and Zavodny, 2008](#))) and in the informal sector (see [Section 6](#)). These findings indicate that firms raise workers' wages following minimum wage increases, regardless of the exact extent of compliance with the law. Finally, it is worth noting that while greater compliance might lead to a larger wage increase (something that workers would care about), it would be of secondary importance for estimating the own-wage employment elasticity—as long as there is a clear increase in the average wage earned by the relevant low-wage group (i.e., a strong first stage).

4.1.2 Amenities

Firms can respond to the minimum wage by reducing non-wage amenities, such as fringe benefits. In the extreme case, if companies could fully offset the increase in wages by spending less on amenities, there would be no net change in labor costs ([Clemens, 2021](#)). Alternatively, in the presence of some market imperfections (for example, monopsony), the provision of amenity could even increase when the minimum wage increases, especially when amenities and wages complement each other in workers' preferences ([Dube et al., 2022](#)).

Unfortunately, direct evidence on amenities is rather limited, especially since we rarely observe all relevant workplace characteristics ([Simon and Kaestner, 2004](#)). use various measures from the National Longitudinal Survey of Youth (NLSY) and the Current Population Survey (CPS) to assess this channel. Exploiting state-level variation in the level of minimum wages and applying a TWFE design, they find no discernible change in fringe benefits (employer provision of health insurance, pension coverage, dental insurance, vacation pay, and training/educational benefits) and working conditions (shift work, irregular shifts, and workplace safety) of low-wage workers. An alternative approach considers the impact on firm-level expenditures on fringe benefits ([Card and Alan, 1995b](#)). and ([Brown, 1999](#)) review the earlier literature, finding a very small or no reduction at all in these expenses. More recently, ([Harasztsosi and Lindner, 2019](#)) find no indication of a decrease in fringe benefits in response to a large increase in the minimum wage in Hungary.

In addition to this broader evidence, the literature paid special attention to two specific elements: employer-provided health insurance and on-the-job training.

Employer-provided health insurance. A particularly important fringe benefit in the U.S. context is employer-provided health insurance ([Marks, 2011](#)). ([Meiselbach and Jean, 2023](#)) and ([Clemens et al., 2018](#)) study the relationship between health insurance and the minimum wages by exploiting various survey data (Current Population Survey, Medical Expenditure Panel Survey, American Community Survey). They document a small reduction in providing health insurance for low-skilled employers, especially at small firms

(Clemens et al., 2018). also provide a back-of-the-envelope estimate implying that roughly 9–16 % of the wage increase of low wage workers is offset via reductions in insurance. This suggests that workers benefit substantially from the policy even after any fringe benefit offset is taken into account.

On-the-job Training. How minimum wages impact on-the-job training is unclear a priori. In the standard competitive framework, general training will be financed by employers in the form of lower wages. In response to the minimum wage, employers will offer less on-the-job training, which could partly offset the positive impact of the policy. However, in the presence of imperfect competition in the labor market, the impact on training is uncertain (Acemoglu and Pischke, 2003). show that when firms earn some rent on workers, it might be more profitable for employers to respond to a minimum wage by increasing worker productivity through additional training rather than laying them off.

The early empirical literature on training often confirmed the prediction of the neoclassical model on the negative consequences of minimum wages on training (Shiller, 1994; Neumark and Wascher, 2001b), although not always (Grossberg and Sicilian, 1999; Acemoglu and Pischke, 2003). provide a critical review of this literature and conduct their own empirical analysis. Contrary to the prediction of the neoclassical model, they find no discernible effect in on-the-job training in the U.S. context. Outside of the U.S. context, (Hara, 2017) finds a significant reduction in firm-provided training, while (Arulampalam et al., 2004) and (Bellmann et al., 2017) find limited effects on-the-job training in the U.K. and Germany, respectively.

Indirect evidence on amenities. While direct evidence on the impact on amenities is limited, there is a sizable body of indirect evidence suggesting that the offset in non-wage amenities is likely to play a limited role in the adjustment to the minimum wage. First, studies consistently find that worker turnover and quit rates decrease in response to the minimum wage (see Section 4.1.10 for details)—a sign that workers like their jobs more when the minimum wage goes up. Consistent with this interpretation, (Holzer et al., 1991) show that minimum wage jobs attract more job applicants than jobs that pay either slightly more or slightly less than the minimum wage. Second, overall job satisfaction increases among low-wage workers following the minimum wage (Bossler and Broszeit, 2017; Güral and Ayaita, 2020), suggesting that the wage increases are not substantially offset by reductions in amenities that workers value. Finally, a large amenity offset would be inconsistent with the significant adjustments along other margins such as output prices that we discuss below. After all, there would be no need (or scope) for raising prices if firms were to compensate most of the wage increase through cutting non-wage compensation.

4.1.3 Substitution with other inputs

In Section 3.3, we concluded that the evidence suggests that the employment and hours responses are likely to be limited in many contexts. A closely related

literature directly inspects the channels through which employment may be affected. In the standard neoclassical framework, one such channel is the substitution between low-wage workers and other inputs in production, including higher-skilled workers and capital (see, e.g., Eq. (1)). Here we review that evidence to better understand the mechanisms, while also considering their implications for low-wage employment overall (as opposed to employment at particular firms, occupations, etc.).

Capital-labor substitution. Based on firm-level evidence from Hungary, firms exposed to minimum wages seem to increase their investment (and hence the capital stock) relative to the unexposed firms (Harasztsosi and Lindner, 2019; Geng et al., 2022). and (Hau et al., 2020) also find an increase in investment and the capital stock in the Chinese manufacturing sector (Fan et al., 2021). find a firm-level increase in robot adoption in China in certain periods but not in others. In both countries, some reduction in employment at highly-exposed firms is found as well (relative to less-exposed ones), which is in line with the prediction of the standard theory: the higher labor cost forces firms to substitute labor with capital, e.g., through increased automation (see Eq. 1). However, it is important to note that these firm-level effects may not reflect the market-level impact on capital stock, since less exposed firms could reduce their investment in response to the minimum wage. Unfortunately, there is little direct evidence on market-level responses in the stock of capital.

Several papers study this capital adjustment indirectly through the lens of a changing occupational structure. Here, the key question has been whether minimum wages lead to a reduction in automatable/routine jobs. Both (Lordan and Neumark, 2018) and (Aaronson and Brian, 2019) find that the minimum wage reduces the employment of the lowest-wage workers in routine automatable occupations. However, the two studies disagree on the overall employment implications of these results (Lordan and Neumark, 2018). argue that the reduction in automatable jobs contributes to a lower overall employment level of low-wage workers, while (Aaronson and Brian, 2019) find that workers laid off from routine jobs are able to move into other expanding low-wage occupations. The broader evidence on employment reviewed in Section 3.3 appears to be more consistent with the interpretation of (Aaronson and Brian, 2019).

Minimum wages may also reduce automation in certain scenarios (Downey, 2021). studies the impact of minimum wages on higher-paying, automatable, jobs. He argues that higher minimum wages decrease the incentive to adopt de-skilling automation technologies (that replace more expensive middle-skill jobs with cheaper lower-skilled ones). His empirical analysis finds that an increase in the minimum wage significantly decreases mid-skill IT employment while increasing routine jobs, consistent with lower automation.

Labor-labor substitution. Besides substituting labor with capital, firms can also replace lower-skilled workers with higher-skilled ones. Such substitution would show up as heterogeneous employment impact of the policy by skill groups, e.g. younger minimum wage workers could be replaced by older,

more experienced ones. We reviewed the evidence on the heterogeneous impact of the policy in [Section 3.3.3](#) where we find little systematic indication of such heterogeneity along key demographic dimensions.

Some papers also study the impact of minimum wages on firms' skill demand as evidenced by the demographic composition within high-impact jobs or firms ([Clemens et al., 2021](#)). find that workers in low-wage occupations are more likely to be older and have a high school degree following minimum wage changes. They also document an increase in skill requirement in low-wage occupations using Burning Glass data on job postings ([Horton, 2018](#)). implements a field experiment on an online platform and shows that imposing a minimum wage on a random subset of firms led those firms to hire more productive workers in an online platform. These pieces of evidence suggest that firms may shift their demand towards more productive minimum wage workers when forced to raise their pay. However, whether such firm-level changes in skill demand lead to greater disemployment in general for the least productive workers is a distinct question. For example, displaced workers from narrow categories of jobs may be absorbed elsewhere.

A separate strand of the literature exploits discontinuities in the level of minimum wages by age to estimate substitution. As we discussed in [Section 3.2.5](#) the evidence from such exemptions should be carefully interpreted since they do not directly speak to the impact of broad-based minimum wage shocks—the focus of this chapter. Even if there is significant labor-labor substitution when similar workers are allowed to be paid differently, such substitution may not be possible in response to an across-the-board minimum wage.

The evidence based on such age exemptions is mixed. In many contexts, firms apparently do not utilize sub-minimum wages to begin with, and pay the same wages to younger workers even when it is possible to pay them less. This is true in the U.S. (see ([Katz and Alan, 1992](#); [Card et al., 1994](#)) who find limited utilization, while ([Neumark and Wascher, 1994](#)) find more utilization in certain states); in the U.K ([Giupponi and Machin, 2022b](#)); and in Finland ([Böckerman and Uusitalo, 2009](#)). In other countries, we do see greater utilization of age-specific floors, leading to reduced wages for younger workers and sometimes to higher employment (see ([Kreiner et al., 2020](#)) in Denmark; ([Kabátek, 2021](#)) in the Netherlands; ([Portugal et al., 2006](#)) and ([Pereira, 2003](#)) in Portugal). These estimates suggest that in some cases, applying the minimum wage policy selectively to different groups could create sizable employment differences and inequality between treated and untreated workers.

4.1.4 Firm-entry and exit

In some models, exit and entry of firms play an important role in mediating the response to wage floors, as the increased cost of labor drives out low-productivity firms. In terms of firm exit, ([Draca et al., 2011](#); [Harasztsosi and Lindner, 2019](#); [Mayneris et al., 2018](#)), and ([Dustmann et al., 2022](#)) find that exposed firms exit at a faster rate following a minimum wage hike. In the U.S.,

(Rohlin, 2011) finds no change in the exit rate, while (Aaronson et al., 2018) and (Chava et al., 2023) find a significant increase. Furthermore, (Luca and Luca, 2019) document an elevated exit rate for restaurants, especially among establishments with very low consumer ratings to begin with. This latter finding is consistent with a “cleansing” effect of the minimum wage, where the most inefficient firms exit the market.

What is the contribution of firm exit to overall employment? This critically depends on whether workers at (likely low-productivity) exiting firms are able to find jobs elsewhere or are more likely to remain without employment. As (Dustmann et al., 2022) show in the context of a model with monopsonistic competition (see their Online Appendix D), even if some small, inefficient firms exit the market, workers may move to more productive, surviving companies.³⁴ In this case, the employment effect of the policy would remain limited, while overall productivity improves. Therefore, the results on firm exit cannot be interpreted as direct evidence of loss in net employment. Empirically, findings from Germany documented by (Dustmann et al., 2022) as well as U.S. evidence by (Rao and Risch, 2024) are consistent with such productive reallocation. In a somewhat different context in South Africa, (Bassier, 2022) also finds a similar reallocation of employment from lower productivity to higher productivity firms in response to a higher sectoral wage floor set by collective bargaining.

The effect on firm entry has also received attention in the literature, as it is unclear *a priori* what to expect in this margin. The standard neoclassical channel suggests that the increased labor cost would dampen firm entry. However, in the presence of some friction in the adaptation of technology, this prediction can be reversed (Aaronson et al., 2018). study the impact of the minimum wage in a putty-clay model. In this framework, newly entering firms can freely choose their input mix, allowing them to substitute costly low-wage workers with capital. Once this choice is made, however, firms operate using a fixed input mix: so incumbent firms’ capital-labor substitution possibilities are restricted.

This framework predicts that short-term employment responses to the minimum wage will be muted, reflecting only scale effects, while the larger effects from substitution will come through firm dynamics. Incumbent firms with *ex post* suboptimal technology will exit slowly, creating entry opportunities for new firms that are able to choose a more efficient mix of inputs. Over time, obsolete labor-intensive technologies will disappear and will be replaced with capital-intensive technologies—leading to reduced employment among low-wage workers. Of course, this is only one theoretical possibility; different types of friction would suggest different compositions of firm entrants. For

³⁴ Imperfect competition in the product market can also lead to a similar type of re-allocation (see, e.g. Rao and Risch, 2024).

example, in a model with monopsonistic competition, higher-productivity firms may be more likely to enter following a minimum wage rise. It is worth noting that the muted short-term response followed by a more amplified long-term one is not a pattern that we see in much of the data (see [Section 3.3.3](#) on evidence on short- and medium-term employment effects).

Moving on to direct empirical findings on firm entry, ([Aaronson et al., 2018](#)) find an increase in entry rate in response to minimum wages, ([Harasztosi and Lindner, 2019](#)) find no change, while ([Rohlin, 2011; Draca et al., 2011](#)), and ([Luca and Luca, 2019](#)) find reduced entry. Therefore, the evidence is mixed but broadly suggests a fall in the entry rate.

Overall, the weight of evidence on firm dynamics suggests that the number of firms probably falls following a minimum wage increase (as the exit rate increases, while the entry rate likely does not rise—at least not strongly). In line with this, ([Orazem and Peter Mattila, 2002](#)) and ([Dustmann et al., 2022](#)) find a decrease in the number of firms resulting from a higher wage floor.

4.1.5 Migration and participation

Minimum wages can also shape the local working population through various mechanisms, including changes in labor force participation and migration responses.

Migration. The relationship between minimum wage levels and migration patterns is theoretically ambiguous. If jobs become more scarce as a result of the policy, one potential adjustment mechanism is for low-skilled workers to relocate to higher wage areas, where the wage floor is less binding. On the other hand, absent a sizable reduction in job finding rates, a higher minimum wage can also act as a magnet for low-skilled workers. For instance, in the monopsonistic competition framework, where many firms are supply-constrained, employers are willing to hire more workers at higher wages—something that could induce migration to locations with higher minimums.

Several studies have analyzed the effect of minimum wages on the composition of the local population. An early study by ([Castillo-Freeman and Richard, 1992](#)) examined the impact of a substantial increase in the minimum wage in Puerto Rico and identified the out-migration of low-skilled workers to the mainland United States as a significant adjustment mechanism. However, we should note that ([Krueger, 1994](#)) points to the fragility of the empirical designs applied in ([Castillo-Freeman and Richard, 1992](#)). More recently, ([Monras, 2019](#)) studied the internal migration pattern in the U.S. context by examining the impact of 441 state- and federal-level minimum wage changes instituted between 1985 and 2012. He finds that minimum wages reduced both the low-skilled employment and the low-skilled population share. The main analysis is complicated by the lack of parallel trends, which might be related to selective timing of federal minimum wage increases (which are parts of the events used in his analysis) ([Minton and Wheaton, 2023](#)). assess the change in

the low-skilled population (instead of the change in the share of low-skilled workers as in (Monras, 2019) and (Castillo-Freeman and Richard, 1992)) using the prediction probability approach. They rely on the American Community Survey, which provides more precise information on population across locations than the Current Population Survey used by (Monras, 2019; Minton and Wheaton, 2023). find an increase in net migration of low-skilled workers towards higher minimum wage states—a finding that is in stark contrast with the earlier U.S. literature.

Some papers also studied the impact on migration in the European context. Since there is no variation in statutory minimums within countries, nation-wide minimum wages could affect internal migration by inducing individuals to move away from (or toward) locations with a lower average wage and more binding policy. However, (Giupponi et al., 2024) and (Ahlfeldt et al., 2018) find no indication of a significant response of internal migration to minimum wages in the U.K. and in Germany, respectively.

A closely related literature studies the impact of the minimum wage on immigrants' location decision. The evidence on this aspect of the policy is somewhat mixed (Orrenius and Zavodny, 2008). and (Cadena, 2014) find that low-skilled immigrants have been discouraged from settling in states with higher minimum wages. In contrast, (Giulietti, 2014) finds that the federal minimum wage changes in 1996–1997 and 2007–2009 induced a sizeable flow of low-skilled immigrants to more affected states (Boffy-Ramirez, 2013). suggest that the impact of the policy depends on the amount of time immigrants have spent in the country: they find that a higher minimum wage attracts migrants who have already been in the U.S. for 2 to 4 years. However, for immigrants who have been in the U.S. for less than 2 years, or longer than 4 years, the minimum wage effect is indistinguishable from zero.

Overall, the evidence suggests that changes in low-skilled population at locations that raise minimum wages are unlikely to be large. The evidence on immigrant population is more mixed, but interestingly, the more mobile immigrants seem to favor places with higher minimum wages. This latter finding is consistent with the prediction of the monopsonistic competition framework.

Participation. Even if the size of the local population does not change, there could be a change in participation rate in response to the minimum wage. The evidence on participation decisions is mixed (Wessels, 2005). and (Luna-Alpizar, 2019) report a decrease in participation among young individuals, and (Lavecchia, 2020) finds a negative (but imprecise) effect among low-skilled individuals. Conversely, (Laws, Athene, 2018) and (Agan and Makowsky, 2021) find that an increase in the minimum wage leads to greater labor force participation (Luna-Alpizar, 2019). finds an increase in participation for prime-age workers with high school education, (Borgschulte and Cho, 2020) and (Hampton and Evan Totty, 2023) find positive effects on individuals near retirement age, while (Godøy et al., 2024) show similar patterns for parents of

young children. Finally, (Cengiz et al., 2022), using the prediction-based approach, find no evidence of changes in participation overall, and find no indication of significant heterogeneity across age, race, and education.

In addition to the change in the participation rate, the effective labor force could also increase through increased effort by the unemployed to find a job. The evidence on the impact of minimum wages on search effort is limited and mixed. A few studies document small responses (Laws, Athene, 2018). finds no clear evidence that minimum wages affect search effort, with some suggestive evidence that the unemployed may decrease it in response to the policy (Adams et al., 2022). finds a very short-lived positive effect only during the month of the minimum wage change, which vanishes soon afterwards. In contrast, recent evidence by (Piqueras, 2023) indicates that the policy has a significant long-term impact on the search effort of low-wage unemployed.

4.1.6 Wage retrenchment of higher-skilled workers

Another channel through which businesses can absorb higher labor costs for low-wage workers is by reducing pay for higher-wage employees. Is this theoretically possible? To the extent that wages reflect rent sharing, higher pay for some could lead to lower pay for others by reducing the surplus being bargained over. This would imply a *negative* wage spillover or ripple effect higher up in the pay distribution, especially within firms. We will review the evidence on spillovers and wage inequality in Section 5. However, the evidence is clear that there are *positive* (but limited) wage spillovers higher up in the distribution. This is also true within firms (Hirsch et al., 2015; Gopalan et al., 2021; Dube et al., 2019).³⁵ (Yao et al., 2023) also find a positive relationship between executive compensation and the minimum wage hikes in China. In conclusion, the available empirical evidence indicates that the increase in the minimum wage for low-wage workers is not financed by a reduction in the wages of their better paid counterparts.

4.1.7 Output prices and consumers reactions

Output prices. There is an extensive literature studying the effect of the minimum wage on output prices. Table 4 lists the published papers at the time of writing. The evidence on prices comes from various sectors of the economy: restaurants, fast-food, manufacturing, and retail. The evidence almost uniformly suggests that output prices increase following a minimum wage change. The magnitude of the price increase is also noteworthy. In the last column of Table 4, we report the pass-through rates: the size of the change in the unit

³⁵ (Clemens et al., 2018) documents a decline in employer-sponsored health insurance in occupations moderately higher up the wage distribution. They calculate that these amenity offsets for above-minimum wage workers represent a larger share of the observed wage increase than analogous offsets from minimum wage workers. Therefore, wage spillover estimates can overstate the spillovers in compensation, and hence the increase in workers' utility.

price of the output relative to the change in unit labor costs. It is worth noting that the latter is typically imputed; as a consequence, the evidence on the pass-through rate should be taken with a grain of salt. That caveat notwithstanding, most papers find full pass-through. In other words, the unit labor cost increase is fully covered by the unit price increase; and in some cases, we see even more than full pass-through.³⁶

The evidence therefore shows that there is a clear and sizable increase in output prices following minimum wage increases. However, the corresponding changes in the quantity and quality of the products are less well understood. In the retail sector, (Leung, 2021) and (Alonso, 2022) find an *increase* in quantities following the changes in the minimum wage, while (Renkin et al., 2022) find no effect on quantities (their point estimate is negative and somewhat noisy). In the case of hotel service, (Agarwal et al., 2024) find no price pass-through, but at the same time a significant reduction in hotel occupancy following minimum wage changes. There are other puzzling aspects of their results, as they find a decline in output prices for unbranded and upscale hotels accompanied by a large reduction in occupancy rate—suggesting that other demand shocks may have coincided with the minimum wage changes they study.

(Cooper et al., 2020) find nominal spending increases more than price gains, suggesting that, on net, consumers purchase more food, both at and away from home. For durable goods, they find that cumulative nominal spending increases roughly in line with prices when the minimum wage rises, suggesting no change in quantity (Harasztosi and Lindner, 2019). evaluate the revenue response to the minimum wage and find a significant increase in the local service sector, but a reduction in the manufacturing sector. This indicates that the quantity response to the output price increase—and hence the scale of production—could vary substantially across sectors. As discussed in Section 3.3.3, researchers seem to find limited dis-employment effect in the local service sector, suggesting that the decline in the scale of production is likely to be limited there. At the same time, there is more evidence of negative employment effects in the tradable/manufacturing sector, in line with the larger reduction in the scale of production there.

The evidence on changes in production quality is also limited, but points in a positive direction (Ruffini, 2022). finds significant improvement in patient health and safety (e.g., measured by patient deaths) in nursing homes after

³⁶ Here we focus on price pass-through. Note, however, that even if output prices were to increase one-to-one with the change unit cost cost (full pass-through), this does not necessarily imply that all of the costs associated with the minimum wage are borne by consumers. The latter depends on consumer reactions, and revenue responses. For example, (Harasztosi and Lindner, 2019) find a close to full price pass-through, but only 75 % of the total wage bill increase is paid for by consumers. The discrepancy between the two can be explained by quantity responses to the price change.

workers' earnings increase due to minimum wage changes ([Brown and Chris, 2023](#)). find improved service quality in the child care context. However, there is much less evidence on service quality in major low-wage sectors such as hospitality and retail.

4.1.8 Input prices and rent

In addition to being passed downstream to customers, the increase in labor costs could also be passed upstream to suppliers. This could be a quantitatively important channel, as 30–75 % of firm-level expenses (depending on industry) are related to the purchase of intermediate goods and services.

In many sectors, such as U.S. retail and agriculture, the primary cost of business is rent. One way firms might finance an increase in labor costs is by reducing the rent they pay. In fact, if rental markets are sufficiently competitive and the land and properties used for lower-skill-intensive production cannot easily be repurposed by businesses relying on higher-skilled workers, the main adjustment to the minimum wage could come from lowering rental or land prices. Unfortunately, we are not aware of any direct evidence on rental price adjustments, and future research should focus on assessing the empirical relevance of this channel. However, in some cases, businesses own the land or stores they use. If minimum wage increases are primarily passed through to landowners, profits would decline for these land-owning firms. We review the evidence on profits in the next section.

In addition to adjusting rents, firms can also lower the prices they pay their suppliers. The extent of pass-through to upstream firms depends on those firms' own exposure to the minimum wage. If minimum wage-intensive firms tend to purchase goods from other minimum wage-intensive suppliers, the extent of pass-through will be limited. Therefore, the sorting of suppliers and buyers based on the type of labor they employ is an important consideration ([Demir et al., 2024](#)). find strong positive assortativity in wages within supplier-buyer relationships, suggesting the possibility of limited pass-through.

Direct evidence on changes in intermediate goods expenses following minimum wage increases is rather limited ([Harasztosi and Lindner, 2019](#)). find that total spending on materials increased in the short term, while in the medium term, the effect on materials is smaller and insignificant. This evidence suggests that lowering supplier prices or rent is not a major margin of adjustment.

4.1.9 Profits

Given the stated redistributive goal of minimum wage policies, the impact on profits is a margin of considerable interest. Economists have developed various approaches to estimate the effect of minimum wages on firm profitability. One approach infers profitability from financial market reactions. The evidence from this approach presents a mixed picture ([Card and Alan, 1995b](#)). find no

changes in the stock market valuation of firms employing low-wage workers around the time of U.S. minimum wage announcements, while (Bell and Machin, 2018) find significant reactions in the U.K. A key limitation of the financial market event study approach is that initial market reactions reflect investors' beliefs about the policy's impact, which may differ from the actual responses that manifest over time. Moreover, these event study estimates of abnormal returns are meaningful only to the extent that these announcements contain information not already anticipated (and thus priced in) by the market. In this regard, the surprise announcement in the U.K. studied by (Bell and Machin, 2018) offers a more informative case.

A second strand of this literature evaluates actual firm-level changes in profitability, typically measured as the profit-to-revenue ratio. Using data from the U.K., (Draca et al., 2011) find a reduction in profits in response to minimum wage policies, while (Mayneris et al., 2018) find no such change in China. A limitation of these estimates is that a decline in the profit-to-revenue ratio could result from either a reduction in profits (the numerator) or an increase in revenue (the denominator), possibly due to higher output prices. To avoid this complication, some studies have considered changes in profits themselves as the outcome (Harasztsosi and Lindner, 2019). exploit a large (60 %) minimum wage increase and find a modest fall in profits, amounting to around 20 % of the labor cost increase at highly exposed firms (relative to less-exposed ones) (Drucker et al., 2021). apply a similar methodology and find a more substantial reduction in profits in Israel.

Three recent studies exploit state-level variation in minimum wages in the U.S., contrasting with the firm-level variation in exposure to nationwide minimum wage shocks used in the studies mentioned in the previous paragraph. In principle, state-level variation provides a more holistic picture of profitability than cross-firm evidence (Vergara, 2023). examine the effect of minimum wages on state-level profit (gross operating surplus) per establishment and find a significant reduction in profits following minimum wage increases. Conversely, (Rao and Risch, 2024) study the impact of the minimum wage on independent (pass-through) businesses in exposed industries and find no indication of a significant reduction in profits. They document an increased rate of firm exit (which obviously represents a reduction in profits for those firms) but also an increase in profits among surviving firms. Finally, (Coviello et al., 2022) investigate the impact of minimum wages on a large U.S. retailer. Depending on the specification, they find either a small and insignificant positive effect or a significant negative effect on profits, with estimates ranging from a 5 % increase to a 16 % decrease. Their estimates, along with the standard errors, can rule out profit reductions greater than 35 %.

To summarize, most (though not all) studies find evidence of some reduction in profits. However, the magnitude of the profit reduction reported in the literature tends to be modest.

4.1.10 Worker turnover and reduction in training costs

A key adjustment channel predicted by various models of dynamic monopsony (Burdett and Dale, 1998; Manning, 2003) is that minimum wages reduce worker turnover (separations). In this context, a higher minimum wage makes relatively lower-paying jobs more attractive, thereby reducing job-to-job separations. A substantial body of research has tested this prediction by empirically evaluating turnover responses.

Table 5 summarizes the key results on turnover from these studies. Of the 14 studies that directly examine the impact on turnover, 12 find a significant reduction in worker turnover rates. In one case, the evidence is mixed, with results depending on the specification, while another study finds no reduction. Therefore, the vast majority of the literature suggests that a reduction in turnover plays an important role in the response to minimum wage increases. This evidence also confirms a key prediction of the dynamic monopsony framework and, more broadly, underscores the role of labor market frictions in understanding the impact of minimum wages.

In addition to changing labor supply to the firm, lower turnover can also be a source of productivity improvement (evidence on overall productivity measures is discussed in more detail below). First, a lower turnover leads to savings in training and recruitment costs (Dube, Arindrajit et al., 2010). Second, workers tend to be more productive when they stay longer at the same job and accumulate more experience. More research is needed to better understand the magnitude of these productivity improvements resulting from lower turnover.

Finally, the reduced turnover results suggest that we should be cautious when interpreting evidence on minimum wage effects on vacancy posting (see, e.g., Kudlyak et al., 2023; Clemens et al., 2021), as these may simply reflect lower separations.³⁷

4.1.11 Productivity

The productivity enhancing effect of minimum wages has long been hypothesized. However, direct evidence on the impact on revenue has been scarce until recently.

Revenue-based productivity. The initial evidence on the productivity channel was based on firm-level revenue measures, such as revenue-based total factor productivity. Notable examples include (Mayneris et al., 2018; Riley et al., 2017), and (Hau et al., 2020), all of whom find a significant increase in revenue-based productivity following minimum wage changes. However, such an increase may simply reflect a rise in output prices rather than an actual increase in the quantity of goods or services produced. This concern is not merely hypothetical; as discussed in Section 4.1.7, the literature on output

³⁷ Reduced turnover can also mean that fewer workers are employed over the course of the year at a firm, even if there is no change in the number of jobs at any given time. This is particularly relevant for interpreting annual payroll data (see, e.g., Rao and Risch, 2024).

prices suggests that firms often respond to minimum wage hikes by raising prices. Researchers have been aware of this issue, and some have attempted to adjust for firm-level price changes by controlling for industry-level price indexes. However, these sector-wide measures are unlikely to capture the substantial variation in firm-level exposure within sectors, making the resulting estimates difficult to interpret.

Quantity-based productivity. Conceptually, a better approach would be to use a quantity-based measure of productivity that removes price effects, providing a clearer picture of actual efficiency improvements. In practice, however, such measures are usually not available for all or most firms in the economy. As a result, studies following this approach have provided evidence for a select set of sectors where such quantity-based efficiency calculations are feasible: ([Ku, 2022](#)) in agriculture, ([Coviello et al., 2022](#)) in retail, and ([Hau et al., 2020](#)) for exporting manufacturing firms. All these papers find evidence of increased quantity in response to the minimum wage. Additionally, a few recent studies have examined the impact on quality: ([Ruffini, 2022](#)) finds an increase in quality in the nursing home sector, while ([Brown and Chris, 2023](#)) report improved service quality in the child-care setting.

Reallocation. Finally, as the discussion on entry and exit indicated, minimum wages could also affect productivity by improving the composition of jobs in the economy ([Mayneris et al., 2018](#); [Dustmann et al., 2022](#)). and ([Rao and Risch, 2024](#)) find evidence for such a channel in Germany and the U.S., respectively. Both studies show that in response to minimum wage increases, small, inefficient firms exited the market, and workers at these firms found employment at more productive firms. However, there could be nuanced welfare consequences from such reallocation. For example, while the exit of low-productivity employers might enhance worker productivity, it could also harm some consumers by reducing variety and diminishing product market competition.

Productivity and employment. How improved productivity affects employment is theoretically ambiguous, as it depends on the price elasticity of output demand. If output demand is highly elastic, increased production can be easily accommodated without changing the output price. On the other hand, in industries where output demand is inelastic, higher per-worker productivity means fewer workers are needed to meet consumer demand. In these sectors, some form of quality improvement that shifts the output demand outward would be necessary to explain limited employment effects.

4.2 Summary of evidence on margins of adjustment

The limited role of employment adjustment, as discussed in [Section 3.3](#), suggests that other margins likely play a more significant role in responding to minimum wage increases. Based on the growing body of evidence reviewed above, we now have a better understanding of how these other margins of

adjustment help firms absorb higher labor costs. Most importantly, passing the increased costs on to consumers in the form of higher prices plays a crucial role. This strategy, combined with the seemingly limited decline in consumer demand in response to price changes, covers a substantial share of the cost increase. While the exact proportion remains an open question, (Harasztsi and Lindner, 2019) suggest that around 80 % of the change in labor cost can be attributed to this channel.

The evidence also indicates that improved efficiency is another key mechanism for absorbing higher labor costs. The productivity-enhancing effects can manifest in various ways, including lower training costs and turnover, improved operational efficiencies, and greater worker-level productivity. These factors help absorb some of the labor cost increase without a significant reduction in employment. Meanwhile, adjustments to profits appear to be limited, although the evidence on this margin is still evolving and somewhat tentative. For example, (Harasztsi and Lindner, 2019) suggest that around 20 % of the labor cost increase could be offset by a reduction in profits.

These conclusions—drawn from studies that directly evaluate the causal impact of minimum wages on various margins of adjustment—are remarkably consistent with what business executives themselves report about their reactions to the policy. For example, executives generally perceive limited scope for reducing employment in response to minimum wage increases (Lester, 1946; Levin-Waldman, 2000; Reich and Laitenen, 2003; Bodnár et al., 2018). At the same time, they almost always emphasize the importance of adjusting output prices and improving management and efficiency. Specifically, in (Lester, 1946)'s seminal study, executives highlight increasing sales efforts and enhancing productivity as the most relevant margins of adjustment. Similarly, (Reich and Laitenen, 2003) report survey evidence on managers in San Francisco, showing that the two most important responses to a minimum wage increase are raising prices and improving efficiency (Hirsch et al., 2015). find that restaurant managers in Alabama and Georgia see a significant role for implementing cost-saving, performance-improving standards in response to the minimum wage. Finally, based on a representative sample of firms in eight Eastern and Central European countries, (Bodnár et al., 2018) document that managers identify raising product prices, cutting non-labor costs, and improving efficiency as the most relevant channels of adjustment.

In summarizing the evidence on these channels of adjustment, it is important to acknowledge the main gaps in our knowledge. We have extensive evidence on employment (including substitution and firm dynamics), output prices, and worker turnover. There is also a growing body of work on fringe benefits, profits, and productivity, but further research in these areas would help refine our quantitative understanding of these adjustment margins. In contrast, the evidence on training is more limited and somewhat outdated, and there is very little research on service quality and consumer demand (e.g., overall sales). Additionally, we lack sufficient understanding of how rents or

the prices of intermediate goods and services respond to minimum wage increases. These under-studied channels represent particularly important areas for future research.

4.3 Modeling implications and open questions

Evidence on these various margins of adjustment suggests that simple explanations for understanding the impact of minimum wages are insufficient. For instance, the competitive neoclassical model correctly emphasizes the role of price pass-through but fails to predict the lack of employment effects—except under the assumption of completely inelastic output demand. In contrast, the presence of imperfect competition in the labor market offers a natural explanation for the limited employment effects but falls short in accounting for a substantial increase in output prices or a limited pass-through to firm owners in the form of reduced profits.³⁸

The evidence on the productivity-enhancing effects of the policy also suggests that insights from institutionalist economists of the 1940s, as well as efficiency wage models, may capture important aspects of the policy response (Hirsch et al., 2015). Economic models of low-wage labor markets often presume full optimization within constraints ensured by competition and free entry of entrepreneurs. However, the pattern of productivity improvements following minimum wage hikes indicates that the discipline of market competition may not be strong enough to ensure that firms operate at the production possibilities frontier. Consequently, firms' X-efficiency could play an important role in understanding the impact of minimum wage policies (Leibenstein, 1966).

Given the current state of the literature, it would be overly ambitious for this review to propose a unifying framework for the low-wage labor market that incorporates all these considerations. Instead, we pose some open questions and puzzles that need to be addressed in future research.

Imperfect competition. Over the past two decades, there has been considerable improvement in modeling the implications of minimum wages in the presence of various types of labor market frictions. This includes models that characterize minimum wage effects in the presence of search frictions (Van Den Berg, 2003; Flinn, 2006, 2010; Engbom and Christian Moser, 2022; Manning, 2003; Dube et al., 2016; Brochu and David, 2013), monopsonistic competition (Dickens et al., 1999; Manning, 2003; Haanwinckel, 2023), oligopsonistic competition (Bhaskar and To, 1999; Berger et al., 2022; Azar et al., 2023), or monitoring problems (Rebitzer and Lowell, 1995). Together, these

³⁸ In the standard monopsony framework, when minimum wages are increased slightly from the optimally set wage, the effect on profit will be zero. That framework also predicts increases in labor supply and lower output prices (see page 2108 in (Brown, 1999)). Interestingly, monopsonistic competition models, such as (Card et al., 2018), predict a rent-sharing type of response to the minimum wage: if all firms are forced to raise wages, firm-level employment will not be affected, but profits will decline.

studies provide a rich and flexible approach to analyzing the impact of the minimum wage on labor market outcomes such as employment level, job flows, and the wage distribution.

In many cases, the presence of friction dampens the employment effects of the policy relative to a frictionless world, but these models often predict considerable dis-employment under realistic parameterization. This result is partly related to the imposition of the free-entry and zero-profit assumption, which leaves very little room for rent sharing or profit reduction to occur in the constrained equilibrium. However, most of these models focus solely on labor market interactions and tend to ignore adjustment on other margins. For example, the models often overlook the extent to which firms can pass through the minimum wage to consumers without a change in output, or the productivity-enhancing effects of the policy. Therefore, a natural next step would be to incorporate these dimensions in existing models to enrich our understanding of how imperfections in the labor market interact with imperfections in the product market.

Productivity. To guide modeling choices, we first need a better understanding of the relative importance of various sources contributing to the productivity-enhancing effects of minimum wages: (i) worker effort (efficiency wage) mechanisms, (ii) turnover reduction, (iii) reductions in X-inefficiency or improved management practices within firms, and (iv) reallocation from low- to high-productivity employers, including the exit of low-productivity firms. Additionally, it is crucial to determine whether these effects are driven by the market-wide nature of minimum wage shocks (i.e., all firms being affected) or if the productivity enhancement would also be present in response to firm-level shocks.

Exploring the interaction between labor market imperfections, such as monopsony power, and X-efficiency is a particularly intriguing direction for future research. In markets characterized by imperfect competition, both in labor and product markets, firms making suboptimal choices can still survive. Consistent with this idea, recent evidence suggests that significant productivity improvements are possible through the implementation of better management practices (Bloom et al., 2013), especially when competitive forces are not strong enough (Bloom et al., 2014). Moreover, there is some evidence that substantial productivity gains can occur even with firm-specific pay raises, and in some cases, firms may benefit more from these efficiency gains than they lose from paying higher wages to their workers (Emanuel and Harrington, 2020).

Relatedly, a growing body of evidence shows that firms often fail to set prices and wages optimally, instead opting for relatively “clunky” decision rules. National chains frequently use uniform pricing in both product and labor markets (DellaVigna and Gentzkow, 2019; Hazell et al., 2022). The presence of such “mispricing” creates additional opportunities for productivity improvement from a wage floor. For instance, (Coviello et al., 2022) document significant productivity effects and an increase in profits following local-level minimum wage hikes for certain stores of a U.S. nationwide retailer that employs a uniform compensation

structure across stores (except for compliance with local minimum wages) along with nationally uniform product pricing.³⁹

Interestingly, the presence of imperfect competition offers a rationale for the prevalence of such optimization frictions. When wages are a choice variable, modest optimization errors have only second-order consequences on profits (due to the envelope theorem), even as they have a first-order impact on workers (see (Dube et al., 2020) for more). More research is needed on the impact of firm-level wage shocks on productivity and how this relationship relates to the pre-existing strength of competition.⁴⁰

Consumer demand. As we have seen, a key channel through which firms respond to the minimum wage is by adjusting output prices. Therefore, understanding how output demand changes in response to minimum wage increases is crucial. The apparent contradiction between limited employment effects and significant output price effects has long been recognized in the literature (see, e.g., Brown, 1999; Aaronson and French, 2007). As Eq. (1) in Section 2.2 shows, a price increase typically leads to a decrease in output demand, which in turn reduces the demand for workers.⁴¹

Therefore, the increase in prices and the limited employment effects are consistent with inelastic output demand (see, e.g., (MacCurdy, 2015)). However, there are several explanations for this inelastic consumer demand, each with different implications for welfare. First, it is possible that consumer demand for products that rely heavily on minimum wage labor is *generally* inelastic. This could be true at the market level even if individual consumers are more price-elastic. For example, in the fast-food industry, a market-level price increase might deter price-sensitive consumers but attract higher-spending consumers who previously avoided longer wait times. As a result, the industry as a whole might experience relatively more inelastic output demand due to a changing consumer base.

Another possibility is that this insensitivity is a *specific* feature of the price increases induced by a higher minimum wage. Below, we outline some leading explanations within this category.

³⁹ (Coviello et al., 2022) find a positive profit effect and a significantly larger increase in productivity when focusing on a subset of stores near the borders of minimum wage changes. When they evaluate the policy's impact on a more representative set by including all stores across all states, they find smaller productivity increases and a slight reduction in profit. This divergence is consistent with uniform prices being optimized for nationwide profits but not necessarily for local markets.

⁴⁰ Valuable insights could be gained from the case study approach of institutional economists or from a better understanding of how decisions within companies are made—a focus of organizational economics (see, e.g., Gibbons and Roberts, 2015).

⁴¹ One interesting explanation for this phenomenon comes from the putty-clay framework described in Section 4.1.4 (Aaronson et al., 2018). shows that the model predicts immediate output price effects but delayed employment reductions. However, much evidence suggests that medium-term employment changes are also limited, indicating that this explanation likely plays a limited role.

First, an important consideration lies in the interaction between product and labor market imperfections ([Bhaskar and To, 1999](#)). show that minimum wages can alleviate imperfections in the labor market while exacerbating imperfections in the output market. Specifically, firm exit can increase market concentration, allowing surviving firms to raise prices by more than the increase in labor costs. This prediction aligns with empirical evidence showing more than full pass-through of labor costs in certain contexts (see [Table 4](#)). In ([Bhaskar and To, 1999](#))'s framework, there is a negative relationship between output prices and output demand. However, in the presence of significant substitution between low-skilled labor and other inputs, they demonstrate that minimum wage increases could boost both employment and output prices. Intuitively, the minimum wage encourages firms to hire more workers as they move along the upward-sloping labor supply curve. If the additional workers allow firms to reduce capital usage, then overall output may decrease. Thus, the model predicts higher employment but lower capital in this knife-edge scenario. However, this prediction does not align with the empirical pattern observed in [Section 4.1.3](#), where capital use appears to increase following minimum wage hikes.

Another possibility is that minimum wage shocks are salient enough to enable firms to implement coordinated price changes. Such minimum wage-induced price increases might be perceived as fairer and more justified, leading consumers to tolerate them more. Additionally, minimum wage increases could improve overall product or service quality in the sector, which might attract additional consumer demand.

Finally, a common explanation for the muted output response relies on income effects. This argument suggests that a higher minimum wage increases the purchasing power of low-skilled workers, who are sometimes assumed to consume low-quality, minimum wage-intensive services and goods. This explanation, sometimes referred to as the "Hungry Teenager Theory" (see, e.g., ([Kennan, 1995](#))), seems less plausible than the other explanations discussed so far. Given that a small share of income is typically spent on low wage-intensive products (e.g., 5–7 % across countries), this limits the extent to which added income could fuel demand for such products. Furthermore, existing evidence on consumption responses to minimum wage increases suggests that low-wage workers tend to spend their extra income on durable goods such as vehicles, rather than on fast food (see ([Aaronsen et al., 2012](#))).

Understanding the source and nature of output demand inelasticity has important welfare implications. For example, if inelastic demand reflects improved quality, the price change may not result in a welfare loss. In contrast, if it indicates a lack of alternatives for consumers to substitute away from low-wage-intensive goods and services, the price effect could represent a pure welfare loss. Gaining a better understanding of the nature of the price response will be an important focus for future research on this topic.

Substitution towards other inputs. Another open question is why the standard substitution channel between lower-skilled workers and other inputs

appears to be dampened in the context of the minimum wage. A simple and straightforward explanation could be related to the production function: in the minimum wage sector, the technology may not allow for substitution between lower-skilled workers and other inputs.

A more fruitful approach is to explore the interaction between technology and imperfect competition in the labor market (Datta and Machin, 2024). provide an excellent recent example of this, showing that in the presence of imperfect competition, firm-specific wage floors can lead to an increase in the employment of low-skilled workers and a decrease in the demand for high-skilled ones—the opposite prediction from the neoclassical framework. The intuition is that under monopsony, a higher relative wage for low-skilled workers due to a wage floor has both labor supply and labor demand effects, which influence relative employment in opposite directions (Datta and Machin, 2024). also estimate the technology parameter after accounting for firms' wage-setting behavior and find that, in their setup, capital and labor are gross complements.

5 Inequality, distributional implications, and downstream effects

We have seen a considerable increase in wage inequality since the early 1980s in many countries, including the United States (Song et al., 2018), United Kingdom (Giupponi and Machin, 2022a), and Germany (Dustmann et al., 2009; Card et al., 2013). In some cases—especially the United States—this occurred concurrently with a fall in the real federal minimum wage (DiNardo et al., 1996; Card and John, 2002).⁴² More recently, we have also seen a reduction in wage inequality in the U.S., concurrently with a rise in state-level minimum wages in parts of the country (Autor et al., 2023). Wage inequality also declined in Germany and the U.K. over the past decade, at the same time minimum wages were introduced or expanded in both countries (Bossler and Schank, 2023; Machin, 2024). These descriptive trends raise the question: have minimum wage policies had a substantial impact on wage inequality?

At a *qualitative* level, the idea that minimum wages reduce wage inequality is uncontroversial. After all, conditional on having a job, a higher minimum wage raises the bottom wage, which will naturally reduce pay dispersion. However, if that were the only effect of the minimum wage policy, its *quantitative* impact on wage inequality would likely be quite modest in most cases, since the share of workers earning exactly the minimum wage is relatively small in most countries. For instance, 7.3 % of the U.S. workforce earned at or

⁴² To understand overall trends in inequality, it is important to distinguish between lower- and upper-tail inequality (Autor et al., 2008). Minimum wage is important to explain lower-tail inequality patterns, but the upper-tail changes are likely to be driven by other forces (e.g. skill bias technological change).

below the federal minimum wage in 2010, just after the last federal increase. This imposes bounds on any direct effect. However, it is also possible that a higher minimum wage leads to a higher wage for those already earning above the new minimum. Such spillovers (or ripple effects) up the pay distribution could considerably increase the scope for minimum wages to affect the wage distribution and inequality. Understanding the size of these spillovers is critical for discerning the policy's importance in determining inequality.

Theoretically, several factors can explain the presence of spillovers. One common idea is that the minimum wage shrinks the wage gap between front-line and higher-paid workers (such as supervisors), which can give rise to concerns related to fairness or incentives. As a result, companies have to adjust the pay of supervisors and others to maintain a wage hierarchy. Existing evidence suggests that such social comparisons matter at the workplace (e.g., (Dube et al., 2019); Gopalan et al., 2021). In addition to within-firm considerations, there may also be market-based mechanisms for spillovers. In an imperfectly competitive labor market, a higher minimum wage can raise the outside options for workers earning slightly above the minimum (see e.g (Flinn, 2010)). If workers at better-paying employers were to lose their jobs and be forced to take a lower-paying position, a higher minimum wage could raise the value workers assign to being unemployed (assuming the probability of finding a job is not too severely harmed). This, in turn, can force better-paying employers to adjust their wages upwards, causing a ripple effect higher up the distribution (see (Flinn, 2006); Butcher et al., 2012; Engbom and Christian Moser, 2022). Spillovers could also emerge from substitution with somewhat higher skilled workers (Teulings, 2000). Finally, there are more specific factors that can amplify wage spillovers, such as compensating differentials (Phelan, 2019), or employers' tendency to set pay at round numbers (Dube et al., 2020).

Empirical evaluations have broadly found evidence consistent with minimum wage spillovers, although the estimated magnitudes sometimes differ across studies (DiNardo et al., 1996). provided an early and useful illustration of how much minimum wages could impact the wage distribution using a novel decomposition method. The modern quasi-experimental literature began with (Lee, 1999), who used variation in the bite of (both federal and state) minimum wage across states in the U.S. between 1979–1988 to quantify the effect of minimum wages on wage inequality. He found sizable ripple or spillover effects, where the rise in the minimum raised pay for the bottom 40 % of wage earners. Strikingly, his findings suggested that nearly all of the growth in inequality at the bottom half of the wage distribution during the 1980s could be explained by the erosion of the federal minimum wage through inflation, similar to decomposition-based results of (DiNardo et al., 1996).

However, there were some puzzling aspects of (Lee, 1999)'s findings, which seemed to suggest that a higher minimum wage not only raised pay at the bottom compared to the median, but also at the top. In his specification, the

median wage appeared in both the dependent variable (log wage at p -th percentile minus log of median wage) and the independent variable (log of minimum wage minus the log of median wage). This raised a concern of division bias, which would tend to yield a spurious positive relationship between the two variables (Lee himself was clear about this possibility).

Subsequent work by (Autor et al., 2016) provides evidence consistent with such division bias. Their evidence, largely based on state-level minimum wage increases over a more recent period, uses an instrumental variables strategy to correct for the division bias. Their IV estimates suggest that a 10 % increase in the minimum raises the 10th percentile wage by around 1.5 %. Additionally, they find some spillover effects extending up to around the 20th or 25th percentile, beyond which the wage effects are close to zero. They also find that minimum wages played an important role in determining the 50/10—which is a measure of wage inequality in the bottom half of the distribution. But compared to (Lee, 1999), the spillovers found in the (Autor et al., 2016) study were smaller, leading to a smaller impact on inequality. Still, they found that maintaining the minimum wage at the 1979 level in real terms would have prevented somewhere between half and three-quarters of the overall increase in the bottom-half wage inequality, depending on the period in question. Moreover, the minimum wage had a larger effect on inequality for female workers, who tend to be lower paid.

More recent work by (Fortin et al., 2021) offers a complementary perspective that focuses on possible heterogeneity in the spillover effects. They point out that the U.S. federal minimum wage in early 1980s was much more binding than state minimum wages studied by (Autor et al., 2016). If the size of the spillovers depends on how binding the minimum wage is, this could offer another explanation behind the relatively larger evidence found in (Lee, 1999; Fortin et al., 2021). develop a hybrid approach that uses both state-level policy variation and the bite of the federal minimum wage, as well as the exact location in the state-level wage distribution where the minimum wage is binding.⁴³ This allows them to better evaluate the changes in the 1980s when the federal minimum wage eroded greatly, but there were only a handful of state-level changes.⁴⁴ Their design suggests that the spillover effects during the 1980s (studied by (Lee, 1999)) were indeed larger than those from subsequent state policies studied by (Autor et al., 2016). Their findings highlight that there may not be “one true estimate” of minimum wage effects, as these can vary based on factors such as the bite of the policy.

⁴³ Their approach shares similarities with (Cengiz et al., 2019) and (Giupponi et al., 2024), though unlike those papers, they are focused on the probability and not frequency distribution of wages—hence abstracting from employment impacts.

⁴⁴ One caveat is that the 1980s constitute a difficult period to study due to confounding shocks we discussed before, as shown, for example, in Fig. 3.

One limitation of much of the evidence on wage spillovers is the inability to distinguish true spillovers from job loss. When the minimum wage rises from \$13 to \$15, if all the jobs under \$15 are destroyed, simply comparing the pre- and post-intervention wage quantiles would show a rise in wages at quantiles initially above \$15. This would happen even if none of the jobs initially paying above \$15 raised their wages. Why? Because the disemployment truncates the wage distribution at \$15, changing the composition of jobs at various quantiles. Most studies in the literature acknowledge this possibility but assume that the disemployment effect is close to zero, thereby limiting any bias ([Cengiz et al., 2019](#)). show that by estimating the impact of the policy on the *frequency* distribution of wages, one can jointly estimate the employment and wage effect—thereby accounting for potential disemployment effects when measuring wage spillovers. They find economically meaningful, but limited, spillover effects from state minimum wages that constitute around 40 % of the overall wage increases from the policy. Additionally, ([Cengiz et al., 2019](#)) show that the measured wage spillovers are not likely driven by measurement error in survey wages, a possibility raised by ([Autor et al., 2016](#)). Evidence of spillovers from payroll data used by ([Gopalan et al., 2021](#)) further supports these conclusions.

Similarly, evidence from the U.K. by ([Giupponi et al., 2024](#)) based on the frequency distribution approach suggests a moderate-sized spillover effect from the National Living Wage policy. Research from Germany by ([Bossler and Schank, 2023](#)) also indicates that the introduction of the national minimum wage in Germany substantially reduced wage inequality, estimating that about half of the recent reduction in wage inequality can be attributed to the minimum wage.⁴⁵

Overall, the body of evidence consistently suggests that higher minimum wages tend to reduce wage inequality in the bottom half of the wage distribution. However, the magnitudes likely vary by the nature of the policy environment.

Distributional-implications. While the impact of minimum wages on *wage inequality* is relatively straightforward, the impact on *income inequality* and poverty is more complicated. First, most poor families have weak ties to the labor market, which limits the scope for minimum wages to raise bottom incomes. The minimum wage could also, of course, affect the probability of work—possibly heterogeneously by latent family incomes. Second, the lowest-wage workers may not always be in the lowest-income families. Finally, a rise in the minimum wage can have complicated interactions with other safety net and tax/transfer policies. As a consequence, a rise in minimum wage can be partly offset by reduced public transfers, thereby reducing the rise in post-tax-and-transfer income at the bottom.

⁴⁵ We review the evidence on spillovers from developing countries in [Section 6](#).

The evidence on poverty reduction is mixed. Some U.S. research have found negative effects on state-level poverty rates ([Addison and Blackburn, 1999](#); [Dube, 2019b](#); [Godoey and Michael Reich, 2021](#)). Other studies have not found evidence of overall poverty reduction ([Burkhauser et al., 2023](#); [Sabia and Robert, 2015](#)). Interestingly, ([Godoey and Michael Reich, 2021](#)) finds that while the average minimum wage increase did not reduce overall poverty rates, more binding increases did. There is also evidence that a higher minimum wage reduces public transfers, suggesting a rise in pre-tax household incomes at the bottom ([\(Dube, 2019b; Vergara, 2023; Reich and R.West, 2015\)](#)).

However, poverty is likely too limited a measure for evaluating distributional impacts. We need to better understand how the minimum wage affects both the pre- and post-tax/transfer income distribution, and their components. Work by ([Neumark et al., 2005](#)) represents an early approach along these lines, studying a short panel of individuals. More recently, ([Dube, 2019b](#)) provides estimates of the policy impact on unconditional income quantiles is an effort in this direction using the TWFE approach and re-centered influence functions. However, more work is needed on this topic, especially using event-based analysis, with attention to possibly heterogeneous impact by how binding the minimum wages are.

Moreover, to the extent that minimum wage workers are both relatively few and somewhat dispersed across the distribution (though more concentrated at the bottom), it is difficult to statistically detect the welfare effects of the policy by studying the entire income distribution. Using a predictive approach to identify families likely to be affected and studying the impact on their incomes could be a fruitful avenue going forward.

Finally, there is even less evidence from outside the U.S., making additional research crucial. One notable exception is ([Giupponi et al., 2024](#)), who study the distributional effects of minimum wage policies in the British context. They use a micro-simulation model of the U.K. tax-transfer system and incorporate their estimated employment and wage effects (including spillovers). Assuming there is no substantial heterogeneity in the response to the minimum wage across household income deciles—which aligns with the heterogeneity results discussed in Section ??—this approach provides an accurate picture of the policy’s distributional effects. While this method requires stronger assumptions than standard reduced-form analysis of household income, it offers valuable insights into the interaction between the tax and transfer system and minimum wages. Additionally, their approach can provide more accurate estimates by filtering out fluctuations in household income driven by the tax and benefit system.

Price effects and welfare. So far, we have discussed the effect of minimum wages on family income distribution. To fully understand the impact of minimum wages on family welfare, we also need to consider consumption responses. As discussed in [Section 4.1.7](#), an important margin of adjustment to the minimum wage is increased output prices, which could affect consumer welfare and should be taken into account.

The first important question is whether the price increase reflects a change in product or service quality. Ideally, we would focus on quality-adjusted price changes, but such data are rarely available. As discussed in [Section 4.1.7](#), our knowledge of quality changes following minimum wage hikes is limited. However, the few studies on this topic suggest some improvements in quality, although these improvements may not fully compensate consumers ([Brown and Chris, 2023](#)).

The second relevant question is the effect of (quality-adjusted) price increases on consumers' budgets. Since minimum wage workers make up a small share of the economy, the increased burden on consumers is relatively modest ([MaCurdy, 2015](#)). estimates that a 20 % increase in the minimum wage raises consumer expenditures by 0.5 %, a figure similar to what ([Harasztosi and Lindner, 2019](#)) estimate for Hungary. It's important to note that these small effects are likely to be an upper bound, given that these calculations typically assume no change in product quality.

The third important question is whether exposure to price shocks differs across the household income distribution. Studying overall consumption responses, both ([MaCurdy, 2015](#)) in the U.S. and ([Harasztosi and Lindner, 2019](#)) in Hungary find little evidence of differential exposure to price shocks across household income quantiles, suggesting that the burden of the minimum wage is shared relatively equally across the income distribution. On the other hand, ([Renkin et al., 2022](#)) use detailed data on grocery store expenditures to analyze how different income groups are affected by economic shocks. Their findings suggest that low-income households experience a disproportionately larger relative impact compared to higher-income households. However, it is unclear whether these findings are specific to the grocery store sector or reflect the more detailed spending data used in their analysis compared to ([MaCurdy, 2015](#)) and ([Harasztosi and Lindner, 2019](#)).

Several studies examine the impact of minimum wage increases on household consumption ([Aaronson et al., 2012](#)). find that households with minimum wage workers tend to increase their consumption more than their earnings in the U.S. context. This response is largely driven by a small number of households making significant purchases, such as vehicles. To finance this increased consumption, these households often turn to borrowing. The increased borrowing aligns with findings from other studies, which suggest that minimum wage hikes lead to greater availability of unsecured credit, a reduction in payday loan usage, decreased delinquency and bankruptcy rates, and improved credit scores ([Dettling and Hsu, 2020; Legal and Eric, 2024](#)). Overall, these findings highlight that the financial health of low-income households improves following minimum wage changes, at least in the U.S. context.

Furthermore, ([Dautović et al., 2024](#)) find similar positive effects in China, where minimum wage increases boosted consumption, particularly in households with children. Similarly, ([Mansoor and O'Neill, 2021](#)) finds that higher

minimum wages tend to raise household consumption in India, especially in areas with high compliance with minimum wage policies. These studies further corroborate the positive impact of minimum wage policies on consumption, underscoring their broader benefits.

Indirect effects of minimum wages. Minimum wages can influence many other socioeconomic outcomes through their direct impact on wages, household income, and consumption. These downstream implications of minimum wage policies are particularly context-dependent. For example, the effect of a minimum wage-induced positive income shock on health outcomes will depend on the structure of the healthcare system and the broader welfare system. As a result, it is unlikely that the downstream effects of minimum wage policies are uniform across different societies and social classes.

With this caveat in mind, it is worth briefly discussing the findings on some key downstream effects of minimum wage policies (Leigh et al., 2019). review evidence on health outcomes as of 2018 and find a beneficial impact of minimum wages on smoking prevalence, but no consistent evidence on other health outcomes (Neumark, 2024). reviews 57 papers on the impact of minimum wages on various health outcomes and 6 papers on the impact on crime. He provides a detailed description of each paper along with a subjective assessment of the credibility of the estimates. The findings suggest that minimum wages positively influence infant and child health, diet and obesity, depression and mental health, suicide rates, and family structure, while the evidence on teen and adult health is more mixed.⁴⁶ The evidence on crime is also mixed, with three studies showing a reduction in crime, one finding no effect, and two reporting an increase in nonviolent crime.

The effect of minimum wage policies on educational outcomes and human capital accumulation could have significant welfare implications. The evidence on high school completion is mixed (Neumark and Wascher, 1995a, 1995b, 2003). report that higher minimum wages depress school enrollment, while (Campolieti et al., 2005) and (Smith, 2021) find no significant changes in dropout rates. Similarly, (Chaplin et al., 2003) find no significant reduction in high school enrollment, particularly for students subject to compulsory schooling laws. This latter finding highlights that the impact of minimum wages on schooling is highly dependent on the education system.

A number of recent studies have examined the impact on post-secondary schooling decisions (Lee, 2020). finds a substantial reduction in enrollment rates at community colleges in response to minimum wage increases. Despite sizable declines in enrollment at two-year colleges, (Schanzenbach et al., 2024) find no reduction in degree completions. This suggests that the social costs associated with lower enrollment could be small (or even socially beneficial),

⁴⁶The health implications of minimum wages seem to align with those of other policies that increase the net income of low-wage workers (see (Godoy, Ken Jacobs, 2021)).

as workers discouraged from enrolling would not have completed their degrees anyway. In the Canadian context, (Alessandrini and Milla, 2024) find a somewhat different pattern: an increase in enrollment at community colleges but lower enrollment at universities. This discrepancy with the U.S. evidence could be explained by the fact that community colleges in Canada focus almost exclusively on vocational education, while most two-year institutions in the U.S. also provide academic training. Again, this points to the highly context-dependent aspect of the minimum wage impact on educational outcomes.

Overall, a large body of evidence produced over the past two decades suggests that the impact of minimum wages extends beyond merely raising the living standards of low-wage workers. At the same time, the indirect effects of minimum wage policies are strongly influenced by the broader institutional structure. Gaining a better understanding of these interactions is an important direction for future research.

6 Minimum wages in developing countries

While much of the research on minimum wage policies has focused on the United States and Europe, minimum wages in developing countries have gained increased attention in recent years. There are several issues specific to the developing country context, which we discuss in this section.

Developing countries often face higher levels of informality in their labor markets, complicating the enforcement and effects of minimum wage laws. Raising the minimum wage could incentivize employers to shift work to the informal sector, potentially harming workers by placing them in jobs that often lack health insurance, social security benefits, or job protection. However, the lack of enforcement in the informal sector does not necessarily mean that firms can simply ignore minimum wages in that sector. As discussed in [Section 4.1.1](#), even in developed economies, enforcement is unlikely to be the primary factor behind compliance (see ([Stansbury, 2024](#))). Therefore, the influence of minimum wages on the informal sector is an empirical question, which we explore in the following discussion.

The second major difference between developed and developing countries is the sectoral composition of minimum wage workers. In most developed economies, minimum wage workers are typically employed in the local service sector, whereas in many developing economies, they are often found in the tradable sector, exporting to the world market. As discussed in [Section 3.3.3](#), the effect of the minimum wage can vary considerably across industries, with more negative employment responses observed in the tradable sector.

Another important consideration is the presence of more significant imperfections and inefficiencies in the developing country context ([Hsieh and Peter, 2009; Bloom et al., 2014](#)). The empirical evidence presented in [Section 4.1.11](#) suggests that firms in developed countries often find ways to improve productivity in response to minimum wage increases. We suspect

that the scope for efficiency improvements is even more substantial in developing economies, making this margin of adjustment potentially more important.

These considerations highlight that it is *a priori* unclear what to expect regarding the impact of minimum wage policies in developing countries. The empirical evidence on the overall employment effects of minimum wages in developing economies is mixed. Two recent reviews have examined the employment effects in low- and middle-income countries (Neumark and Munguía Corella, 2021). review 61 papers on the impact of minimum wage policies in developing contexts and find mixed evidence on employment effects. They observe more consistently negative employment effects in formal sectors, particularly where the minimum wage is more binding and strictly enforced. On the other hand, (Broecke et al., 2017) review the impact of minimum wages on employment in 14 major emerging economies and find that minimum wages have minimal impact on employment, with evidence of reporting bias toward statistically significant negative results. More vulnerable groups (e.g., youth and low-skilled workers) are found to be slightly more negatively affected, and there is some evidence that higher minimum wages lead to increased informal employment.

However, it is important to note that both reviews focus on the employment elasticity with respect to the minimum wage rather than the own-wage elasticity. This distinction makes it difficult to compare estimates across groups, studies, or with the extensive evidence from developed economies that we reviewed in Section 3.3.

Less attention has been given to studying various other margins of adjustment in developing countries. However, a few papers, which we reviewed in Section 4, present findings consistent with those observed in developed countries. For instance, (Lemos, 2006) finds a positive effect on output prices in Brazil, while (Mayneris et al., 2018) and (Huang et al., 2014) find evidence of significant productivity improvements in China.

The evidence from developing countries also suggests that minimum wage increases can substantially alleviate inequality (Bosch and Manacorda, 2010). find that the erosion of minimum wages in Mexico contributed to a significant increase in wage inequality between the late 1980s and early 2000s, particularly at the bottom of the distribution. In India, research exploiting province-level variation finds that minimum wages have contributed substantially to a decline in wage inequality since the beginning of the century (Khurana et al., 2023a). In Brazil, minimum wage increases compressed wages in both the formal and informal sectors (Lemos et al., 2004; Derenoncourt et al., 2021). Notably, the estimated spillover (ripple) effects of the Brazilian minimum wage policy appear to be significantly larger, amplifying the contribution of minimum wages to reducing inequality (Engbom and Christian Moser, 2022). estimate that minimum wages affected pay up to the 90th percentile. However, (Haanwinckel, 2023), after controlling for other supply and demand shocks,

finds significant but more muted effects, extending up to the second decile from the bottom.

Interestingly, minimum wages also affect inequality in the informal sector, even though informal workers are not legally entitled to the minimum wage (see (Lemos et al., 2004); Derenoncourt et al., 2021). This phenomenon, sometimes referred to as the “lighthouse” effect, underscores the broader influence of minimum wage policies beyond their immediate legal scope. While understanding why minimum wages influence wages in sectors with limited enforcement requires further research, these findings suggest that firms cannot easily circumvent minimum wage regulations by shifting to the informal sector.

To summarize, the impact of minimum wage policies in developing countries is influenced by various factors, such as a high degree of informality, the sectoral composition of low-wage workers, and more pronounced market imperfections. At the same time, while the empirical literature is still growing, the evidence gathered so far from developing countries does not suggest that the core impacts differ considerably from those observed in developed economies.

7 Conclusion and future directions

Minimum wage policies have been the subject of extensive research for over a century, and there remains considerable interest in understanding their impact. Given the policy’s fundamental relevance for testing basic economic theories, it has been at the forefront of economic thinking, reflecting (and influencing) the evolution of both theories and empirical methods. At the same time, during this period the debate has often centered on the question of whether we should have a minimum wage at all.

We reviewed the most important evidence on how minimum wages shape the labor market and other downstream outcomes, focusing on contributions made since the beginning of the 21st century. During this period, a substantial body of empirical evidence has accumulated regarding the employment effects of the policy and its broader economic and social impacts. While the evidence is not unanimous, a reasonable conclusion from the existing literature is that minimum wage policies have had limited direct employment effects while significantly increasing the earnings of low-wage workers—at least at certain levels and in particular economic contexts. We see this as suggesting that minimum wages can be beneficial in many situations and should be considered a key economic tool for intervening in low-wage labor markets and improving economic outcomes.

However, the fact that minimum wages are effective at certain levels and in certain contexts does not imply they always work. Research in the 21st century should focus on determining the appropriate levels of the minimum wage, rather than debating the existence of the policy itself. Nonetheless, answering

this question presents a number of challenges that researchers will need to overcome to make progress.

First, we need a better understanding of the heterogeneous impact of the policy across different economic contexts. Identifying turning points where the minimum wage begins to significantly affect employment dynamics is an essential next step. A common theme throughout our chapter is the difficulty in characterizing heterogeneous effects. Estimating the average causal response to minimum wages is challenging enough, and sometimes requires strong assumptions for identification. Obtaining reliable estimates by subgroups (such as different types of workers or policy levels) typically require even stronger assumptions and may lack statistical power. Additionally, due to publication bias towards statistically significant results, the search for heterogeneous treatment effects may be more prone to false discovery. Therefore, research on heterogeneity requires careful and transparent analyses, and a variety of econometric and data challenges need to be addressed.

Second, a particularly important recent advance in the minimum wage literature is the use of a more transparent event study design rather than the relatively opaque TWFE panel regression. Future work should rely on event study designs to provide more precise estimates for each major minimum wage event. Accumulating knowledge about these events will be crucial for properly understand the nature of heterogeneity in minimum wage effects. The literature can also advance by providing more details about the treated and potential control units and how they were selected.

Third, to determine the appropriate levels of the minimum wage, we need to improve our models of low-wage labor markets. The standard competitive model often fails to capture the complexity of real-world labor dynamics, suggesting a need for more nuanced economic frameworks. Building such models requires a better understanding of the various margins of adjustments, especially in areas with relatively thin literature, as identified in [Section 4](#). These economic models should account for consumers' demand responses and the productivity-enhancing effects of the policy, along with incorporating labor market imperfections. However, it is important to avoid past mistakes. Models should be designed to match the best available evidence, instead of primarily being used for extrapolation. At a fundamental level, core theories of the labor market should be falsifiable, and we should not dismiss empirical findings that are difficult to reconcile with existing models.

Lastly, to set the optimal level of the minimum wage, it is crucial to define what “appropriate” means. This is a complex question, as many of the welfare consequences of the policy may depend on its interactions with other policy instruments and the political feasibility of changing them. Future research should better incorporate the interaction of minimum wages with key taxes, benefits, and government policies into the analysis.

In conclusion, while the study of minimum wage policies has made significant strides, there remains ample room for further research to refine our

understanding of their economic impacts. By addressing the areas outlined and advancing methodological approaches, researchers can contribute to more informed policy decisions that balance wage fairness with economic efficiency.

Appendix A Additional results

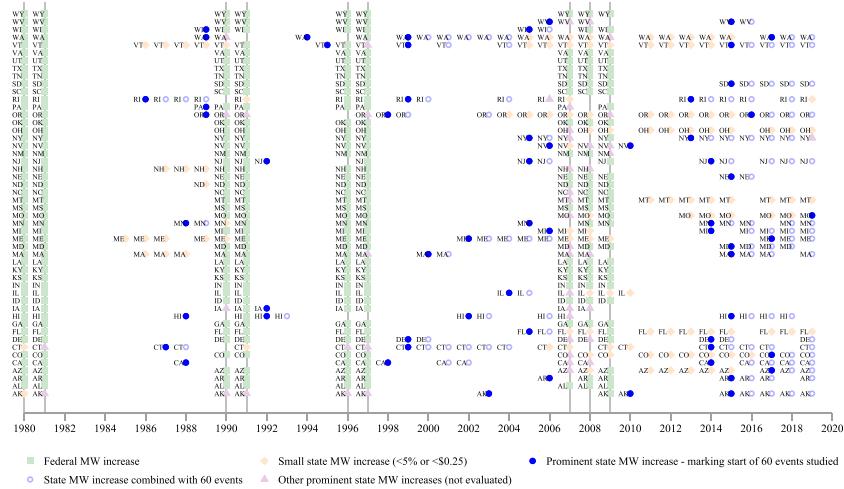


FIG. A1 All minimum wage increases from 1980–2019. Notes: This figure plots all instances of the minimum wage increasing in any state. The blue solid circles represent the start of 60 events that we study in our main analysis. The blue hollow circles are state MW increases that are evaluated as part of the post-periods of these 60 events. There can also be state MW increases that are neither the start of an event, nor are evaluated as part of another event. These are either very small increases (some of them due to indexation) - represented by orange diamonds; or they are prominent increases, represented by purple triangles. Most of the latter are in years when the federal minimum wage also increased - the only two exceptions are RI 2006 and NY 2019, which are both close enough to a prominent increase to not be counted as the start of a separate event, but more than 6 periods away from the last event-start year. The green squares show instances of increases due to a federal minimum wage increase. Any state that does not have a green square in a federal MW increase year either had its own increase in that year (small or prominent) which made its binding MW higher than the federal, or had a MW higher than the federal in the previous year as well, which made the federal increase redundant. More details are in Appendix Section 12.

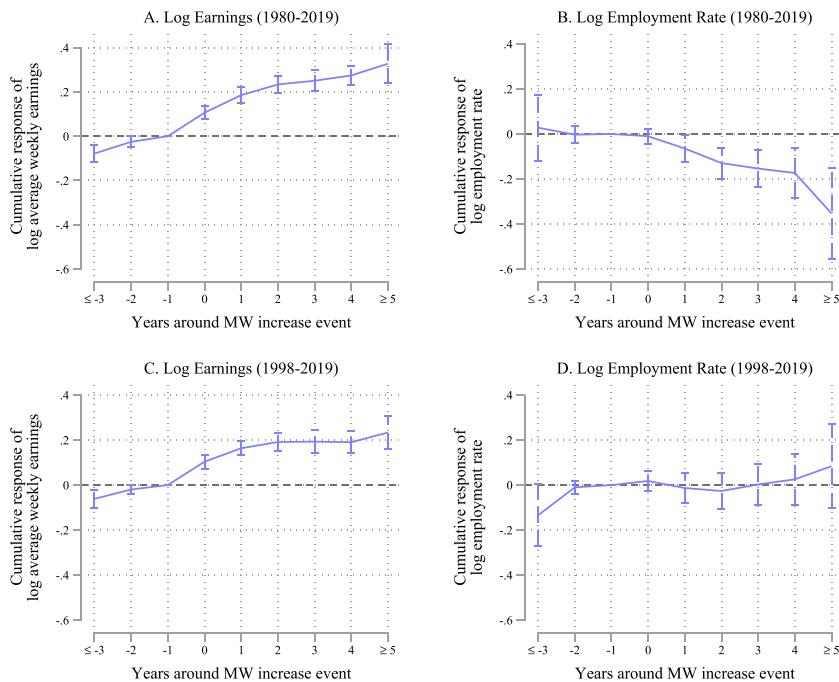


FIG. A2 Impact of a log-point increase in MW on log employment rate and log average earnings in the restaurant sector over time for TWFE-logMW specification with distributed lags. Notes: This figure plots the cumulative response of log employment rate and log average weekly earnings (2023 dollars) for restaurant workers to a log-point increase in the minimum wage (i.e., elasticities) using a distributed lag model with 2 leads and 5 lags of log minimum wage. The cumulative responses are normalized relative to event date (-1). 95 % confidence intervals are based on standard errors clustered by state, and all regressions use state population weights. Panels A and C show the impact on earnings in the full and post-1998 time periods. Panels B and D show impacts on employment.

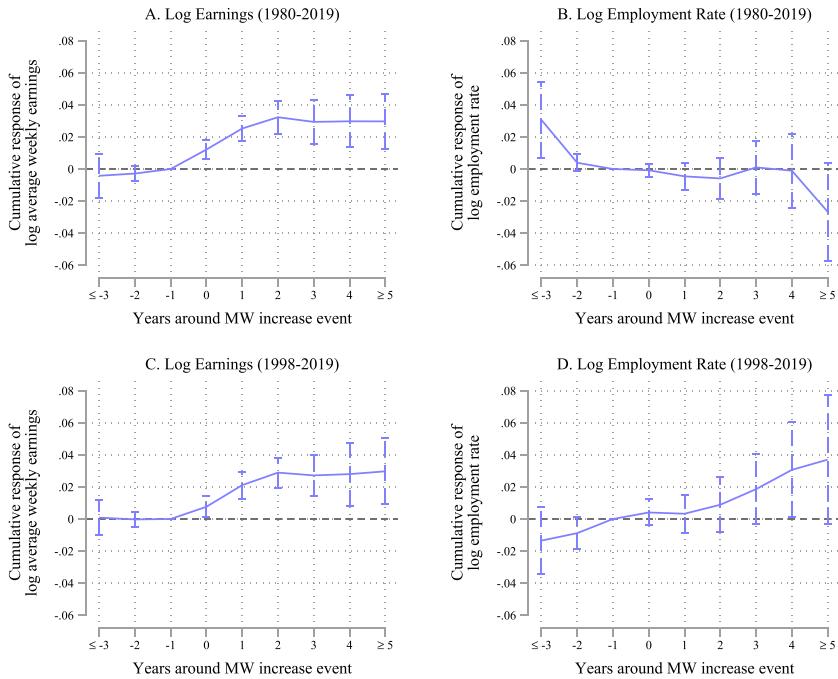


FIG. A3 Impact of a binary increase in MW on log employment rate and log average earnings in the restaurant sector over time for TWFE-binary specification with distributed lags. Notes: This figure plots the cumulative response of log employment rate and log average weekly earnings (2023 dollars) for restaurant workers to a binary increase in the minimum wage (i.e., semi-elasticities) using a distributed lag model with 2 leads and 5 lags of log minimum wage. The cumulative responses are normalized relative to event date (-1). 95 % confidence intervals are based on standard errors clustered by state, and all regressions use state population weights. Panels A and C show the impact on earnings in the full and post-1998 time periods. Panels B and D show impacts on employment.

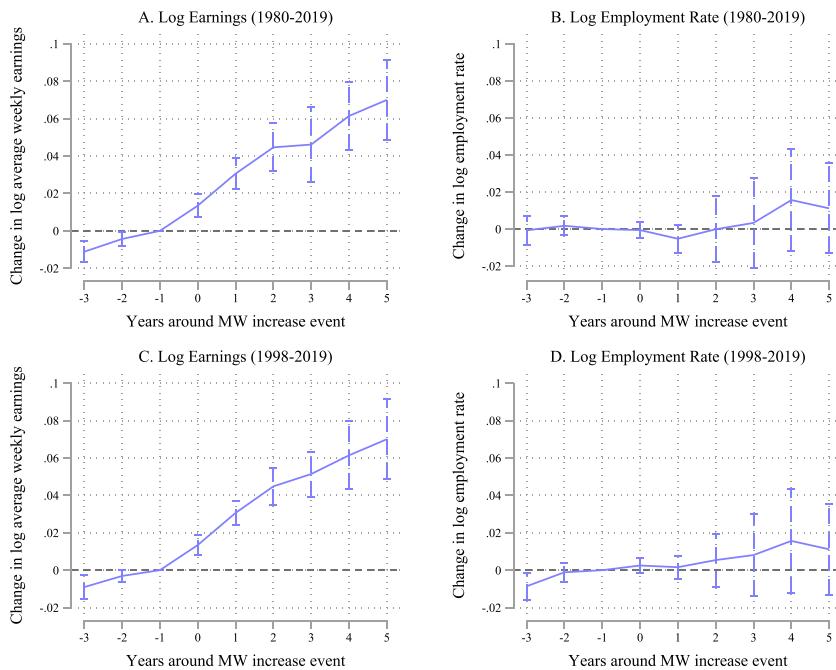


FIG. A4 Impact of an increase in MW on log employment rate and log average earnings in the restaurant sector over time for event-study specification. Notes: This figure shows the impact of a minimum wage increase on restaurant workers' log average weekly earnings in 2023 dollars (Panel A and Panel C) and log employment rate (Panel B and Panel D) in an event-study design. The top two panels are for the full period (1980–2019), while the bottom two only use data post-1998. The plotted points are coefficients from separate regressions using Eq. 7. Standard errors are clustered by state, and all regressions use state population weights.

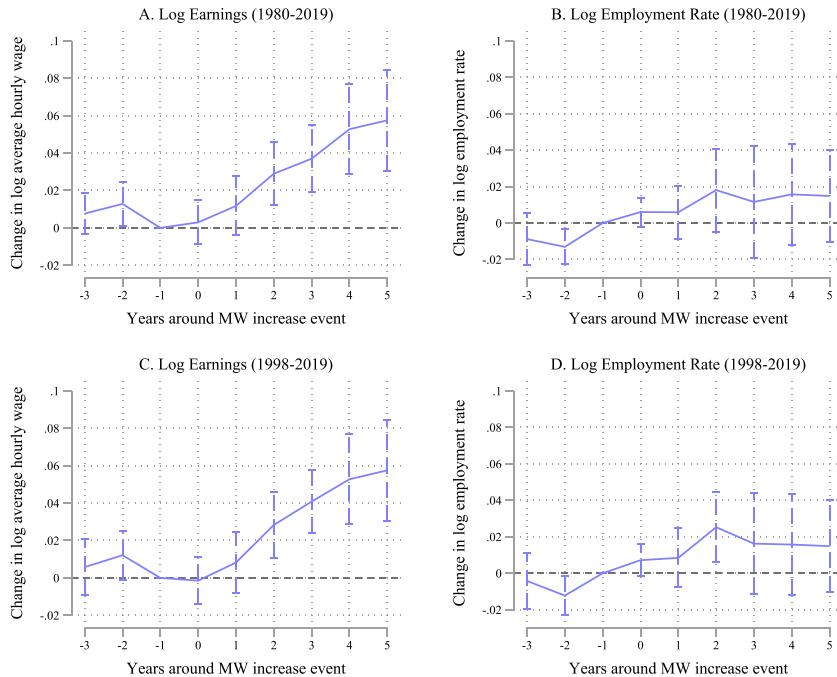


FIG. A5 Impact of an increase in MW on log employment rate and log average earnings in the high recall sample over time for event-study specification. Notes: This figure shows the impact of a minimum wage increase on the high recall sample's log average hourly wage in 2023 dollars (Panel A and Panel C) and log employment rate (Panel B and Panel D) in an event-study design. The top two panels are for the full period (1980–2019), while the bottom two only use data post-1998. The plotted points are coefficients from separate regressions using Eq. 7. Standard errors are clustered by state, and all regressions use state population weights.

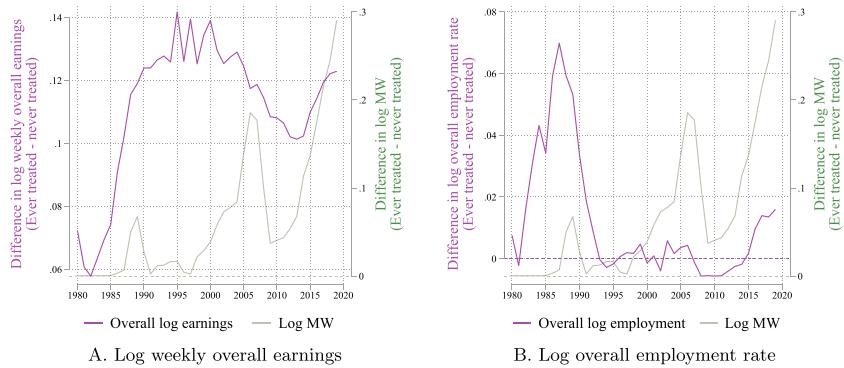


FIG. A6 Difference in minimum wage, overall employment rate and overall log average earnings between ever-treated and never-treated states. Notes: This figure plots the population-weighted difference in overall log earnings/log employment between 35 ever treated and 15 never treated states on one axis, and the difference in log MW between these states on the other. Log average weekly earnings (in 2023 dollars) are shown in Panel A, and log employment rate in Panel B. Ever treated states had at least one state MW increase, over and above any federal increases from 1980–2019. Never treated states had no such increase.

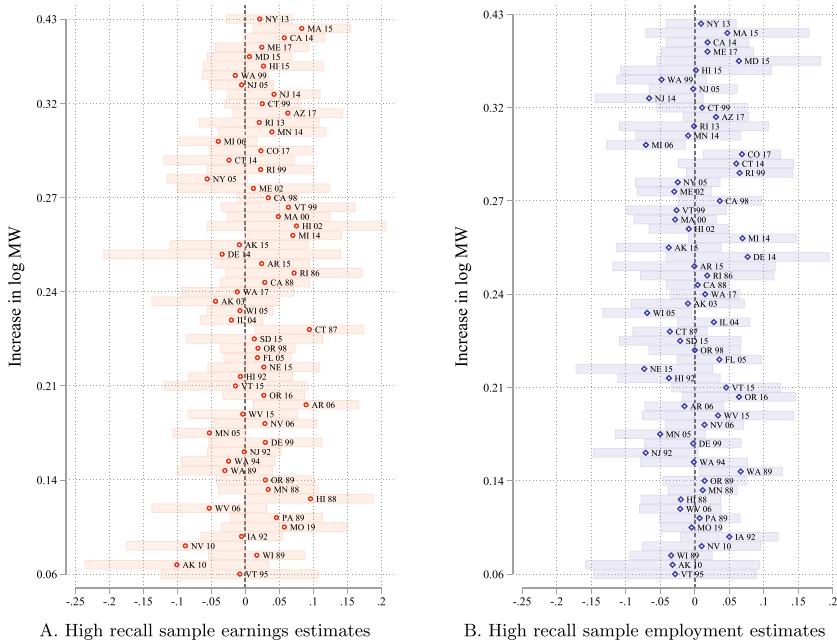


FIG. A7 Event-by-event employment and earnings estimates for high-recall sample. Notes: These figures plot the event-by-event earnings and employment estimates for the high recall sample. Thus, each point estimate is from an event-study regression (Eq. 5) with a particular combined event and its corresponding clean controls. The dependent variable is log average hourly wage (2023 dollars) for Panel A, and log employment rate for Panel B. All regressions use state population weights. The 95 % confidence intervals use Ferman-Pinto standard errors. We identify 60 events, where successive minimum wage increases are combined into a single event (see Appendix 12). The events are sorted by total increase in log minimum wage.

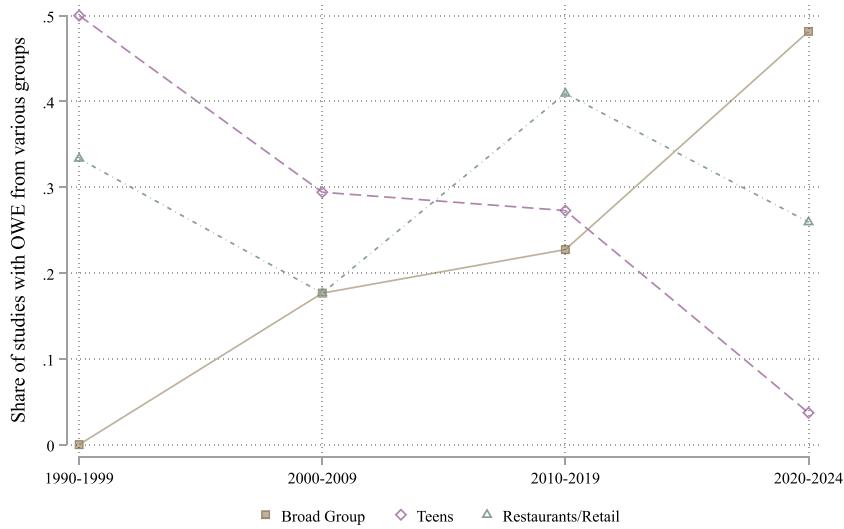


FIG. A8 Evolution of MW studies over time: the share of groups being studied over time. Notes: This figure plots the share of MW studies in the ([Dube and Zipperer, 2024](#)) database studying various groups over time. Starting from the 1990s, the share of studies on teens has steadily declined, while a much larger number of studies now focus on broader groups. Studies on restaurant/retail sectors have declined somewhat, but not as much as teens.

TABLE A1 Details for the 61 combined minimum wage events, by year of the initial increase.

	Number of events	Post period length (years)	No. of clean control states	Mean increase in log MW
1986	1	4	36	.14
1987	1	3	36	.19
1988	3	2	36	.159
1989	4	1	36	.116
1992	3	4	35	.15
1994	1	2	41	.142
1995	1	1	45	.057
1998	2	6	39	.171
1999	5	6	38	.196
2000	1	6	33	.232
2002	2	5	28	.188
2003	1	4	28	.235
2004	1	3	29	.177
2005	5	2	30	.201

2006	5	1	31	.195
2010	1	5	21	.067
2011	1	6	19	.089
2013	1	6	22	.192
2014	7	6	23	.196
2015	9	5	24	.198
2016	1	4	24	.125
2017	4	3	24	.236
2019	1	1	24	.091
Overall	61	3.8	29.8	.179

This table provides information by cohort on the 61 events used in the event-study analysis. The first column reports the number of events in that cohort/year. The second column gives the length of the post-period for that cohort in years. Post-periods are a maximum of 6 years unless interrupted by a federal MW increase year or end-of-sample (except for 2010 - this exception is explained in the text). The third column reports number of states that serve as clean controls for the cohort - a clean control needs to have no state MW increase in the three years before the cohort year, and also no state MW increase in the relevant post-period. Finally, the last column reports the average log MW increase (averaged log MW in the post-period - log MW in year $t-1$, where t is the event year) across all events in a cohort. The last row gives the total number of events and means by event (not by cohort) of post-period length, clean control states, and log MW increase.

TABLE A2 Effects of increased MW on restaurant employment - Quasi event study estimates.

	(1) Short pre-period (2010–2013)	(2) Long pre-period (1980–1987; 1990–2000; 2009–2013)
Log MW	0.141 ^b (0.027)	0.170 ^b (0.029)
Log wages	0.043 ^b (0.010)	0.071 ^b (0.013)
Log employment	0.006 (0.012)	-0.050 ^a (0.024)
OWE	0.134 (0.265)	-0.706 ^a (0.350)

Each cell represents a coefficient from a separate regression. The first three rows in the first column report estimates from Eq. 8, i.e., this is like an event study regression where the event start date is 2014 for all 35 ever-treated states, and the sample consists of observations from only 2010–2019. The pre-period is thus 2010–2013 and the post-period is 2014–2019. The second column extends the pre-period to also include all years before 2010 in which the difference in log MW between ever-treated and never-treated states is no greater than the difference in 2010–2013. This way, we end up including 1980–1987; 1990–2000; and 2009. Both columns show effects on log MW, log wages, and log employment; and the last column reports own-wage elasticity which is the ratio of the log employment and log earnings estimates. Standard errors in parentheses are clustered by state, and all regressions use state population weights.

^a $p < 0.05$

^b $p < 0.01$

TABLE A3 Heterogeneity analysis and robustness for impacts of increased MW.

	(1) All events	(2) Long events	(3) Long post events	(4) Top tercile events	(5) Equally weighted ATT	(6) Control for low prob emp	(7) TWFE binary
A. Restaurant sample (Data source: QCEW)							
Log wages	0.031 ^b (0.005)	0.051 ^b (0.007)	0.047 ^b (0.004)	0.040 ^b (0.005)	0.033 ^b (0.006)	0.032 ^b (0.006)	0.031 ^b (0.008)
Log employment	-0.002 (0.004)	0.006 (0.008)	0.001 (0.007)	0.003 (0.005)	-0.0002 (0.005)	-0.001 (0.005)	-0.049 ^b (0.018)
OWE	-0.079 (0.135)	0.111 (0.159)	0.031 (0.150)	0.071 (0.120)	-0.008 (0.154)	-0.039 (0.140)	-1.586 ^a (0.679)
B. High recall sample (Data source: CPS)							
Log wages	0.020 ^b (0.007)	0.036 ^b (0.009)	0.032 ^b (0.007)	0.022 ^a (0.010)	0.022 ^b (0.008)	0.021 ^b (0.007)	-0.004 (0.005)
Log employment	0.007 (0.007)	0.018 (0.011)	0.017 ^a (0.009)	0.004 (0.009)	0.007 (0.008)	0.003 (0.007)	0.003 (0.006)
OWE	0.319 (0.308)	0.505 (0.342)	0.547 (0.286)	0.191 (0.371)	0.303 (0.371)	0.154 (0.332)	-0.752 (1.969)

Notes: This table reports various robustness checks for the original event-study estimates reported in Column 1 run as per Eq. 5. Columns 2, 3, and 4 run the same regression, but with a subset of the original 61 events. Column 2 only keeps events that have at least a three year clean pre-period and a three year clean post-period. Column 3 has a weaker constraint, and keeps all events that have at least a three year post-period. Having a three year post-period means not having any federal increases until at least two years after the event, and having a three year pre-period means having no federal increases in the three years preceding the event. Column 4 restricts attention to the top tercile of events in terms of the total minimum wage increase caused by the combined event. Finally, Column 5 and 6 report an equally weighted ATT, using regression adjustment; Column 6 controls for low probability sample employment rate. Finally, Column 7 reports estimates from a TWFE regression with the cumulative number of MW increase events in a state as the independent variable (this is a variation on Eq. 3). All columns have effects on log wages, effects on log employment, and own-wage elasticity which is the ratio of the latter to the former, for both restaurant and high recall samples. For wages, the restaurant sample uses real weekly log earnings, while the high recall sample uses real hourly log wage. Standard errors in parentheses are clustered by state, and all regressions use state population weights.

^a $p < 0.05$.

^b $p < 0.01$.

Appendix B Bias from heterogeneous pre-existing trends: a simulation study

Recent literature on difference-in-differences (DiD) has highlighted how, under staggered treatment adoption, heterogeneous treatment effects combined with negative weighting can lead to spurious estimated dynamics when using Two-Way Fixed Effects (TWFE) models with distributed lags (see ([Sun and Abraham, 2021](#))). However, another key source of bias is the interaction between the negative weighting problem in TWFE models and pre-treatment trends. This interaction can lead to highly misleading estimates of dynamic responses that (1) obscure actual pre-existing trends and (2) falsely suggest treatment effects. These issues are further exacerbated by a related concern: unlike DiD models with defined event windows, the TWFE model requires a stronger assumption that potential outcomes follow parallel trends over the entire sample period ([Marcus and Sant'Anna, 2021; Roth et al., 2023](#)).

In this Appendix, we use a simple Monte Carlo simulation with two treatment cohorts and staggered adoption to illustrate these issues. We demonstrate that heterogeneous violations of the parallel trends assumption can cause TWFE distributed lag models to generate highly misleading inferences, even when the true treatment effects are constant (and zero). Importantly, in our example, the average violation of parallel trends is small and occurs well before the treatment events, outside the defined event window. Consequently, while the TWFE model's stronger parallel trends assumption results in biased estimates, modern difference-in-differences event study estimators, such as those proposed by ([Callaway and Sant'Anna, 2021; Cengiz et al., 2019](#)), and ([Dube et al., 2023](#)) provide unbiased estimates.

Simulation design.

The basic setup for the data generating process (DGP) reflects a staggered adoption process:

- Time runs from $t = 1$ to $t = 15$.
- There are 50 states, divided into three groups: 30 never-treated states, 10 early-adopter states (treated at time $t = 9$), and 10 late-adopter states (treated at time $t = 11$).
- The true treatment effect β is equal to zero.
- Errors ν_{st} at the s -by- t level are normally distributed with a standard deviation of 0.05.

The key feature of the DGP here is that early-adopter states have a **negative latent trend** until time $t = 6$. In contrast, late-adopter states have a **positive latent trend** per year until $t = 6$. There are no latent trends after $t = 6$ for either group. All trends are relative to the never-treated group. The opposite signs in the latent trends means that the overall violation of parallel trends is small. However, as we will see, in combination with the staggered adoption this can lead to a large bias in TWFE estimates.

These assumption can be expressed formally using a potential outcomes model. For state s at time t , there are state fixed effects γ_s , time fixed effects τ_t , and treatment D_{st} , which takes the value 1 if state s is treated at time t , and 0 otherwise. The treatment effect is denoted by β . The potential outcomes Y_{st}^0 are described as follows:

$$Y_{st}^0 = \gamma_s + \tau_t + \text{Latent trend}_{st} + \nu_{st}$$

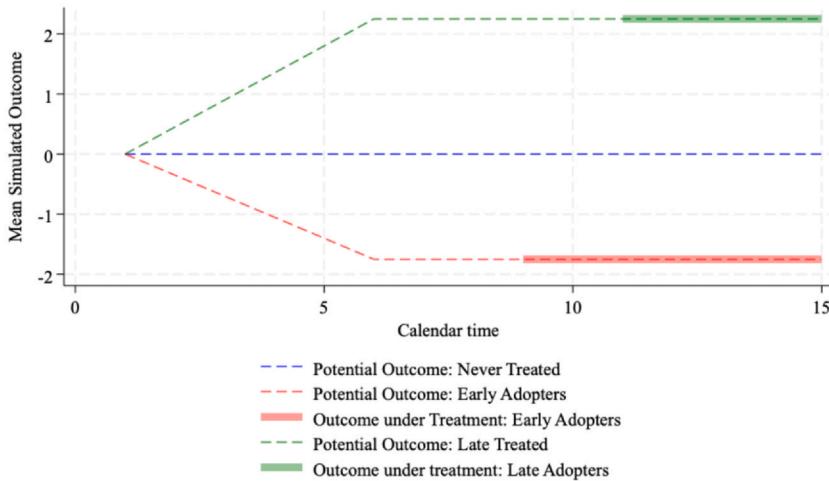


FIG. B1 Mean potential outcome, and outcome under treatment for the 3 groups.

The heterogeneous latent trends (i.e., violations of parallel trends) for the early and late-treated groups are specified as:

$$\text{Latent trend}_{st} = \begin{cases} -0.35 \times t & \text{if } s \in \text{early adopter and } t < 6 \\ -1.75 & \text{if } s \in \text{early adopter and } t \geq 6 \\ 0.45 \times t & \text{if } s \in \text{late adopter and } t < 6 \\ 2.25 & \text{if } s \in \text{late adopter and } t \geq 6 \\ 0 & \text{if } s \in \text{never treated} \end{cases}$$

The overall outcome under this data generating process can be expressed as:

$$Y_{st} = \gamma_s + \tau_t + \beta D_{st} + \text{Latent trend}_{st} + \nu_{st}$$

Time effects and state effects are assumed to be distributed normally with mean zero, and standard deviation of 0.05.

Fig. B1 plots the mean potential outcomes in dashed lines and outcomes under treatment in solid lines for these three groups of states. We can see the heterogeneous violations of parallel trends before $t = 5$, but afterward the potential outcomes follow parallel trends. In addition, the effect of the treatment is constant (in this case, zero), and so the potential and actual mean outcomes coincide in the units treated following treatment.

This setup illustrates a violation of the assumption of parallel trends before time $t = 5$, since the early and late adopting groups exhibit opposite-signed latent trends. However, after time $t = 5$, the assumption of parallel trends holds, as the change in expected potential outcomes is the same in all 3 groups.

Event study versus TWFE distributed lag estimates.

We begin by estimating the dynamic treatment effects using well-suited difference-in-differences (DiD) event study designs, specifically the (Callaway and Sant'Anna, 2021) and LP-DiD estimators. Recall that LP-DiD, when applied without re-weighting—as in this implementation—is equivalent to the stacked regression estimator proposed by (Cengiz et al., 2019). Recall that both treatment events occur after $t = 6$, at periods 9 and 11. This means

that if we focus on an event window covering 3 years before and 3 years after treatment, DiD methods such as Callaway-Sant'anna or LP-DiD would estimate a null effect, as the true causal effect is zero, and there is no bias due to violations of parallel trends within the event windows.

[Fig. B2](#) presents event study estimates from 250 simulations, focusing on an event window from periods -3 to 3 . The solid lines represent the mean estimates from the simulations, while the shaded areas show the $2 \times SD$ interval around the mean, where SD is the standard deviation of the estimates across simulations. Both the LP-DiD and Callaway-Sant'Anna estimators accurately identify the absence of pre-treatment trends, consistent with the true data-generating process. Additionally, both estimators correctly detect zero post-treatment effects, as expected, since the treatment has no impact beyond the event period.

Next, we apply a TWFE distributed lag model to the same scenario. This model includes 2 leads of treatment, the contemporaneous treatment, and 3 lags of treatment. We normalize all coefficients relative to event time -1 , meaning that estimates for event time -3 or earlier are combined (or “binned”) into a single coefficient, as are estimates for event time 3 or later. We assume no changes in treatment status before time 1 or after time 14 , allowing us to define leads and lags for all 750 observations. Cumulative responses are then calculated by summing the effects over event time.

The mean impulse responses from the TWFE distributed lag model also show no obvious pre-treatment trends, including in the binned -3 event time; this is not itself a problem, as the opposite signed early-period violations tend to mostly cancel each other out. However, the model suggests a small negative “short-run” effect, followed by a large negative “long-run” effect, despite the true effect being zero throughout. This spurious result arises because the negative weighting problem inherent in TWFE models interacts with a violation of parallel trends before time $t = 5$, leading to a highly misleadingly shaped impulse response function.

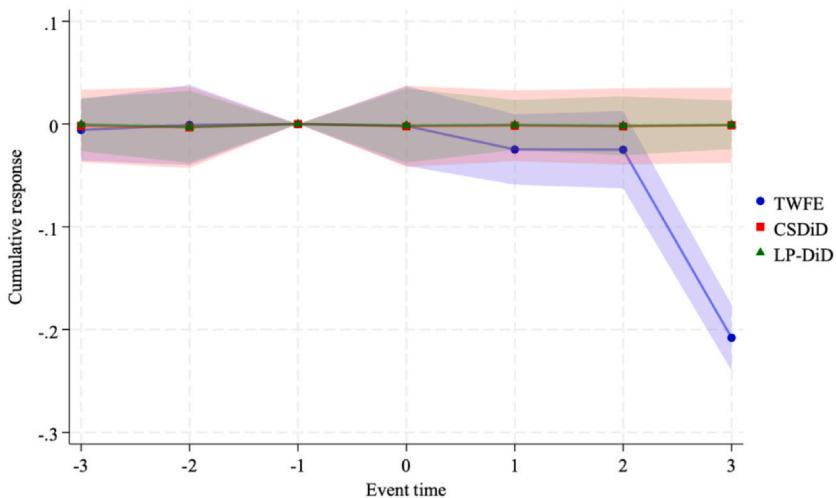


FIG. B2 TWFE-distributed lag versus Event Study estimates using Callaway-Sant'Anna, and Local Projections DiD (equivalently Stacked DiD). Notes: Results from $N = 100$ simulations. The shaded regions represent $2 \times SD$ intervals based on the estimated treatment effects.

The most concerning issue is that while the source of the bias lies in a parallel trends violation substantially prior to the event window, the erroneous inference manifests as a lagged treatment effect. Moreover, we would not be able to detect this problem by inspecting the leading coefficients. This is the mirror image of the problem identified by ([Sun and Abraham, 2021](#)), where heterogeneous treatment effects produce misleading pre-trends, even when the parallel trends assumption is satisfied in reality.

Overall, this simulation demonstrates how heterogeneous pre-trends can result in biased inferences when using TWFE distributed lag models in a highly non-transparent fashion. With staggered treatment timing, even if the parallel trends assumption holds within the event window and the treatment effect is homogeneous, the combination of negative weighting and violations of parallel trends outside the event window produces spurious impulse response estimates. This is a consequence of the stronger form of parallel trends assumption needed for consistency under the TWFE model: that it holds not only within the event window but across the whole sample period.

Fortunately, modern difference-in-differences estimators, such as Callaway-Sant'anna and LP-DiD, which utilize clean control groups and define a clear event window, are better at guarding against these biases and provide more reliable results.

Fortunately, modern difference-in-differences estimators, such as Callaway-Sant'anna and LP-DiD, which utilize clean control groups and define a clear event window, are better at guarding against these biases and provide more reliable results.

Appendix C Data sources for cross-country Kaitz indices

Brazil.

Minimum wage data can be accessed from the *Instituto de Pesquisa Econômica Aplicada* or IPEA. Average wage for 2002–2016 also comes from IPEA, from the survey *Pesquisa Mensal de Emprego* (PME). PME was a monthly employment survey of 6 metropolitan areas (Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo and Porto Alegre) that was discontinued in 2016 and was replaced with the PNAD (below) with covers the whole country.

For 2012–2023, we access average wage data from the household survey *Pesquisa Nacional por Amostra de Domicílios Contínua* (PNAD).

To harmonize these two sources, we created a wage variable that is equal to the average wage from PNAD for 2012+. Then, for 2002–2011, we used the metropolitan average wage from PME multiplied by the average ratio of the overall to metropolitan average wage for the overlapping years (2012–2016).

China.

Average wage and employment numbers for China come from the Chinese National Bureau of Statistics (NBS) interactive data explorer. Some earlier data is not available there, but was gathered manually from the China Statistical Yearbook - average wage data for 1995–1999 was sourced from the 2010 Statistical Yearbook.

NBS only reports average wage for workers employed in urban areas, so we use urban average wages. We use the following indicators from NBS: Average Wage of Employed Persons in Urban Units(yuan), Average Wage of Employed Persons in Urban Private Units (yuan), Urban Employed Persons(10000 persons), and Urban Employed Persons, Private Enterprises(10000 persons). Note that “Urban units” does not include private, so the indicator Average Wage of Employed Persons in Urban Units(yuan) is the average wages for workers in the non-private sector.

We calculate the proportion employed in each year and use that to take a weighted average of the average wage in the private and non-private sector to get the national annual average wage.

The China NBS stops reporting the number of urban employed persons in private enterprises in 2019. So, after 2019, we hold the proportion of urban employment in private enterprises constant (i.e. we use the 2019 value for 2020+).

Finally, NBS only reports private (urban) average wages starting in 2009. We calculate the private average wage prior to 2009 ($t < 2009$), by multiplying the average wage in year t by the ratio between the average private wage to average non-private wage in 2009.

Monthly nominal minimum wage data is from the ILO. We multiply the monthly minimum wage by 12 to get the annual minimum.

India.

We use regional data used in (Khurana et al., 2023b) and (Khurana and K. Mahajan, 2020) provided to us by the authors - the data is on average daily wages and regional population for the years 1983–84, 1993–94, 1999–00, 2004–05, 2007–08, 2009–10, 2011–12, 2017–18 and 2018–19. Average wage data is for age ranges 15–59. We also use daily minimum wage data for unskilled agricultural workers (the lowest minimum wage) for the years 1999–00, 2004–05, 2007–08, 2009–10, 2011–12, 2017–18 and 2018–19 provided to us by Khurana and Mahajan. We use minimum wage data from Menon and Rodgers for the years 1983 and 1993 available here.

We create a Kaitz for India by first calculating region-specific Kaitz and then using a population-weighted average of these for each year to get the national Kaitz.

The workers' wage data is based on surveys conducted from July to June, and minimum wage data (provided by Mahajan and Khurana) is from January to December. For surveys conducted from July to December in year t , the minimum wages for t are used, and for surveys from January to June in year $t + 1$, the minimum wages for $t + 1$ are used. Therefore, the $t-t+1$ wages data correspond to the average of the t and $t+1$ minimum wage data, that is, the 1999–2000 worker's average wages data is linked to the average of the 1999 and 2000 administrative minimum wage data. Wages for the years 1983–84, 1993–94 are paired with minimum wage data from Rodgers and Menon for the years 1983 and 1993, respectively. In the figure, we plot Kaitz based on the first year in which the survey was conducted, so, for example, the Kaitz associated with wages and population numbers for the 1993–94 household survey is plotted as the Kaitz in 1993.

Minimum wage data from Rodgers and Menon is based on industry and occupation groups. Because there is not always an “unskilled” agriculture minimum wage for each region in their data, we take the lowest minimum wage in the agriculture sector of each region as the minimum wage for the Kaitz index.

The data for the years 1999–00, 2004–05, 2007–08, 2009–10, 2011–12, 2017–18 and 2018–19 pertains to 18 major states of India according to the Census 2001. For the newly created states from the old states, administrative minimum wages are allocated district-wise from the year these states began reporting. For example, Telangana, which was carved out of Andhra Pradesh in 2014, uses the minimum wages of Telangana for the districts in Andhra Pradesh that were transferred to Telangana since the time Telangana started reporting the minimum wages.

For the years 1983–84, 1993–94, we merge wage and population data from Khurana and Mahajan with minimum wage data from Menon and Rodgers. The result of this is that in 1993–94, there are 16 states represented, but in 1983–84, we only have data to

construct Kaitz for 9 states. In order to correct for this, we calculate Kaitz indexes in 1999 based on the subset of overlapping states between 1999 and 1983 (or 1993). Then we then multiply the Kaitz indexes in 1983 (or 1993) by the ratio of the 1999 Kaitz using all states and the 1999 Kaitz using only overlapping states. For 1983, all states are in both datasets (9 states). In 1993, one state, Orissa, does not overlap and, thus, is dropped, so there are 15 overlapping observations.

OECD countries.

We use OECD data from France, Germany, Spain, U.K. and the US, on the Kaitz index, constructed as the ratio of the minimum wage to mean wages of full-time workers. We assume that the average full-time wage is around 10 % higher than the overall average wage. So we scale the OECD Kaitz index by 1.1 to construct an estimated Kaitz for all workers (not just full-time workers).

United States.

For the U.S., we supplement the OECD data by additionally constructing a state-minimum-wage based Kaitz measure. Specifically, we construct two measures as follows.

1. Kaitz with federal minimum wage: we do the same as for other OECD countries, and use the Kaitz index reported by OECD and multiply times 1.1. The OECD reports the Kaitz index starting in 1974. Pre-1973, we use the average hourly wage of production and non-supervisory workers in order to calculate Kaitz from 1964–1972.
2. Kaitz with state minimum wage: we use the Kaitz index from the OECD to construct a Kaitz index using state-level minimum wages starting in 1980. We use state-level minimum wage data over time from ([Vaghul and Zipperer, 2016](#)). We then construct an annual, weighted mean of the state minimum wage, where the weights are the counts of wage earners in each state-year cell. These counties are calculated using the CPS data (NBER extracts for 1980–1981, and IPUMS extracts for 1981–2022). We then multiply the OECD (federal) Kaitz by the average state MW/federal MW ratio to obtain the Kaitz using state minimums.

Appendix D Constructing historical QCEW restaurant data

Our results on the restaurant and overall sample come from Quarterly Census of Employment and Wages (QCEW) data. In this Appendix, we describe the process of cleaning and constructing this data set.

Harmonizing restaurant data from different classification systems (NAICS and SIC).

We downloaded the annual version of QCEW data for 1980–2019—however, not all of it is available under the same classification. Data from 1990 onwards is available under the North American Industry Classification System (NAICS), while data from 1980–2000 is available under the Standard Industrial Classification (SIC).

We use two groups of outcome variables: overall earnings and employment by state-year, and earnings and employment in the restaurant sector by state-year. There are separate SIC and NAICS based versions for both groups. For the overall outcomes, we use SIC-based variables before 1990, and NAICS-based variables from 1990 onward. As these are overall numbers, they should not differ across the two classification systems.⁴⁷

⁴⁷ In practice, there are some very minor differences, but the mean difference between employment level and total wages in NAICS and SIC data in 1990 (the first year for which both overlap) is less than 0.1 %.

However, there are differences in restaurant earnings and employment we get from both classification systems in overlap years (1990–2000). For NAICS, we take data for the three-digit code 722 (“Food services and drinking places”), while for SIC, we use code SIC_0G581 (“Eating and drinking places”). To create a harmonized series for restaurant earnings and employment for 1980–2019, we use the following procedure. We first calculate the ratio of NAICS employment (or earnings) to SIC employment (or earnings) in 1990—the first year when we have data for both classification systems. Next, we multiply the SIC-based outcomes (employment or earnings) during 1980–1989 by this ratio to obtain our new *aligned* series for these years. For 1990 and later years, we simply use the NAICS data.

Fig. D1 provides an illustration of this exercise. The top panel plots restaurant employment rate in “ever-treated” and “never-treated” states from both the NAICS and SIC datasets. We can see that, while close, the levels in the two datasets are not exactly equal in the overlapping years of 1990–2000. The bottom panel plots the aligned series where the post-1990 lines remain the same, but the pre-1990 lines have been aligned as described above to prevent any discontinuities.

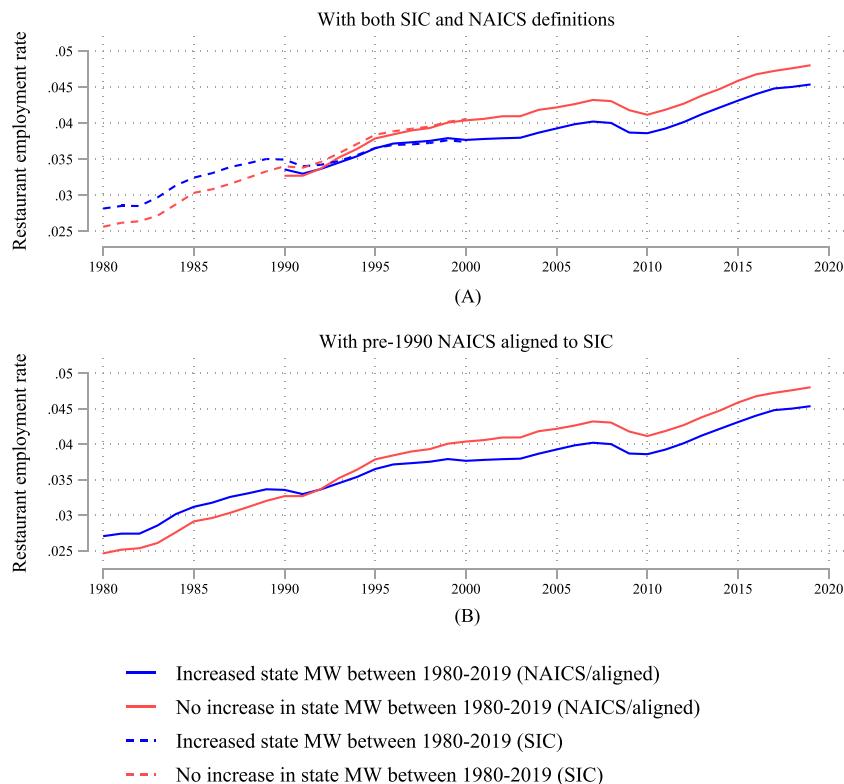


FIG. D1 Example of aligning NAICS definition with SIC definition. Notes: This figure gives an example of the NAICS/SIC alignment process. We have data as per the SIC definition from 1980–2000, and as per the NAICS definition from 1990 onwards. The top panel plots restaurant employment in ever treated and never treated states as per both definitions, subject to data availability. The bottom panel imputes data for 1980 as per the NAICS definition by comparing the NAICS/SIC ratio in 1990 (the first year where we have overlapping data). The text in this section explains this process in more detail.

Dealing with missing data for Alaska, Delaware, and Rhode Island in the 1990s.

Three states—Alaska, Delaware, and Rhode Island—have missing or incomplete data for the restaurant sector (NAICS code 722) in the 1990s. For these states, we impute data for the 3-digit restaurant sector using available data from 4-digit sub-sectors. In the NAICS classification system, there are five 4-digit sub-sectors within the broadly defined restaurant sector (the description for NAICS code 722 is “Food services and drinking places”). These 5 sub-sectors are 7221 (Full-service restaurants), 7222 (Limited-service eating places), 7223 (Special food services), 7224 (Drinking places), and 7225 (Restaurants and other drinking places). Out of these, data on 7221 and 7222 is consistently available through the 1990s for all three of Alaska, Delaware, and Rhode Island. The other three sub-sectors are not available in one or more of the three. Thus, we use only 7221 and 7222 for our imputation. Clearly, just adding these two sub-sectors is not enough as they leave out some components of 722, which would create non-comparability of levels around 2000. We solve this using a similar method as the NAICS-SIC alignment as outlined above. For all three states, we take data for 2001 (which is the first year with full data), and take the ratio of employment under the 3-digit code 722 to the sum of employment under the 4-digit codes 7221 and 7222. We then multiply the 7221 + 7222 employment numbers from 1990–2000 by this factor. This gives us an estimate for employment for the full 3-digit sector in the 1990s for these states. We do the same for earnings (wage bill). Using this estimated NAICS data for 1990, we can apply the NAICS-SIC scaling described above to these three states as well, thus giving us a full harmonized series from 1980–2019.

Other data.

We use the Current Population Survey (CPS) data to calculate earnings and employment for various groups (explained more in Appendix F). We also use CPS data to calculate the 16+ population in each state and year. Finally, we use state-level minimum wage data from ([Vaghul and Zipperer, 2016](#)) (the details are provided in Appendix E).

Appendix E Construction of 60 state-level minimum wage events

Identifying events.

The event-study analysis reported in [Table 1](#) uses 60 prominent state minimum wage increases, which are shown in [Fig. A1](#). We use the following process to identify these events:

1. We first classify as “initial events” all instances of a *state* minimum wage increase of at least \$0.25 and 5%.
2. Of these initial events, we retain those that occurred in years without a federal minimum wage increase as “admissible events.” This is because an event starting in a federal minimum wage year cannot have a “clean” post-period (see below for more details on clean controls). As a result, we exclude initial events in 1980, 1981, 1990, 1991, 1996, 1997, 2007, 2008, and 2009.
3. Among the admissible events, we classify an event as a “provisional combined event” (or the start of a potentially multi-year minimum wage increase) if no admissible event occurred in the preceding three years. This ensures that multi-year minimum wage increases are counted as a single event.
4. However, in some cases, this process may misidentify the true start year of a multi-year minimum wage increase. Specifically, if a multi-phase increase begins with a small increment, we might incorrectly conclude that the event started later than it actually did. To correct for this, we implement the following algorithm:
 - o For all provisional combined events, we check if there was a minimum wage increase (large or small) in the preceding year, as long as the previous year was not a federal minimum wage increase year.

- o If there was such an increase, we then check whether the state had indexation at that time. If it did not, the increase is considered a legislated (albeit small) minimum wage increase, and we adjust the event start date to the previous year. (We use data on indexation from (Brummund and Michael, 2020).)
 - o If the state did have indexation, we check if the small MW increase was a legislated increase. If it was indeed a legislated increase, we again move the event start date to the previous year.
 - o We repeat this process until no combined event start years have a legislated non-federal-year minimum wage increase in the preceding year.
5. The preceding procedure identifies 64 events. However, since the maximum post-period is six years (as discussed below), there are five cases where an event falls within the post-period of an earlier event. These instances are: Alaska 2010 and 2015; Delaware 2014 and 2019; New Jersey 2014 and 2019; Oregon 1998 and 2003; and Rhode Island 1999 and 2004. We must decide whether to treat the latter year as part of the post-period for the earlier event, or to let the latter year be a standalone event. We choose the first option (removing the latter event) if the minimum wage increase in the post-period of the latter event is no larger than the increase in the post-period of the former event. If the increase for the latter event is larger, we keep both events but reduce the post-period of the former event, so the latter event no longer overlaps with the former's post-period. In practice, we choose the first option for Delaware, New Jersey, Oregon, and Rhode Island—i.e., the latter events are dropped and considered part of the earlier event. For Alaska, we keep both the 2010 and 2015 events; as a result, we restrict the post-period for Alaska 2010 to five years instead of six. This results in a final list of 60 combined events, which are plotted by state and time in the Fig. A1.

Post-periods and clean controls.

To implement Eq. 5, we need to calculate the differences between average outcomes over (up to) six years following the event and the outcome in the year prior to the event. In some cases, the post-period is shortened either by a federal minimum wage increase or the end of the sample period in 2019.

For example, 1998 events have a full six-year post-period (1998–2003). In contrast, 2005 events only have a two-year post-period (2005–2006, since 2007 includes a federal minimum wage increase), while 2016 events have a four-year post-period (2016–2019). Therefore, we group post-periods by event-year cohorts—every event in the same year has the same post-period.⁴⁸ Table A1 summarizes the distribution of post-periods for each event-year cohort.

Post-periods also enter into the construction of clean controls. In a given year, a state is considered a clean control if it has *no* state MW increase (large or small) in the three preceding years, and *no* state MW increase in the post-period (as defined by that year's event-year cohort). For instance, a clean control for an event in 1998 must have had no state minimum wage increase during the 1995–2003 period, whereas a clean control for an event in 2005 only needs to have no increases during the 2002–2006 period.⁴⁹

⁴⁸ The one exception is Alaska in 2010, which has a five-year post-period, while the other 2010 event (Nevada) has a six-year post-period. This ensures that the 2016 Alaska event is not included in the evaluation of the 2010 Alaska event.

⁴⁹ For Alaska 2010, while we reduce the post-period to five years instead of six, clean controls must still be “clean” for six years, like other events in the 2010 cohort.

Appendix F Construction of probability groups using demographic predictors

Here we provide a brief summary of the approach. The details are provided in (Cengiz et al., 2022), which we follow to construct the probability groups.

We construct a model that predicts the likelihood of an individual being a minimum wage worker based on demographic variables, using data from the Current Population Survey over the years 1979–2019. For the minimum wage information, we select state-level events for which the given state did not experience significant minimum wage events in the past 20 quarters and which will experience a prominent change in the subsequent 12 quarters. In terms of demographic variables, the ones used in the prediction model are: age, education, sex, rural residency, marital status, race, Hispanic, and veteran status.

To find the best predictors we use only data from the 20 quarters preceding the minimum wage changes. The data is divided into two mutually exclusive samples: a training sample and a test sample. We apply a gradient boosted trees algorithm (Friedman, 2001) in the training sample, choosing the parameters using cross-validation and assessing model performance in the test sample. The trained model is then employed to estimate predicted probabilities for the comprehensive dataset. This procedure results in an individual-level dataset containing predicted probabilities of being a minimum wage worker, and encompassing all states and time periods over 1979–2019.

We then use that dataset to create three groups of individuals according to their predicted probabilities, defined as: (i) high probability – sample of top 10 % of predicted probability distribution; (ii) high-recall – broader sample comprising 75 % of all minimum wage workers which includes the high probability group; (iii) low probability – sample of individuals that are not contained in the high-recall group.

References

- Aaronson, D., 2001. Price pass-through and the minimum wage. *Rev. Econ. Stat.* 83 (1), 158–169.
- Aaronson, D., Brian, J.P., 2019. Wage shocks and the technological substitution of low-wage jobs. *Econ. J.* 129 (617), 1–34.
- Aaronson, D., French, E., MacDonald, J., 2008. The minimum wage, restaurant prices, and labor market structure. *J. Hum. Resour.* 43 (3), 688–720.
- Aaronson, D., French, E., Sorkin, I., To, T., 2018. Industry dynamics and the minimum wage: a putty-clay approach. *Int. Econ. Rev.* 59 (1), 51–84.
- Aaronson, D., Agarwal, S., French, E., 2012. The spending and debt response to minimum wage hikes. *Am. Econ. Rev.* 102 (7), 3111–3139.
- Abowd, J.M., Kramarz, F., Lemieux, T., Margolis, D.N., 2000. Minimum wages and youth employment in France and the United States. *Youth Employment and Joblessness in Advanced Countries*. University of Chicago Press, pp. 427–472.
- Acemoglu, D., Pischke, J.-S., 2003. Minimum wages and on-the-job training. *Worker Well-Being and Public Policy*. Emerald Group Publishing Limited, pp. 159–202.
- Adams, Camilla, Meer, J., Will Sloan, C., 2022. The minimum wage and search effort. *Econ. Lett.* 212, 110288.
- Addison, J.T., Blackburn, M.K., 1999. Minimum wages and poverty. *ILR Rev.* 52 (3), 393–409.
- Agan, A.Y., Makowsky, M.D., 2021. The minimum wage, EITC, and criminal recidivism. *J. Hum. Resour.*

- Agarwal, S., Meghana, A., Kosová, R., 2024. Minimum wage increases and employer performance: role of employer heterogeneity. *Manag. Sci.* 70 (1), 225–254.
- Ahlfeldt, G.M., Roth, D., Seidel, T., 2018. The regional effects of Germany's national minimum wage. *Econ. Lett.* 172, 127–130.
- Allegretto, S., Reich, M., 2018. Are local minimum wages absorbed by price increases? Estimates from internet-based restaurant menus. *ILR Rev.* 71 (1), 35–63.
- Allegretto, S., Dube, A., Reich, M., 2011. Do minimum wages really reduce teen employment? Accounting for heterogeneity and selectivity in state panel data. *Ind. Relations A J. Econ. Soc.* 50 (2), 205–240.
- Allegretto, S., Dube, A., Reich, M., Zipperer, B., 2017. Credible research designs for minimum wage studies: a response to Neumark, Salas, and Wascher. *ILR Rev.* 70 (3), 559–592.
- Alessandrini, D., Milla, J., 2024. Minimum-wage effects on human capital accumulation: evidence from Canadian Data. *J. Hum. Cap.* 18 (2).
- Alonso, C., 2022. Beyond labor market outcomes. *J. Hum. Resour.* 57 (5), 1690–1714.
- Angrist, J.D., Graddy, K., Imbens, G.W., 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Rev. Econ. Stud.* 67 (3), 499–527.
- Arulampalam, W., Alison, L.B., Mark, L.B., 2004. Training and the new minimum wage. *Econ. J.* 114 (494), C87–C94.
- Ashenfelter, O., Card, D., 1981. Using longitudinal data to measure minimum wage effects. CEP Discussion Paper, Cent. Economic Performance, LSE.
- Ashenfelter, O., Robert, S.S., 1979. Compliance with the minimum wage law. *J. Political Econ.* 87 (2), 333–350.
- Ashenfelter, O., Štěpán Jurajda, 2022. Minimum wages, wages, and price pass-through: the case of McDonald's restaurants. *J. Labor. Econ.* 40 (S1), S179–S201.
- Autor, D.H., Lawrence, F.K., Melissa, S.K., 2008. Trends in US wage inequality: revising the revisionists. *Rev. Econ. Stat.* 90 (2), 300–323.
- Autor, D.H., Manning, A., Smith, C.L., 2016. The contribution of the minimum wage to U.S. wage inequality over three decades: a reassessment. *Am. Econ. J. Appl. Econ.* 8 (1), 58–99.
- Autor, D., Dube, A., McGrew, A., 2023. The unexpected compression: competition at work in the low wage labor market. Technical report, National Bureau of Economic Research.
- Azar, J., Huet-Vaughn, E., Marinescu, I., Taska, B., Wachter, T. von, 2023. Minimum wage employment effects and labour market concentration. *Rev. Econ. Stud.* rdad 091.
- Bailey, M.J., DiNardo, J., Stuart, B.A., 2021. The economic impact of a high national minimum wage: Evidence from the 1966 fair labor standards act. *J. labor. Econ.* 39 (S2), S329–S367.
- Baker, M., Benjamin, D., Stanger, S., 1999. The highs and lows of the minimum wage effect: A time-series cross-section study of the Canadian law. *J. labor. Econ.* 17 (2), 318–350.
- Basker, E., Khan, M.T., 2016. Does the minimum wage bite into fast-food prices? *J. Labor. Res.* 37, 129–148.
- Bassier, I., 2022. Collective bargaining and spillovers in local labor markets.
- Bell, B., Machin, S., 2018. Minimum wages and firm value. *J. Labor. Econ.* 36 (1), 159–195.
- Bellmann, L., Mario Bossler, H.-D.G., Hübler, O., 2017. Training and minimum wages: first evidence from the introduction of the minimum wage in Germany. *IZA J. Labor. Econ.* 6, 1–22.
- Belman, D., Paul, J.W. 2014. What does the minimum wage do?
- Berger, D.W., Kyle, F.H., Mongey, S., 2022. Minimum wages, efficiency and welfare, 'Technical report. National Bureau of Economic Research.
- Bernhardt, Annette, Spiller, M., Theodore, N., 2013. Employers gone rogue: explaining industry variation in violations of workplace laws. *Ind. Labor. Relat. Rev.* 66, 808.

- Bhaskar, V., To, T., 1999. Minimum wages for Ronald McDonald Monopsonies: a theory of monopsonistic competition. *Econ. J.* 109 (455), 190–203.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., Roberts, J., 2013. Vol. 128 February 2013 Issue. *Q. J. Econ.* 1, 51.
- Bloom, N., Lemos, R., Sadun, R., Scur, D., Reenen, J.V., 2014. JEEA-FBBVA lecture 2013: The new empirical economics of management. *J. Eur. Economic Assoc.* 12 (4), 835–876.
- Blundell, R., Duncan, A., Meghir, C., 1998. Estimating labor supply responses using tax reforms. *Econometrica* 827–861.
- Böckerman, P., Uusitalo, R., 2009. Minimum wages and youth employment: evidence from the finnish retail trade sector. *Br. J. Ind. Relat.* 47 (2), 388–405.
- Bodnár, K., Ludmila Fadejeva, S.I., et al., 2018. How do firms adjust to rises in the minimum wage? Survey evidence from Central and Eastern Europe. *IZA J. Labor. Policy* 7, 1–30.
- Boeri, T., Giupponi, G., Krueger, A.B., Machin, S., 2020. Solo self-employment and alternative work arrangements: A cross-country perspective on the changing composition of jobs. *J. Econ. Perspect.* 34 (1), 170–195.
- Boffy-Ramirez, E., 2013. Minimum wages, earnings, and migration. *IZA J. Migr.* 2, 1–24.
- Boone, C., Dube, A., Goodman, L., Kaplan, E., 2021. Unemployment insurance generosity and aggregate employment. *Am. Econ. J. Econ. Policy* 13 (2), 58–99.
- Borgschulte, M., Cho, H., 2020. Minimum wages and retirement. *ILR Rev.* 73 (1), 153–177.
- Borusyak, K., Jaravel, X., Spiess, J., 2024. Revisiting event-study designs: robust and efficient estimation. *Rev. Econ. Stud.*
- Bosch, M., Manacorda, M., 2010. Minimum wages and earnings inequality in urban Mexico. *Am. Econ J Appl. Econ.* 2 (4), 128–149.
- Bossler, M., Hans-Dieter, G., 2020. Employment effects of the new german minimum wage: evidence from establishment-level microdata. *ILR Rev.* 73 (5), 1070–1094.
- Bossler, M., Broszeit, S., 2017. Do minimum wages increase job satisfaction? Micro-data evidence from the new German minimum wage. *LABOUR* 31 (4), 480–493.
- Bossler, M., Schank, T., 2023. Wage inequality in Germany after the minimum wage introduction. *J. Labor. Econ.* 41 (3), 813–857.
- Brochu, P., David, A.G., 2013. The impact of minimum wages on labour market transitions. *Econ. J.* 123 (573), 1203–1235.
- Brochu, P., David, A.G., Lemieux, T., Townsend, J., 2023. The minimum wage, turnover, and the shape of the wage distribution.
- Broecke, S., Forti, A., Vandeweyer, M., 2017. The effect of minimum wages on employment in emerging economies: a survey and meta-analysis. *Oxf. Dev. Stud.* 45 (3), 366–391.
- Brown, C., 1999. Minimum wages, employment, and the distribution of income. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, 3, Part B, 1st edition. Elsevier, pp. 2101–2163.
- Brown, C., Gilroy, C., Kohen, A., 1982. The effect of the minimum wage on employment and unemployment. *J. Econ. Lit.* 20, 2.
- Brown, J.H., Chris M.H., 2023. Minimum wage, worker quality, and consumer well-being: evidence from the child care market.
- Brummund, P., Michael, R.S., 2020. Does employment respond differently to minimum wage increases in the presence of inflation indexing? *J. Hum. Resour.* 55, 3.
- Buchanan, J., 1996. Commentary on the minimum wage. *Wall Str. J.*, April. 25, A20.
- Burauel, P., Caliendo, M., Grabka, M.M., Obst, C., Preuss, M., Schröder, C., 2020. The impact of the minimum wage on working hours. *Jahrbücher für Nationalökonomie und Statistik* 240 (2–3), 233–267.

- Burdett, K., Dale, T.M., 1998. Wage differentials, employer size, and unemployment. *Int. Econ. Rev.* 257–273.
- Burkhauser, R.V., McNichols, D., Sabia, J.J., 2023. Minimum wages and poverty: new evidence from dynamic difference-in-differences estimates. Technical report. National Bureau of Economic Research.
- Butcher, T., Dickens, R., Manning, A., 2012. Minimum wages and wage inequality: some theory and an application to the UK.
- Cadena, B.C., 2014. Recent immigrants as labor market arbitrageurs: Evidence from the minimum wage. *J. Urban. Econ.* 80, 1–12.
- Caetano, C., Callaway, B., Payne, S., Sant'Anna Rodrigues, H., 2022. Difference in differences with time-varying covariates. arXiv preprint arXiv:2202.02903.
- Caliendo, M., Wittbrodt, L., Schröder, C., 2019. The causal effects of the minimum wage introduction in Germany – an overview. *Ger. Econ. Rev.* 20 (3), 257–292.
- Caliendo, M., Fedorets, A., Preuss, M., Schröder, C., Wittbrodt, L., 2018. The short-run employment effects of the German minimum wage reform. *Labour Econ.* 53, 46–62 European Association of Labour Economists 29th annual conference, St.Gallen, Switzerland, 21-23 September 2017.
- Callaway, B., Sant'Anna, P.H.C., 2021. Difference-in-differences with multiple time periods. *J. Econom.* 225 (2), 200–230.
- Campoli, M., Fang, T., Gunderson, M., 2005. How minimum wages affect schooling-employment outcomes in Canada, 1993–1999. *J. Labor. Res.* 26 (3), 533–545.
- Card, D., 1992a. Do minimum wages reduce employment? A case study of California, 1987–89. *Ind. Labor. Relat. Rev.* 46 (1), 38–54.
- Card, D., 1992b. Using regional variation in wages to measure the effects of the federal minimum wage. *ILR Rev.* 46 (1), 22–37.
- Card, D., Alan, B.K., 2000. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: reply. *Am. Econ. Rev.* 90 (5), 1397–1420.
- Card, D., Alan, B.K., 1999a. The effect of the minimum wage on shareholder wealth, Working paper, Princeton University, Department of Economics, Industrial Relations Section.
- Card, D., Alan, B.K., 1994b. Minimum wages and employment: a case study of the New Jersey and Pennsylvania fast food industries. *Am. Economic Rev.* 84 (4), 772–793.
- Card, D., Alan, B.K., 1995a. Myth and measurement: the new economics of the minimum wage. Princeton University Press, New Jersey.
- Card, D., Alan, B.K., 1995b. Time-series minimum-wage studies: a meta-analysis. *Am. Econ. Rev.* 85 (2), 238–243.
- Card, D., John, E.D.N., 2002. Skill-biased technological change and rising wage inequality: Some problems and puzzles. *J. labor. Econ.* 20 (4), 733–783.
- Card, D., Cardoso, A.R., Heining, J., Kline, P., 2018. Firms and labor market inequality: Evidence and some theory. *J. Labor. Econ.* 36 (S1), S13–S70.
- Card, D., Heining, J., Kline, P., 2013. Workplace heterogeneity and the rise of West German wage inequality. *Q. J. Econ.* 128 (3), 967–1015.
- Card, D., Lawrence, F.K., Alan, B.K., 1994. Comment on David Neumark and William Wascher, Employment effects of minimum and subminimum wages: panel data on state minimum wage laws. *Ind. Labor. Relat. Rev.* 47 (3), 487–497.
- Castillo-Freeman, A.J., Richard, B.F., 1992. When the minimum wage really bites: The effect of the U. S.-level minimum on Puerto Rico. University of Chicago Press, Chicago, pp. 177–212.
- Cengiz, D., Dube, A., Lindner, A., Zipperer, B., 2019. The effect of minimum wages on low-wage jobs. *Q. J. Econ.* 134 (3), 1405–1454.

- Cengiz, D., Dube, A., Lindner, A., Zentler-Munro, D., 2022. Seeing beyond the trees: using machine learning to estimate the impact of minimum wages on labor market outcomes. *J. Labor. Econ.* 40 (S1), S203–S247.
- Chaplin, D.D., Mark, D.T., Andreas, D.P., 2003. Minimum wages and school enrollment of teenagers: a look at the 1990's. *Econ. Educ. Rev.* 22 (1), 11–21.
- Chava, S., Oettl, A., Singh, M., 2023. Does a one-size-fits-all minimum wage cause financial stress for small businesses? *Manag. Sci.* 69 (11), 7095–7117.
- Clemens, J., 2021. How do firms respond to minimum wage increases? Understanding the relevance of non-employment margins. *J. Econ. Perspect.* 35 (1), 51–72.
- Clemens, J., Wither, M., 2019. The minimum wage and the Great Recession: evidence of effects on the employment and income trajectories of low-skilled workers. *J. Public. Econ.* 170, 53–67.
- Clemens, J., Michael, R.S., 2021. The heterogeneous effects of large and small minimum wage changes: evidence over the short and medium run using a pre-analysis plan. Technical Rep. National Bureau of Economic Research.
- Clemens, J., Michael, R.S., 2022. Understanding “wage theft”: Evasion and avoidance responses to minimum wage increases. *Labour Econ.* 79, 102285.
- Clemens, J., Lisa, B.K., Meer, J., 2018. The Minimum Wage, Fringe Benefits, and Worker Welfare Working. National Bureau of Economic Research, pp. 24635.
- Clemens, J., Lisa, B.K., Meer, J., 2021. Dropouts need not apply? The minimum wage and skill upgrading. *J. Labor. Econ.* 39 (S1). <http://dx.doi.org/10.1086/711490>.
- Collier, R.B., Collier, D., 2002. Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America. University of Notre Dame Press.
- Congressional Budget Office, (CBO), 2019. The effects on employment and family income of increasing the federal minimum wage. <https://www.cbo.gov/system/files/2019-07/CBO-55410-MinimumWage2019.pdf>, Accessed: 2024-07-30.
- Connolly, S., Gregory, M., 2002. The national minimum wage and hours of work: implications for low paid women. *Oxf. Bull. Econ. Stat.* 64 (ement), 607–631.
- Cook, M.L., 2010. Politics of Labor Reform in Latin America: Between Flexibility and Rights. Penn State Press.
- Cooper, D., Luengo-Prado, M.J., Parker, J.A., 2020. The local aggregate effects of minimum wage increases. *J. Money, Credit. Banking* 52 (1), 5–35. <https://doi.org/10.1111/jmcb.12684>
- Couch, K.A., David, C.W., 2001. The response of hours of work to increases in the minimum wage. *South. Econ. J.* 68 (1), 171–177.
- Coviello, D., Deserranno, E., Persico, N., 2022. Minimum wage and individual worker productivity: evidence from a large US retailer. *J. Political Econ.* 130 (9), 2315–2360.
- Currie, J., Bruce, C.F., 1996. The minimum wage and the employment of youth evidence from the NLSY. *J. Hum. Resour.* 31 (2), 404–428.
- Datta, N., Machin, S., 2024. Government contracting and living wages; minimum wages. Technical Report, Cent. Economic Performance, LSE.
- Dautović, E., Hau, H., Huang, Y., 2024. Consumption response to minimum wages: evidence from Chinese households. *Rev. Econ. Stat.* 1–47.
- De Chaisemartin, C., d'Haultfoeuille, X., 2018. Fuzzy differences-in-differences. *Rev. Econ. Stud.* 85 (2), 999–1028.
- De Chaisemartin, C., d'Haultfoeuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. *Am. Econ. Rev.* 110 (9), 2964–2996.
- DellaVigna, S., Gentzkow, M., 2019. Uniform pricing in us retail chains. *Q. J. Econ.* 134 (4), 2011–2084.

- Demir, B., Fiebler, A.C., Xu, D.Y., Yang, K.K., 2024. O-ring production networks. *J. Political Econ.* 132 (1), 200–247.
- Derenoncourt, E., Montialoux, C., 2021. Minimum wages and racial inequality. *Q. J. Econ.* 136 (1), 169–228.
- Derenoncourt, E., Gérard, F., Lagos, L., Montialoux, C., 2021. Racial inequality, minimum wage spillovers, and the informal sector.
- Detting, L.J., Hsu, J.W., 2020. Minimum wages and consumer credit: effects on access and borrowing. *Rev. Financial Stud.* 34 (5), 2549–2579.
- Dickens, R., Machin, S., Manning, A., 1998. Estimating the effect of minimum wages on employment from the distribution of wages: A critical view. *Labour Econ.* 5 (2), 109–134.
- Dickens, R., Machin, S., Manning, A., 1999. The effects of minimum wages on employment: Theory and evidence from Britain. *J. Labor. Econ.* 17 (1), 1–22.
- DiNardo, J., Nicole, M.F., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. *Econometrica* 64 (5), 1001–1044.
- Dolado, J., Francis Kramarz, S.M., Alan Manning, D.M., Teulings, C., 1996. Minimum wages: the European experience. *Econ. Policy* 23 (3), 317–357.
- Dolton, P., Bondibene, C.R., Wadsworth, J., 2012. Employment, inequality and the UK national minimum wage over the medium-term. *Oxf. Bull. Econ. Stat.* 74 (1), 78–106.
- Doucouliagos, H., Tom, D.S., 2009. Publication selection bias in minimum-wage research? A meta-regression analysis. *Br. J. Ind. Relat.* 47 (2), 406–428.
- Downey, M., 2021. Partial automation and the technology-enabled deskilling of routine jobs. *Labour Econ.* 69, 101973.
- Draca, M., Machin, S., Van Reenen, J., 2011. Minimum wages and firm profitability. *Am. Econ. J. Appl. Econ.* 3 (1), 129–151.
- Drucker, L., Mazirov, K., Neumark, D., 2021. Who pays for and who benefits from minimum wage increases? Evidence from Israeli tax data on business owners and workers. *J. Public. Econ.* 199, 104423.
- Dube, A., 2019a. Impacts of minimum wages: review of the international evidence. *Independent Report*. UK Gov. Publ. 268–304.
- Dube, A., 2019b. Minimum wages and the distribution of family incomes. *Am. Econ. J. Appl. Econ.* 11 (4), 268–304.
- Dube, A., Manning, A., Naidu, S., 2020. Monopsony and employer mis-optimization explain why wages bunch at round numbers.
- Dube, A., Daniele Girardi, Ò.J., Alan, M.T., 2023. A local projections approach to difference-in-differences. Working. National Bureau of Economic Research, pp. 31184.
- Dube, A., Giuliano, L., Leonard, J., 2019. Fairness and frictions: the impact of unequal raises on quit behavior. *Am. Econ. Rev.* 109 (2), 620–663.
- Dube, A., Naidu, S., Reich, A.D., 2022. Power and dignity in the low-wage labor market: Theory and evidence from wal-mart workers. Technical report. National Bureau of Economic Research.
- Dube, A., Naidu, S., Reich, M., 2007. The economic effects of a citywide minimum wage. *ILR Rev.* 60 (4), 522–543.
- Dube, A., Lester, T.W., Reich, M., 2016. Minimum wage shocks, employment flows, and labor market frictions. *J. Labor. Econ.* 34 (3), 663–704.
- Dube, A., Lester, T.W., Reich, M., 2010. Minimum wage effects across state borders: estimates using contiguous counties. *Rev. Econ. Stat.* 92 (4), 945–964.
- Dube, E.F., Arindrajit, Reich, M., 2010. Employee replacement costs, *Institute for Resarch on Labor and Employment*. UC. Berkeley Working Pap. 201–10.

- Dube, A., Lindner, A., 2021. City limits: what do local-area minimum wages do? *J. Econ. Perspect.* 35 (1), 27–50.
- Dube, A., Zipperer, B., 2024. Own-wage elasticity: quantifying the impact of minimum wages on employment.
- Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., Vom Berge, P., 2022. Reallocation effects of the minimum wage. *Q. J. Econ.* 137 (1), 267–328.
- Dustmann, C., Ludsteck, J., Schönberg, U., 2009. Revisiting the German wage structure. *Q. J. Econ.* 124 (2), 843–881.
- Emanuel, N., Harrington, E., 2020. The payoffs of higher pay: elasticities of productivity and labor supply with respect to wages.
- Engbom, N., Christian, M., 2022. Earnings inequality and the minimum wage: Evidence from Brazil. *Am. Econ. Rev.* 112 (12), 3803–3847.
- Fan, H., Hu, Y., Tang, L., 2021. Labor costs and the adoption of robots in China. *J. Econ. Behav. Organ.* 186, 608–631.
- Ferman, B., Pinto, C., 2019. Inference in differences-in-differences with few treated groups and heteroskedasticity. *Rev. Econ. Stat.* 101 (3), 452–467.
- Fishback, P.V., Seltzer, A.J., 2021. The rise of American minimum wages, 1912–1968. *J. Econ. Perspect.* 35 (1), 73–96.
- Flinn, C.J., 2006. Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica* 74 (4), 1013–1062.
- Flinn, C.J., 2010. The Minimum Wage and Labor Market Outcomes. MIT Press.
- Fortin, N.M., Lemieux, T., Lloyd, N., 2021. Labor market institutions and the distribution of wages: the role of spillover effects. *J. Labor. Econ.* 39 (S2), S369–S412.
- Freeman, R., 1989. Labor Markets in Action. Harvard University Press, Cambridge.
- Freeman, R.B., 1990. Employment and earnings of disadvantaged young men in a labor shortage economy, Working Paper 3444, National Bureau of Economic Research.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 1189–1232.
- Friedman, M., 1953. The methodology of positive economics. *Essays on Positive Economics*. University of Chicago Press, Chicago.
- Gardner, J., Thakral, N., Tô, T., Luther, Y., 2024. Two-stage differences in differences.
- Geng, H., Huang, Y., Lin, C., Liu, S., 2022. Minimum wage and corporate investment: evidence from manufacturing firms in China. *J. Financial Quant. Anal.* 57 (1), 94–126.
- Gittings, R.K., Schmutte, I.M., 2016. Getting handcuffs on an octopus: Minimum wages, employment, and turnover. *ILR Rev.* 69 (5), 1133–1170.
- Giuliano, L., 2013. Minimum wage effects on employment, substitution, and the teenage labor supply: evidence from personnel data. *J. Labor. Econ.* 31 (1), 155–194.
- Giulietti, C., 2014. Is the minimum wage a pull factor for immigrants? *ILR Rev.* 67 (3_suppl), 649–674.
- Giupponi, G., Machin, S., 2022a. Labour market inequality. *IFS Deaton Rev.* 15.
- Giupponi, G., Machin, S., 2022b. Company Wage Policy in a Low-Wage Labor Market, CEP Discussion Paper CEPDP1869, Centre for Economic Performance, LSE.
- Giupponi, G., Joyce, R., Lindner, A., Waters, T., Wernham, T., Xu, X., 2024. The employment and distributional impacts of nationwide minimum wage changes. *J. Labor. Econ.* 42 (S1), S293–S333.
- Godoey, A., Reich, M., 2021. Are minimum wage effects greater in low-wage areas? *Ind. Relations A J. Econ. Soc.* 60 (1), 36–83.
- Godøy, A., Jacobs, K., 2021. The downstream benefits of higher incomes and wages. Fed. Reserve Bank. Boston Community Dev. Discussion Pap. 21 -21.

- Godøy, A., Reich, M., Wursten, J., Allegretto, S., 2024. Parental labor supply: evidence from minimum wage changes. *J. Hum. Resour.* 59 (2), 416–442.
- Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. *J. Econom.* 225 (2), 254–277.
- Gopalan, R., Barton, H.H., Kalda, A., Sovich, D., 2021. State minimum wages, employment, and wage spillovers: Evidence from administrative payroll data. *J. Labor. Econ.* 39 (3), 673–707.
- Goraus-Tańska, K., Lewandowski, P., 2019. Minimum wage violation in central and eastern Europe. *Int. Labour Rev.* 158 (2), 297–336.
- Grimshaw, D., Miozzo, M., 2003. Minimum Wages and Pay Equity in Latin America 12 International Labour Office.
- Grossberg, A.J., Sicilian, P., 1999. Minimum wages, on-the-job training, and wage growth. *South. Econ. J.* 65 (3), 539–556.
- Grossman, J., 1978. Fair labor standards act of 1938: maximum struggle for a minimum wage. *Monthly Labor. Rev.* 101 (6), 22–30.
- Gülal, F., Ayaita, A., 2020. The impact of minimum wages on well-being: evidence from a quasi-experiment in Germany. *J. Happiness Stud.* 21, 2669–2692.
- Haanwinckel, D., 2023. Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution. Technical report. National Bureau of Economic Research.
- Hamermesh, D.S., 1995. Labour demand and the source of adjustment costs. *Econ. J.* 105 (430), 620–634.
- Hammond, M.B., 1915. Wages boards in Australia: IV. Social and economic results of wages boards. *Q. J. Econ.* 563–630.
- Hampton, M., Evan Totty, 2023. Minimum wages, retirement timing, and labor supply. *J. Public. Econ.* 224, 104924.
- Hara, H., 2017. Minimum wage effects on firm-provided and worker-initiated training, *Labour Economics*, 47, 149–162, EALE conference issue 2016.
- Harasztosi, P., Lindner, A., 2019. Who pays for the minimum wage? *Am. Econ. Rev.* 109 (8), 2693–2727.
- Hau, H., Huang, Y., Wang, G., 2020. Firm response to competitive shocks: evidence from China's minimum wage policy. *Rev. Econ. Stud.* 87 (6), 2639–2671.
- Hazell, J., Patterson, C., Sarsons, H., Taska, B., 2022. National wage setting. University of Chicago, Becker Friedman Institute for Economics Working Paper. pp. 2022–2150.
- Hicks, J., 1932. *The Theory of Wages*. Macmillan, London.
- Hirsch, B.T., Bruce, E.K., Zelenska, T., 2015. Minimum wage channels of adjustment. *Ind. Relations A J. Econ. Soc.* 54 (2), 199–239.
- Holzer, H.J., Lawrence, F.K., Alan, B.K., 1991. Job queues and wages. *Q. J. Econ.* 106 (3), 739–768.
- Horton, J.J., (Forthcoming). Price floors and employer preferences: Evidence from a minimum wage experiment. *Am. Econ. Rev.*
- Hsieh, C.-T., Peter, J.K., 2009. Misallocation and manufacturing TFP in China and India. *Q. J. Econ.* 124 (4), 1403–1448.
- Huang, Y., Loungani, M.P., Wang, G., 2014. Minimum wages and firm employment: evidence from China: international monetary fund.
- ILO, 2017. 'Minimum wage policy guide.'
- ILO, OECD, IMF, World Bank, 2012. Boosting jobs and living standards in G20 countries.
- Jacobs, K., Reich, M., Challenor, T., Farmand, A., 2024. Gig passenger and delivery driver pay in five metro areas.

- Jardim, E., Mark, C.L., Plotnick, R., van Inwegen, E., Vigdor, J., Wething, H., 2017. Minimum wage increases, wages, and low-wage employment: evidence from Seattle. NBER Working Paper No. 23532.
- Jardim, E., Mark, C.L., Plotnick, R., van Inwegen, E., Vigdor, J., Wething, H., 2022. Minimum-wage increases and low-wage employment: evidence from seattle. *Am. Econ. J. Econ. Policy* 14 (2), 263–314.
- Jha, P., Neumark, D., Rodriguez-Lopez, A., 2024. What's across the border? Re-evaluating the cross-border evidence on minimum wage effects. *J. Political Econ. Microecon.*, Forthcom.
- Kabátek, J., 2021. Happy birthday, you're fired! The effects of an age-dependent minimum wage on youth employment flows in the Netherlands. *ILR Rev.* 74 (4), 1008–1035.
- Kaitz, H., 1970. Experience of the past: the national minimum. *Youth Unemployment Minim. Wages* 30–54.
- Karabarbounis, L., Lise, J., Nath, A., 2022. Minimum wages and labor markets in the Twin Cities. Technical report. National Bureau of Economic Research.
- Katz, L.F., Alan, B.K., 1992. The effect of the minimum wage on the fast-food industry. *ILR Rev.* 46 (1), 6–21.
- Katz, L.F., Kevin, M.M., 1992. Changes in relative wages, 1963–1987: supply and demand factors. *Q. J. Econ.* 107 (1), 35–78.
- Kellogg, M., Mogstad, M., Pouliot, G.A., Torgovitsky, A., 2021. Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *J. Am. Stat. Assoc.* 116 (536), 1804–1816.
- Kennan, J., 1995. The elusive effects of minimum wages. *J. Econ. Lit.* 33 (4), 1950–1965.
- Khurana, S., K. Mahajan, 2020. Evolution of wage inequality in India (1983-2017): the role of occupational task content. WIDER working paper.
- Khurana, S., Mahajan, K., Sen, K., 2023a. Minimum wages and changing wage inequality in India. Technical report. Institute of Labor Economics (IZA).
- Khurana, S., Mahajan, K., Sen, K., 2023b. Minimum wages and changing wage inequality in India, IZA discussion papers.
- Kreiner, C.T., Reck, D., Ebbesen Skov, P., 2020. Do lower minimum wages for young workers raise their employment? Evidence from a danish discontinuity. *Rev. Econ. Stat.* 102 (2), 339–354.
- Krueger, A.B., 1994. The effect of the minimum wage when it really bites: a reexamination of the evidence from Puerto Rico. Working. National Bureau of Economic Research, pp. 4757.
- Ku, H., 2022. Does minimum wage increase labor productivity? Evidence from piece rate workers. *J. Labor. Econ.* 40 (2), 325–359.
- Lavecchia, A.M., 2020. Minimum wage policy with optimal taxes and unemployment. *J. Public. Econ.* 190, 104228.
- Laws, Athene, 2018. Do minimum wages increase search effort? Technical report, Faculty of Economics, University of Cambridge.
- Lee, C.H., 2020. Minimum wage policy and community college enrollment patterns. *ILR Rev.* 73 (1), 178–210.
- Lee, D., Saez, E., 2012. Optimal minimum wage policy in competitive labor markets. *J. Public. Econ.* 96 (9-10), 739–749.
- Lee, D.S., 1999. Wage inequality in the United States during the 1980s: rising dispersion or falling minimum wage? *Q. J. Econ.* 114 (3), 977–1023.
- Legal, D., Eric, R.Y., 2024. The effect of minimum wages on consumer bankruptcy. *J. Econ. Business* 129 Inequality in Consumer Credit and Payments, 106171.
- Leibenstein, H., 1966. Allocative efficiency vs. X-efficiency. *Am. Econ. Rev.* 56 (3), 392–415.

- Leigh, J.P., Wesley, A.L., Du, J., 2019. Minimum wages and public health: a literature review. *Preventive Med.* 118, 122–134.
- Lemos, S., 2006. Anticipated effects of the minimum wage on prices. *Appl. Econ.* 38 (3), 325–337.
- Lemos, S., Rigobon, R., Lang, K., 2004. Minimum wage policy and employment effects: evidence from brazil [with comments]. *Economia* 5 (1), 219–266.
- Leonard, T., 2000. The very idea of applying economics: The modern minimum-wage controversy and its antecedents. *History Political Econ.* 32 (5), 117–144.
- Lester, R.A., 1946. Shortcomings of marginal analysis for wage-employment problems. *Am. Econ. Rev.* 36 (1), 63–82.
- Leung, J.H., 2021. Minimum wage and real wage inequality: evidence from pass-through to retail prices. *Rev. Econ. Stat.* 103 (4), 754–769.
- Levin-Waldman, O.M., 2000. The effects of the minimum wage: a business response. *J. Econ. Issues* 34 (3), 723–730.
- Liu, S., Thomas, J.H., Regmi, K., 2016. Impact of the minimum wage on youth labor markets. *Labour* 30 (1), 18–37.
- Lordan, G., Neumark, D., 2018. People versus machines: the impact of minimum wages on automatable jobs. *Labour Econ.* 52, 40–53.
- Luca, D.L., Luca, M., 2019. Survival of the fittest: the impact of the minimum wage on firm exit. Working. National Bureau of Economic Research, pp. 25806.
- Luna-Alpizar, J.L., 2019. Worker Heterogeneity Asymmetric Eff. Minim. Wages. CERGE-EI Working Papers.
- Machin, S., 2024. Wage controversies: real wage stagnation, inequality and labour market institutions. *LSE Public. Policy Rev.* 3, 2.
- Machin, S., Manning, A., Rahman, L., 2003. Where the minimum wage bites hard: introduction of minimum wages to a low wage sector. *J. Eur. Econ Assoc.* 1 (1), 154–180.
- Machlup, F., 1946. Marginal analysis and empirical research. *Am. Econ. Rev.* 36 (4), 519–554.
- MacCurdy, T., 2015. How effective is the minimum wage at supporting the poor? *J. Political Econ.* 123 (2), 497–545.
- Malan, T., 1978. Wage control and minimum wages in Africa. *Afr. Insight* 8 (1), 3–17.
- Manning, A., 2003. Monopsony in Motion: Imperfect Competition in Labor Markets. Princeton University Press.
- Manning, A., 2021. The elusive employment effect of the minimum wage. *J. Econ. Perspect.* 35 (1), 3–26.
- Mansoor, K., O'Neill, D., 2021. Minimum wage compliance and household welfare: an analysis of over 1500 minimum wages in India. *World Dev.* 147, 105653.
- Marks, M.S., 2011. Minimum wages, employer-provided health insurance, and the non-discrimination law. *Ind. Relations A J. Econ. Soc.* 50 (2), 241–262.
- Marshall, A., 1897. The old generation of economists and the new. *Q. J. Econ.* 11 (2), 115–135.
- Mayneris, F., Poncet, S., Zhang, T., 2018. Improving or disappearing: firm-level adjustments to minimum wages in China. *J. Dev. Econ.* 135, 20–42.
- McGuinness, S., Redmond, P., 2018. Estimating the Effect of an Increase in the Minimum Wage on Hours Worked and Employment in Ireland. IZA Discussion. Institute of Labor Economics (IZA), pp. 11632.
- Meer, J., West, J., 2016. Effects of the minimum wage on employment dynamics. *J. Hum. Resour.* 51 (2), 500–522.
- Meiselbach, M.K., Jean, M.A., 2023. Do minimum wage laws affect employer-sponsored insurance provision? *J. health Econ.* 92, 102825.

- Metcalf, D., 1999. The Low Pay Commission and the national minimum wage. *Econ. J.* 109 (453), 46–66.
- Meyer, B., 2016. Learning to love the government: trade unions and late adoption of the minimum wage. *World Politics* 68 (3), 538–575.
- Meyer, R.H., David, A.W., 1983. The effects of the minimum wage on the employment and earnings of youth. *J. Labor. Econ.* 1 (1), 66–100.
- Mill, J.S., 1848. *Principles of Political Economy*. John W. Parker.
- Ministry of Manpower in Singapore, 2018. Opinion Editorial by Minister Josephine Teo on Minimum Wage.
- Minton, R., Wheaton, B., 2023. Minimum Wages and Internal Migration.
- Monras, J., 2019. Minimum wages and spatial equilibrium: theory and evidence. *J. Labor. Econ.* 37 (3), 853–904.
- Neumark, D., 2024. The effects of minimum wages on (almost) everything? A review of recent evidence on health and related behaviors. *Labour* 38 (1), 1–65.
- Neumark, D., Munguía Corella, L.F., 2021. Do minimum wages reduce employment in developing countries? A survey and exploration of conflicting evidence. *World Dev.* 137, 105165.
- Neumark, D., Peter Shirley, 2022. Myth or measurement: What does the new minimum wage research say about minimum wages and job loss in the United States? *Ind. Relations A J. Econ Soc.* 61 (4), 384–417.
- Neumark, D., Wascher, W., 1992. Employment effects of minimum and subminimum wages: panel data on state minimum wage laws. *ILR Rev.* 46 (1), 55–81.
- Neumark, D., Wascher, W., 1994. Employment effects of minimum and subminimum wages: reply to card, katz, and krueger. *Ind. Labor. Relat. Rev.* 47 (3), 497–512.
- Neumark, D., Wascher, W., 1995a. Minimum wage effects on employment and school enrollment. *J. Bus. Econ. Stat.* 13 (2), 199–206.
- Neumark, D., Wascher, W., 1995b. Minimum-wage effects on school and work transitions of teenagers. *Am. Econ. Rev.* 85 (2), 244–249.
- Neumark, D., Wascher, W., 2000. 'Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: comment. *Am. Econ. Rev.* 90 (5), 1362–1396.
- Neumark, D., Wascher, W., 2001a. Minimum wages and training revisited. *J. Labor. Econ.* 19 (3), 563–595.
- Neumark, D., Wascher, W., 2001b. Using the EITC to help poor families: new evidence and a comparison with the minimum wage. *Natl Tax. J.* 54 (2), 281–317.
- Neumark, D., Wascher, W., 2003. Minimum wages and skill acquisition: Another look at schooling effects. *Econ. Educ. Rev.* 22 (1), 1–10.
- Neumark, D., Wascher, W., 2008. Minimum Wages. MIT Press, Cambridge, MA.
- Neumark, D., Wascher, W., 2011. Does a higher minimum wage enhance the effectiveness of the earned income tax credit? *ILR Rev.* 64 (4), 712–746.
- Neumark, D., Ian Salas, J.M., Wascher, W., 2014. Revisiting the minimum wage—employment debate: throwing out the baby with the bathwater. *Ilr Rev.* 67 (3_suppl), 608–648.
- Neumark, D., Schweitzer, M., Wascher, W., 2004. Minimum wage effects throughout the wage distribution. *J. Hum. Resour.* 39 (2), 425–450.
- Neumark, D., Schweitzer, M., Wascher, W., 2005. The effects of minimum wages on the distribution of family incomes: a nonparametric analysis. *J. Hum. Resour.* 40 (4), 867–894.
- Newell, P., 2009. A historical overview of the fair labor standards act. *Fla. Coast. Law Rev.* 10 (4), 675.
- Nordlund, W.J., 1997. *The quest for a living wage: the history of the federal minimum wage program* (48). Greenwood publishing group.

- Okudaira, H., Takizawa, M., Yamanouchi, K., 2019. Minimum wage effects across heterogeneous markets. *Labour Econ.* 59, 110–122.
- Orazem, P.E., Peter Mattila, J., 2002. Minimum wage effects on hours, employment, and number of firms: The Iowa case. *J. Labor. Res.* 23, 1.
- Orrenius, P.M., Zavodny, M., 2008. The effect of minimum wages on immigrants' employment and earnings. *ILR Rev.* 61 (4), 544–563.
- Osterman, P., 2011. Institutional labor economics, the new personnel economics, and internal labor markets: a reconsideration. *ILR Rev.* 64 (4), 637–653.
- Pereira, S.C., 2003. The impact of minimum wages on youth employment in Portugal. *Eur. Econ. Rev.* 47 (2), 229–244.
- Phelan, B.J., 2019. Hedonic-based labor supply substitution and the ripple effect of minimum wages. *J. Labor. Econ.* 37 (3), 905–947.
- Piqueras, J., 2023. Search effort and the minimum wage.
- Portugal, P., Cardoso, A.R., 2006. Disentangling the minimum wage puzzle: an analysis of worker accessions and separations. *J. Eur. Econ Assoc.* 4 (5), 988–1013.
- Rao, N., Risch, M., 2024. Who's afraid of the minimum wage? Measuring the impacts on independent businesses using matched US tax returns.
- Rebitzer, J.B., Lowell, J.T., 1995. The consequences of minimum wage laws some new theoretical ideas. *J. Public. Econ.* 56 (2), 245–255.
- Reich, M., Laiten, A., 2003. Raising low pay a high income economy: Econ. a San. Francisco Minim. wage.
- Reich, M., West, R., 2015. The effects of minimum wages on food stamp enrollment and expenditures. *Ind. Relations: A J. Econ. Soc.* 54 (4), 668–694.
- Renkin, T., Montialoux, C., Siegenthaler, M., 2022. The pass-through of minimum wages into U.S. retail prices: evidence from supermarket scanner data. *Rev. Econ. Stat.* 104 (5), 890–908.
- Riley, R., Bondibene, C.R., 2017. Raising the standard: minimum wages and firm productivity. *Labour Econ.* 44, 27–50.
- Rohlin, S.M., 2011. State minimum wages and business location: evidence from a refined border approach. *J. Urban. Econ.* 69 (1), 103–117.
- Rositani, A., 2017. Work and wages in the code of Hammurabi. *Work. Wages Code Hammurabi* 47–71.
- Roth, J., Pedro, H.C.S.'A., Bilinski, A., Poe, J., 2023. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *J. Econ.* 235 (2), 2218–2244.
- Rothstein, J., 2010. Is the EITC as good as an NIT? Conditional cash transfers and tax incidence. *Am. Econ. J. Econ. Policy* 2 (1), 177–208.
- Ruffini, K., 2022. Worker earnings, service quality, and firm profitability: evidence from nursing homes and minimum wage reforms. *Rev. Econ. Stat.* 1–46.
- Sabia, J.J., Robert, B.N., 2015. Minimum wages, poverty, and material hardship: new evidence from the SIPP. *Rev. Econ. Househ.* 13, 95–134.
- Sabia, J.J., 2009. The effects of minimum wage increases on retail employment and hours: new evidence from monthly CPS data. *J. Labor. Res.* 30, 75–97.
- Schanzenbach, Whitmore, D., Turner, J.A., Turner, S., 2024. Raising state minimum wages, lowering community college enrollment. *Rev. Econ. Stat.* 1–29.
- Shiller, R.J., 1994. Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks. OUP, Oxford.
- Sidgwick, H., 1886. Economic socialism. *History Economic Thought Artic.* 50, 620–631.
- Simon, K.I., Kaestner, R., 2004. Do minimum wages affect non-wage job attributes? Evidence on fringe benefits. *ILR Rev.* 58 (1), 52–70.

- Smith, A.A., 2021. The minimum wage and teen educational attainment. *Labour Econ.* 73, 102061.
- Song, J., David, J.P., Guvenen, F., Bloom, N., von Wachter, T., 2018. Firming up inequality. *Q. J. Econ.* 134 (1), 1–50.
- Soundararajan, V., 2019. Heterogeneous effects of imperfectly enforced minimum wages in low-wage labor markets. *J. Dev. Econ.* 140, 355–374.
- Stansbury, A., 2024. Incentives to comply with the minimum wage in the US and UK. *ILR Rev.* Forthcom.
- Starr, G., 1981. Minimum wage fixing: international experience with alternative roles. *Int. Labour Rev.* 120, 545.
- Stewart, M.B., 2002. Estimating the impact of the minimum wage using geographical wage variation. *Oxf. Bull. Econ. Stat.* 64 (ement), 583–605.
- Stewart, M.B., Joanna, K.S., 2008. The other margin: do minimum wages cause working hours adjustments for low-wage workers? *Economica* 75 (297), 148–167.
- Stigler, G.J., 1946. The economics of minimum wage legislation. *Am. Econ. Rev.* 36 (3), 358–365.
- Sun, L., Jesse, M.S., 2022. A linear panel model with heterogeneous coefficients and variation in exposure. *J. Econ. Perspect.* 36 (4), 193–204.
- Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *J. Econom.* 225 (2), 175–199.
- Teulings, C.N., 2000. Aggregation bias in elasticities of substitution and the minimum wage paradox. *Int. Econ. Rev.* 41 (2), 359–398.
- Totty, E., 2017. The effect of minimum wages on employment: A factor model approach. *Econ. Inq.* 55 (4), 1712–1737.
- Umkehrer, M., Berge, P., 2020. Evaluating the minimum-wage exemption of the long-term unemployed in Germany. *ILR Rev.* 73.
- Vaghul, K., Zipperer, B., 2016. Historical state and sub-state minimum wage data. Wash. Cent. Equitable Growth Working Pap. <http://cdn.equitablegrowth.org/wp-content/uploads/2016/09/02153029/090716-WP-Historical-min-wage-data.pdf>.
- Van Den Berg, G.J., 2003. Multiple equilibria and minimum wages in labor markets with informational frictions and heterogeneous production technologies. *Int. Econ. Rev.* 44 (4), 1337–1357.
- Vergara, D., 2023. Minimum wages and optimal redistribution: The role of firm profits.
- Webb, S., 1912. The economic theory of a legal minimum wage. *J. Political Econ.* 20 (10), 973–998.
- Wessels, W., 2005. Does the minimum wage drive teenagers out of the labor force? *J. Labor. Res.* 26 (1), 169–176.
- Wiltshire, J., 2022. Walmart supercenters and monopsony. Power: How Large, Low-Wage Empl. Impacts Local Labor Markets.
- Wiltshire, J.C., McPherson, C., Reich, M., Sosinskiy, D., 2023. Minimum wage effects and monopsony explanations.
- Wooldridge, J.M., 2012. *Introductory Econometrics: A Mod. Approach*. South-Western Cengage Learn.
- Wursten, J., Reich, M., 2023. Racial inequality in frictional labor markets: evidence from minimum wages. *Labour Econ.* 82, 102344.
- Yao, W., Qian, Y., Yang, H., Xu, W., 2023. Does minimum wages affect executive compensation? – Evidence from China. *Pacific-Basin Financ. J.* 80, 102080.
- Zavodny, M., 2000. The effect of the minimum wage on employment and hours. *Labour Econ.* 7 (6), 729–750.

Further reading

- Aaronson, D., French, E., 2007. Product market evidence on the employment effects of the minimum wage. *J. Labor. Econ.* 25 (1), 167–200.
- Gibbons, R., Roberts, J., 2015. Organizational economics. *Emerg. Trends Soc. Behav. Sci.* American Cancer Society, pp. 1–15.
- Kertesi, G., Köllö, J., 2003. Fighting “Low Equilibria” Doubling Minim. Wage?? Hungary’s Experiment.
- Kudlyak, M., Tasci, M., Tuzemen, D., 2023. Minimum wage increases and vacancies. *Working 2022*, 10.

This page intentionally left blank

Chapter 5

The micro and macro economics of short-time work[☆]

Pierre Cahuc*

Sciences Po & IZA & CEPR

*Corresponding author. e-mail address: pierre.cahuc@sciencespo.fr

Chapter Outline

1 Introduction	386	5.1 The social willingness to pay for STW	414
2 Overview of STW schemes	388	5.2 The Net public cost of STW	416
2.1 The spread of STW since the 1920s	388	6 Effects of STW at the macro-economic level	419
2.2 STW and other workforce retention measures	389	6.1 STW before the Great recession of 2008-2009	419
2.3 The design of STW schemes	389	6.2 STW in the Great recession of 2008-2009	420
3 STW take-up	393	6.3 STW in the Covid-19 crisis	421
3.1 Take-up by type of workers and firms	393	6.4 The role of the timing of STW regulation and eligibility criteria	422
3.2 Take-up in large recessions	401	7 Effects on firms	423
3.3 Take-up outside of large recessions	401	7.1 Employment and hours of work	423
3.4 Take-up and labor market regulation	403	7.2 Firm productivity, profitability and firm survival	426
3.5 STW design and administrative capacity	404	7.3 Job reallocation and productivity	427
4 The theoretical models of STW	406	8 Effects on workers trajectories	428
4.1 Normative approach	406	9 Conclusion	429
4.2 Positive approach	407	References	430
5 The efficiency of STW	412		

☆ I thank Christian Dustmann, Antoine Ferey, Pat Kline, Thomas Lemieux, and the participants in the Handbook of Labor Economics Conference in Berlin in October 2023 for their insightful comments. I thank Zana Babaie (Department of Social Protection, Ireland), Britta Gehrke, Pierre Gramme (ADEM Statistiques et Etudes, Luxembourg), Rasmus GrØndahl (Danish Agency for Labor Market and Recruitment), Takao Kato, Daniel Kopp, Claudio Lucifora, Veronika Murauer (Arbeitsmarktforschung und Berufsinformation/Bereich Statistik, Austria), Michael Siegenthaler, Oskar Nordström Skans, Petri Syvanen (Ministry of Economic Affairs and Employment, Finland), Michel van Smoorenburg (UWV Public Employment Service Netherlands) for providing help to data access. I also thank Natalia Bermudez Barrezueta and Giulia Tarullo for excellent research assistance.

1 Introduction

Short-Time Work (STW) programs are government initiatives aimed at preserving employment within companies temporarily facing economic difficulties. Through STW, employees maintain employment at reduced hours, receiving a portion of their usual wage augmented by the program. Although wages during STW periods generally fall below usual salaries, they surpass typical unemployment benefits. The financial burden of supplementing income under STW is shared between the employer and the government.

STW was introduced as early as the 1920s in Germany and has been gradually adopted by other countries, particularly in continental Europe. It gained increased prominence during the Great Recession of 2008–2009, during which 25 OECD countries utilized it, and 7 introduced it among them. The very slight increase in unemployment in Germany, combined with the use of “kurzarbeit” (the German STW scheme) during the Great Recession, sparked a renewed interest in this type of intervention. This resurgence in attention is illustrated by a column by Paul Krugman published in the *New York Times* in 2009, where he wrote: In Country A, employment has fallen more than 5 %, and the unemployment rate has more than doubled. In Country B, employment has fallen only half a percent, and unemployment is only slightly higher than it was before the crisis. Don’t you think Country A might have something to learn from Country B?” The answer was a clear yes’ and Country B was Germany who “came into the Great Recession with strong employment protection legislation. This has been supplemented with a “short-time work scheme,” which provides subsidies to employers who reduce workers’ hours rather than laying them off. These measures didn’t prevent a nasty recession, but Germany got through the recession with remarkably few job losses.” In this chapter, we will see that this enthusiasm should be assessed in light of the studies that have highlighted other factors contributing to Germany’s strong performance during this period.

STW experienced another expansion in many countries during the Covid-19 pandemic, where the proportion of workers covered by STW reached historical levels, going up to 35 % at the peak of the recession, in May-June 2020, in France ([OECD, 2021](#)). Such levels were unparalleled compared to those in the past, which could be close to 5 %. However, during the Covid-19 pandemic, OECD countries exhibited a wide range of approaches to STW, highlighting questions about the efficacy of each country’s strategic decisions during this period. In particular, the response of continental European countries stood out due to their intense utilization of STW schemes. The unprecedented use of this mechanism underscores its deep integration within their strategy for managing economic crises. The situation was markedly different across the Atlantic. The United States, resorting to STW only sparingly, encountered a sharp rise in unemployment. But this surge was quickly mitigated, as the majority of the sudden spike in unemployment was due to temporary layoffs, a

significant proportion of which resulted in rehires (Hall and Kudlyak, 2022). This raises the question of the relevance of using STW.

STW programs can be justified by the inefficiencies associated with job destruction. Job destruction decisions are socially efficient if they result in the elimination of jobs whose social value is negative. The social value of a job includes not only the gains for the employer and the worker but also those for the community, particularly through the impact on public finances, on competing businesses, on crime, and on the quality of social relations. The operation of the economy is such that job destructions often have multiple reasons to be socially inefficient. Employers generally do not take into account all the externalities induced by their decisions. Financial constraints may force them to destroy jobs that could be profitable in the long term. Conversely, very stringent employment protection regulation may compel them to retain workers who could be more productive in other companies.

Evaluating the effectiveness of STW programs requires assessing to what extent, and at what cost, STW prevents socially inefficient job destruction. This evaluation reveals two distinct conceptual challenges.

The first challenge concerns assessing the gains achieved following the job destructions avoided thanks to short-time work, compared with the cost of this measure. It is necessary to consider all the gains and costs. Therefore, a comprehensive assessment of STW's effectiveness must consider direct participant benefits and broader economic effects. STW immediately relieves income loss for employed workers and helps firms conserve their specific human capital during temporary downturns and liquidity challenges. Such preservation aids companies by lowering recruitment costs upon recovery and reducing social costs associated with unemployment. It also results in benefits for society as a whole by limiting the negative consequences of unemployment. Nonetheless, STW poses potential drawbacks: the subsidization of hours not worked might inadvertently decrease work hours for employees not at risk of job loss otherwise, especially if hours of work and firms performance are challenging to verify. STW risks keeping employees tethered to low-productivity firms, curtailing better career prospects and stifling the shift of labor to more productive sectors. The program's impact also heavily depends on the institutional backdrop. For instance, if unemployment benefits are generous and permit earnings from short-term or part-time work, STW's financial relief might only be marginal.

The second challenge concerns the possibility to rely on alternative income protection and employment promotion strategies, which may be more effective than STW, depending on the circumstances. Enhancing unemployment benefits could protect against income losses while still motivating job searches in dynamic sectors. Additionally, incentivizing companies to maintain their workforce through an experience-rated unemployment insurance financing system, combined with specific layoff policies and state-backed business loans for firms in financial distress, could offer other support framework.

Despite the nearly century-old existence of STW, economists' academic output on this topic remained sparse until recently. The initial significant studies emerged in the 1980s and 1990s, aiming to empirically document STW's impact on employment stabilization during economic downturns, relying on macroeconomic or sectoral data. Concurrently, more theoretical contributions, rooted in a normative perspective, analyzed the pros and cons associated with integrating STW into unemployment insurance. The academic output significantly grew post the Great Recession. The increased reliance on STW during this period, coupled with improved data accessibility, spurred empirical studies evaluating STW's impact. The Covid-19 pandemic provided an additional boost. This article presents the contributions and discusses the limitations of this academic output.¹

[Section 2](#) is devoted to describing the main features of various STW arrangements, tracing the history of their adoption and their distinctions from other workforce retention mechanisms. [Section 3](#) presents the STW utilization rates, emphasizing the difference in STW utilization across recession periods and non-recession periods, the influence of labor market institutions, and administrations' capacity to manage highly complex arrangements. Theoretical models analyzing STW's impact and its integration with unemployment insurance are introduced in [Section 4](#). The various implications of STW, sporadically analyzed by theoretical models, will enable us to discuss STW's efficiency, considering the institutional context and the existence of other insurance mechanisms in [Section 5](#). Empirical studies dedicated to STW's impact at the macroeconomic level, company level, and individual trajectories are respectively presented in [Sections 6, 7, and 8](#). [Section 9](#) concludes by highlighting the areas that future research should explore to address the most significant gaps in our understanding of STW.

2 Overview of STW schemes

This section begins by describing the proliferation of STW schemes across various countries over time. It then notes that STW is not the only job retention scheme available, before outlining the key features of STW schemes implemented in different countries.

2.1 The spread of STW since the 1920s

STW was first implemented before World War II in Germany (*Kurzarbeit*, 1927) and Belgium (*Chômage temporaire*, 1933).² Its use spread progressively, in Italy in 1941 (*Cassa Integrazione Guadagni*), France in 1951 (*Activité*

¹ [Giupponi et al. \(2022\)](#) and [Bermudez et al. \(2023\)](#) also provide recent surveys of this literature.

² We list here the countries for which we found information on the first date of introduction of STW.

Partielle), Austria in 1968 (*Kurzarbeiterhilfe*), Japan (*Koyo Chosei Joseikin*) and Luxembourg (*Chômag partiel pour difficultés économiques conjoncturelles*) in 1975, Spain in 1980 (*Expediente Temporal de Regulación de Empleo*), Denmark in 1981 (*Arbejdssfordeling*), Canada (*Work Sharing*) and Switzerland (*Indemnité en cas de réduction de l'horaire de travail*) in 1982, the USA³ in 1992 (*Short-time compensation*) — See Table 1. The spread accelerated during the Great Recession of 2008–2009 and the COVID-19 recession. Before the 2008–2009 crisis, short-time work schemes were existing in 18 of OECD countries. By 2009, these schemes operated in 25 OECD countries, including most of the Continental European countries. Only 5 countries lacked such schemes: Australia, Greece, United-Kingdom, Iceland, Sweden. The countries which created new schemes during the crisis (usually at the end of 2008 or the beginning of 2009) are the Czech Republic, Hungary, Mexico, the Netherlands, New Zealand, Poland and the Slovak Republic.

The diffusion of STW schemes has therefore been significant. However, as we will see below, the take-up of STW is very heterogeneous within the group of countries which have introduced this type of system, particularly during periods of recession. This is largely due to the fact that there are other labor retention mechanisms.

2.2 STW and other workforce retention measures

STW differs from other schemes aimed at retaining the workforce in companies facing temporary difficulties. STW schemes differ significantly from furloughs, which are mandatory leaves of absence present in Denmark and the United Kingdom (Adams-Prassl et al., 2020) during the Covid-19 crisis. Unlike furloughs, which require workers to reduce their hours to zero, STW schemes primarily focus on adjusting work hours. They encourage employers to modify schedules rather than halt employment, even if only temporarily.

Temporary wage subsidies, used for instance in Australia, Canada, Estonia, Ireland, New Zealand and the Netherlands during the Covid-19 crisis, are another job retention scheme that differs from STW. These subsidy schemes not only subsidize hours worked but can also be used to top up the earnings of workers on reduced hours and are typically reserved for firms experiencing a significant decline in revenue. Fig. 1 provides an overview of the use of these schemes by OECD countries during the Covid-19 crisis.

2.3 The design of STW schemes

STW schemes differ significantly in their design and regulations across countries, but they also share common features. To qualify for these programs,

³The program started in 1978 in California before receiving permanent federal authorization as part of the Unemployment Compensation Amendment Act in 1992).

TABLE 1 Year of introduction of STW and furlough programs in OECD countries.

Scheme	Implementation	Year of introduction	
		Before COVID-19	During COVID-19
Short-time work	Temporary	Czech Republic (Vzdeljte se), Hungary (ESF-financed short-time working scheme), Mexico (Programa de Empleo Temporal Ampliado), the Netherlands (Deeltijd-WW), New Zealand (Job support scheme), Poland (Guaranteed employee benefits fund), Slovak Republic (Support for maintenance of employment)	Czech Republic (Antivirus B), Hungary (Job protection wage subsidy), Lithuania (Wage subsidies during idle hours), Sweden (Korttidsarbete)
Permanent		Austria (Kurzarbeiterhilfe - 1968), Belgium (Che temporaire - 1933), Canada (Work Sharing - 1982), Denmark (Arbejdssfordeling - 1981), France (Activitrielle - 1951), Germany (Kurzarbeit - 1927), Ireland (Short-TimeWork Support - 1950*), Italy (Cassa Integrazione Guadagni - 1941), Japan (Koyo Chosei Joseikin - 1975), Korea (Employment retention subsidy scheme - 1995), Luxembourg (Che partiel - 1975), Mexico (Programa de Empleo Temporal - 1995), Norway (Permittering - 1977*), Portugal (Suspensu redu temporaria da presta de trabalho - 1983) Spain (Expediente Temporal de Regulaci Empleo - 1980), Switzerland (Indemnité réduction de l'horaire de travail - 1982), US (Short-time compensation - 1992)	

Furlough	Temporary		Denmark (Midlertidig Inkompensation), Greece (Syn- ergasia), Latvia (Dikstaves pabalsts), Slovenia (Anti-Corona Law), United Kingdom (Coronavirus job retention scheme)
	Permanent	Finland (Lomautus - 2001*)	

Notes: This Table reports the year of introduction of short-time work (STW) and furlough schemes in OECD countries. The type of support is distinguished into *temporary* and *permanent*. Temporary support is defined as the limited duration of the scheme in the labor regulation. Permanent support is defined as the enduring implementation of the scheme. The existence of the program is distinguished into *Before COVID-19* and *During COVID-19*. *Before COVID-19* refers to the period before 2011. *During COVID-19* refers to the period from March to July 2020. For Mexico, the original STW program introduced in 1995 for workers of rural areas was extended in 2009 for workers of urban areas. Information *during COVID-19* for Australia, Estonia, New Zealand, Poland, and the Slovak Republic is omitted because a Wage Subsidy (WS) program was implemented during that period. For Ireland, the WS program (*Employment Wage Subsidy Scheme*) was also implemented during COVID-19. For The Netherlands, the pre-existing STW scheme was suspended during the COVID-19 period until October 2021, while a WS program was introduced (*Noodfonds Overbrugging Werkgelegenheid*). STW program's name and year of introduction in parentheses. In this notation, (*) stands for approximate year of program introduction.

Sources: National sources; [Hijzen and Venn \(2011\)](#); OECD (2020, 2021).

firms typically need to satisfy various eligibility criteria which can be related to force majeure events, technical accidents, supply problems, or economic motives. In this latter case, firms must provide evidence of economic downturn, like reduced production or sales, and show that collective agreements are in place allowing for the adoption of STW. They must also either consult with their employees or obtain individual agreements.

STW programs primarily cater to firms, rather than individual workers. Traditionally, STW offers a form of retrospective compensation to these firms. Employers seek approval to access funds from the STW program, subject to certain criteria. Once approved, employers can reduce their employees' hours and make up the difference by advancing payments from their own resources. Afterward, they can claim a reimbursement from the social security system. This approach provides companies with significant flexibility. Instead of committing in advance to specific numbers or patterns of reduced hours for their workforce, employers can make decisions based on their real-time needs and report the actual figures for reimbursement afterward.

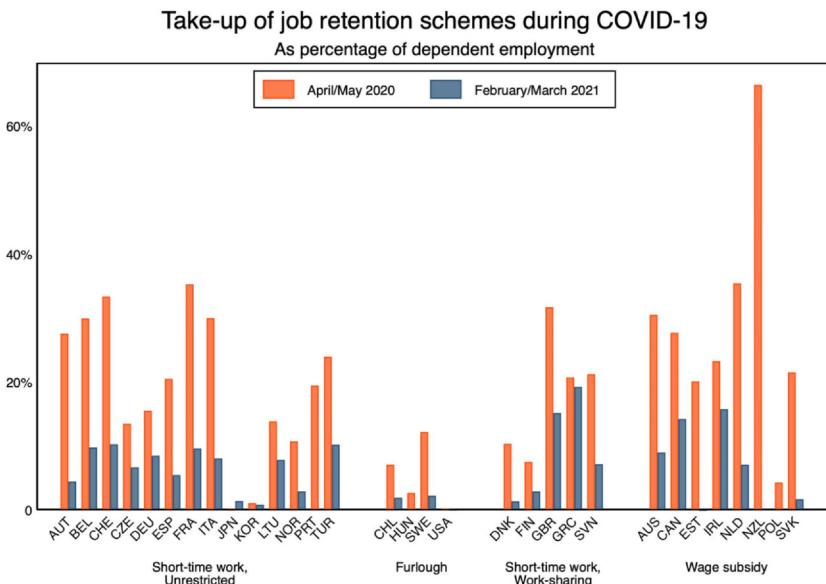


FIG. 1 OECD job retention schemes during the Covid-19 crisis. **Notes:** This figure reports the take-up of job retention schemes in OECD countries during COVID-19, each computed as the ratio of the number of beneficiaries to dependent employment. Job retention schemes comprise “Short-time work- unrestricted” (for schemes without rules setting limits on the reduction in working time); “Furlough” (if no partial reduction in working time is allowed); “Short-time work- work-sharing” (if limits are set on the maximum reduction in working time); ‘Wage subsidy’ (granted according to a drop in the wage bill or in business activity). Data on employment refers to 2020Q1. Figures for the months of April/May 2020 (orange) and for the months of February/March 2021 (blue). Source: [OECD \(2021\)](#).

Some countries, such as Denmark, Finland, Ireland, and Spain extend STW benefits to all workers, regardless of their employment status. However, in most countries, workers must have a minimum employment contribution record to qualify for STW. This condition often excludes workers on fixed-term contracts or those working limited hours. During the Great Recession and the Covid-19 crisis, many countries eased these eligibility restrictions, especially for workers on non-permanent contracts.

STW schemes come with certain conditions, typically requiring specific actions from both firms and employees. For instance, companies might have to commit to retaining their employees for a set period after the STW benefits cease. Additional conditions might include the need for employees to search for temporary or part-time jobs, the formulation of a recovery strategy by the firm, or the provision of training opportunities for employees.

The extent of working-hour reductions under STW is generally temporary and can be either total or partial, contingent upon the severity of the economic downturn. In countries like Germany, employers face fixed costs per worker

under STW, such as social security contributions, which remain constant regardless of the number of hours worked. This structure minimizes the temptation to use STW as a fully subsidized furlough scheme with total hour reductions.

Every country has set a maximum duration for STW compensation, emphasizing its inherently temporary nature. To ensure that the system is not exploited, many countries require workers and employers to cover a portion of the compensation cost for each reduced hour. This shared responsibility serves as a deterrent against misuse. As working hours decrease further from their standard level, the income generally decreases progressively in most countries. Workers can also lose social security entitlements for hours non-worked but compensated by short-time work subsidies.

[Fig. 2](#) breaks down the cost of unworked hours for employers, workers, and the government within the framework of STW and furlough schemes at the onset of the Covid-19 crisis, in May-June 2020, for workers earning an average wage. On average, the cost to workers, in the form of reduced income for unworked hours, amounts to 24 % of the labor cost. The income losses of workers are lower than those of workers compensated by unemployment insurance, even if the replacement ratios for unworked hours can be comparable, since workers continue to work a non-zero number of hours in the company. The average employer's contribution is 7 %. Overall, the government bears the majority of the cost, i.e. 69 %. However, the share covered by the government varies significantly from one country to another. It reaches 114 % in the United States, resulting in a replacement ratio for unworked hours of 121 %.⁴ The contribution of the government drops to 40 % in Japan.

3 STW take-up

This section provides an analysis of the factors influencing the adoption of STW across countries, both during and outside of recession periods. It underscores the pivotal role played by labor market institutions such as employment protection regulations and the coverage of collective agreements which are related to the design and the usage of STW.

3.1 Take-up by type of workers and firms

STW aims to protect the jobs of companies subjected to temporary shocks. As such, it particularly concerns certain types of companies and workers. Its usage also varies depending on the economic situation and the

⁴This exceptional situation reflects the pandemic policy response of the U.S. government which extended the potential benefit duration from 26 to 39 weeks and more than doubled typical benefit levels, leading most unemployed workers to receive more income from unemployment than they had from their prior jobs ([Ganong et al., 2021](#)).

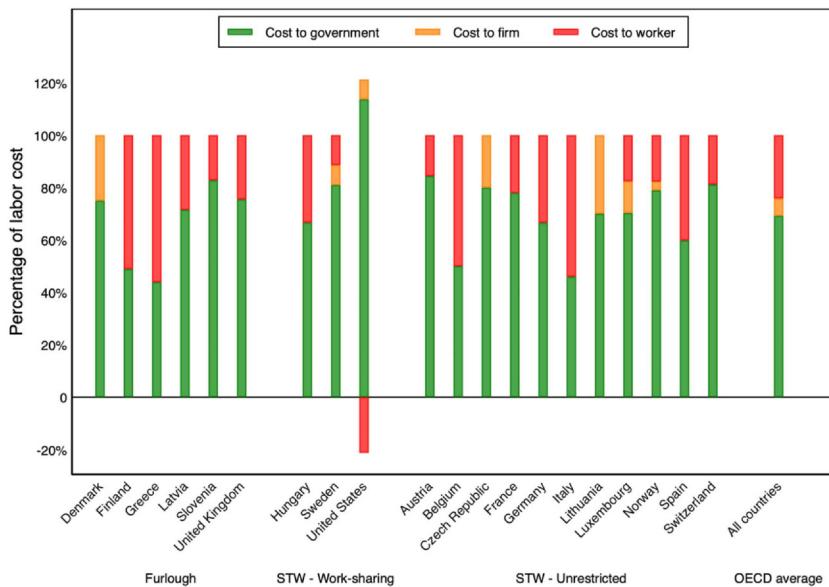


FIG. 2 The cost of hours not worked for the government, the firm, and the worker at the average wage level in STW and furlough schemes. **Notes:** This Figure presents the costs of hours not worked for the government, the firm, and the worker across OECD countries and by different types of short-time work and furlough schemes. OECD countries retained are those with a STW program in place in May/June 2020. Three types of schemes are considered: Furlough; STW – Work-sharing; STW – Unrestricted. Furlough refers to programs that only allow for a full suspension of worker's working time. STW – Work-sharing includes STW programs in which restriction on the worker's maximum reduction in working time are set. STW–Unrestricted stands for STW programs with no limits on the reduction in working time. OECD average reports the average for the selected OECD countries. Costs are computed as the percentage of the labor costs, at the average wage level, for the country-specific maximum permissible reduction in working time. Information refers to the period May/June 2020. Sources: [OECD \(2021\)](#).

institutional environment. The Wage Dynamics Network survey (WDN3), which covers a sample of about 25,000 firms in 25 European countries and ask questions about wages and employment provides interesting information about the usage of STW. WDN3 asked employers the following questions in 2014–2015: “During 2010–2013 did you need to significantly reduce your labor input or alter its composition? If YES, which of the following measures did you use to reduce your labor input or alter its composition when it was most urgent?”. Among those measures, employers can reply: “Subsidized reduction of working hours”. The responses to this survey, analyzed by ([Lydon et al., 2019](#)), show that the STW take-up rate in European countries is significant in this post-Great Recession period when some countries are still facing an economic slowdown — see Fig. 3. Over a four-year period, 8.3 % of companies and

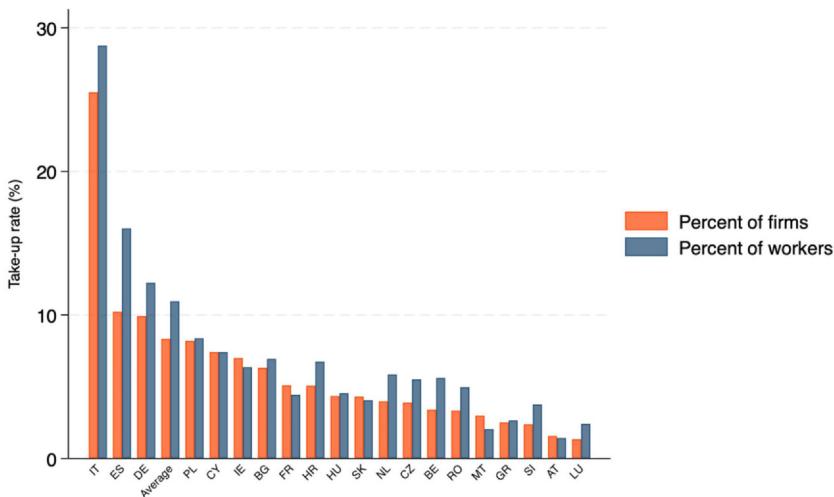


FIG. 3 STW take-up rate in European countries for the period 2010–2013 **Notes:** This figure reports take-up rates of STW as percent of firms and percent of workers for selected countries over the period 2010–2013. Information on STW use is drawn from the third wave of the Wage Dynamics and Network (WDN) survey. For Estonia, Latvia, Lithuania, Portugal, and United Kingdom, the survey does not provide information on STW take-up. The sectors of Arts, Public sector services, and Electricity, gas and water are excluded from calculations. The percent of workers refers to the equivalent in terms of employment of those firms that are in STW. The survey does not have information on STW at the worker level. Source: Wage Dynamic Network, European Central Bank.

11 % of workers used STW.⁵ These proportions exceed 25 % in Italy, which was still in recession during this period.⁶ It is much lower in countries that emerged more quickly from the recession, such as Belgium or Germany, which nevertheless have a well-established STW system.

Firms that declare that they needed to significantly reduce their labor input or alter its composition use STW much more frequently than other firms — Fig. 4. About 30 % of these firms used it over the entire 2010–2013 period. STW is therefore a variable frequently used by companies faced with the need to reduce the amount of work they use. This proportion is particularly high in Italy, where 65 % of companies declare having used it, and in Germany where this proportion reaches 43 %. The WDN3 survey shows that the use of STW has an impact on employment adjustment (Lydon et al., 2019). find that firms using STW are significantly less likely to lay off permanent workers in response to a negative shock. However, this relation between STW take-up and job separation does not hold for temporary workers. Relating the STW take-up

⁵ These figures represent averages across countries.

⁶ See Behavioral Finance & Financial Stability Global Crises Data by Country.

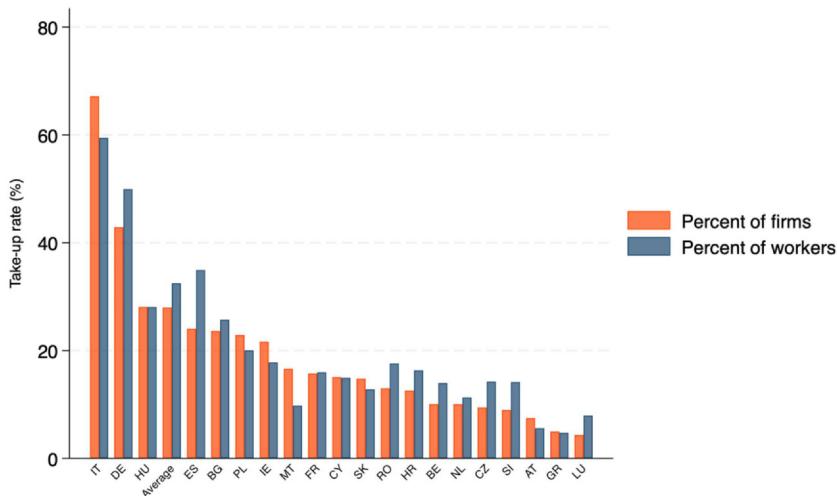


FIG. 4 STW take-up rate in European countries for the period 2010–2013 for firms which significantly reduce their labor input or alter its composition **Notes:** This figure reports take-up rates of STW as percent of firms and percent of workers for firms which significantly reduce their labor input or alter its composition for selected countries over the period 2010–2013. Information on STW use is drawn from the third wave of the Wage Dynamics and Network (WDN) survey. For Estonia, Latvia, Lithuania, Portugal, and United Kingdom, the survey does not provide information on STW take-up. The sectors of Arts, Public sector services, and Electricity, gas and water are excluded from calculations. The percent of workers refers to the equivalent in terms of employment of those firms that are in STW. The survey does not have information on STW at the worker level. Source: Wage Dynamic Network, European Central Bank.

measure in the micro data to aggregate data on employment and output trends, (Lydon et al., 2019) show that sectors with a high STW take-up exhibit significantly less cyclical variation in employment.

Companies in the manufacturing and construction sectors use STW more frequently as their activity is more fluctuating — see Table 2. Their usage rate is on average twice as high as that of the market services sectors. However, since these sectors represent a relatively small proportion of employment, the proportion of companies using STW mainly comes from the services sectors. There is no systematic relationship between firm size and the use of STW when considering countries as a whole, as shown in Table 3, but this relationship varies among individual countries, as detailed in Table 4. Notably, large firms in Belgium and Germany exhibit a higher take-up of STW, although to a lesser extent compared to other countries. In contrast, in Italy, small firms demonstrate the highest take-up rate. These variations can be attributed to the specific design of each country's STW program and the sectoral composition of firms of different sizes.

Workers who resort to STW differ depending on whether the economy is in recession. Generally, workers with stable, indefinite-term employment, or those

TABLE 2 STW take-up by firms across and within sectors.

	Percent of firms			Percent of workers	
	Not in STW	In STW	In STW (within sector)	Not in STW	In STW
Manufacturing	13.7	23.3	13.4	28.1	44.1
Construction	12.7	20.3	12.7	7.0	10.0
Trade	29.0	23.4	6.8	22.1	14.7
Business services	42.7	32.6	6.5	40.4	30.7
Financial intermediation	1.9	0.3	1.6	2.4	0.5
Total	100.0	100.0	8.3	100.0	100.0
				11.0	11.0

Notes: This table reports take-up rates of STW as percent of firms and percent of workers by sector of economic activity for selected countries over the period 2010–2013. Information on STW use is drawn from the third wave of the Wage Dynamics and Network (WDN) survey. The sectors of Arts, Public sector services, and Electricity, gas and water are excluded from calculations. The percent of workers refers to the equivalent in terms of employment of those firms that are in STW. The survey does not have information on STW at the worker level. The take-up rate is calculated by dividing the number of firms who have used STW at least once in the period by the average number of firms of the corresponding category.

TABLE 3 STW take-up by firms across and within size.

	Percent of firms			Percent of workers	
	Not in STW	In STW	In STW (within size category)	Not in STW	In STW (within size category)
Less than 5 employees	45.6	47.4	8.6	4.5	3.5
5–19 employees	36.4	33.8	7.8	15.8	10.9
20–49 employees	9.4	11.0	9.6	15.3	14.3
50–199 employees	4.0	4.1	8.4	26.1	20.9
200 employees and +	4.5	3.7	7.0	38.3	50.3
Total	100.0	100.0	8.3	100.0	100.0

Notes: This table reports take-up rates of STW as percent of firms by firm size for selected countries over the period 2010–2013. Information on STW use is drawn from the third wave of the Wage Dynamics and Network (WDN) survey. The percent of workers refers to the equivalent in terms of employment of those firms that are in STW. The survey does not have information on STW at the worker level. The take-up rate is calculated by dividing the number of firms who have used STW at least once in the period by the average number of firms of the corresponding category.

TABLE 4 STW take-up by firms by firm size.

Country	Firm size				Total
	Less than 5 employees	5–19 employees	20–49 employees	50–199 employees	
AT	0.0 %	0.5 %	0.9 %	3.2 %	1.4 %
BE		3.4 %	2.8 %	5.1 %	5.9 %
BG		5.6 %	9.8 %	2.6 %	10.4 %
CY	2.1 %	10.6 %	23.3 %	0.0 %	4.4 %
CZ		3.6 %	2.9 %	5.5 %	6.6 %
DE	9.2 %	11.2 %	11.4 %	10.6 %	14.5 %
ES		10.4 %	9.3 %	7.8 %	18.9 %
FR		2.4 %	8.9 %	3.9 %	22.3 %
GR		2.2 %	4.2 %	2.7 %	2.1 %
HR		4.4 %	9.6 %	1.8 %	8.3 %
HU		3.8 %	6.5 %	4.3 %	4.4 %
IE	5.6 %	9.4 %	6.1 %	6.3 %	2.8 %
IT		34.7 %	25.9 %	22.0 %	27.2 %
LU	0.8 %	1.4 %	3.8 %	6.0 %	0.0 %
MT		3.8 %	1.9 %	3.4 %	0.0 %

Continued

Table 4 STW take-up by firms by firm size.—Cont'd

	Firm size				Total
	Less than 5 employees	5–19 employees	20–49 employees	50–199 employees	200 employees and +
NL	3.7 %	4.1 %	6.1 %	5.4 %	4.0 %
PL	8.1 %	8.7 %	4.6 %	5.5 %	8.2 %
RO			2.7 %	4.1 %	3.3 %
SI			2.0 %	5.2 %	4.2 %
SK			4.4 %	3.4 %	4.3 %
Total	8.6 %	7.8 %	9.6 %	8.4 %	7.0 %

Notes: This table reports take-up rates of STW as percent of firms by firm size for selected countries over the period 2010–2013. Information on STW use is drawn from the third wave of the Wage Dynamics and Network (WDN) survey. The take-up rate is calculated by dividing the number of firms who have used STW at least once in the period by the average number of firms of the corresponding category.

eligible for unemployment compensation, qualify for STW during regular periods. In countries like Belgium, France, or Germany, STW allows for the retention of stable workers by reducing their working hours when activity slows down, particularly due to seasonal fluctuations. Thus, under these circumstances, STW limits the use of fixed-term employment, replacing temporary jobs with permanent ones. When major recessions, such as the Great Recession of 2008–2009 or the Covid-19 recession, occur, public authorities tend to expand STW access to workers on fixed-term contracts. This expansion was observed in countries like Finland, France, and Switzerland during the Covid-19 crisis (Giupponi et al., 2022; Hijzen and Salvatori, 2022; OECD, 2021).

3.2 Take-up in large recessions

STW, which aims to support employment in companies facing temporary declines in activity, is intended to be used more intensively during large recessions. Outside of large recessions, the fraction of the labor force using STW is below 1 % of workers in dependent employment in most countries, with the exception of Belgium and Italy — see Fig. 4. The use of STW increased significantly during the last two major recessions, reaching an unprecedented level during the Covid-19 crisis. The average annual percentage of workers who used STW during the Covid-19 crisis approached 25 % in Austria, Belgium, and the United Kingdom. Comparing this to the rates during the Great Recession of 2008–2009, which were below 1 % in Austria, amounted to 4 % in Belgium, and none in the UK, clearly highlights the exceptional nature of the Covid-19 crisis in terms of STW usage. The STW take-up rate also grew considerably during this period compared to the Great Recession in Germany, Finland, France, and Luxembourg, where STW had been in place for several decades before the Great Recession.

3.3 Take-up outside of large recessions

During recessions, it is predominantly businesses facing unforeseen and temporary drops in activity that benefit from STW. On the other hand, in regular times, it is often companies subject to anticipated seasonal fluctuations in activity that recurrently use STW. These companies usually belong to the manufacturing sector and are large. The difference in the reasons for using STW is illustrated in Fig. 5, which shows that over 60 % of the non-worked hours subsidized by STW are used by businesses that resort to it recurrently for at least three consecutive years outside of recession periods in France from 2007 to 2014. In contrast, this proportion dropped to less than 30 % during the depths of the Great Recession in 2009.

In this context, STW subsidies are concentrated on large businesses. The top 1 % of users, who are mostly systematic users, consumed over 50 % of the short-time work hours between 2006 and 2014. These major consumers were consistently active throughout the 2008–2014 period. This focus of subsidies on a limited number of players can affect STW regulations. In

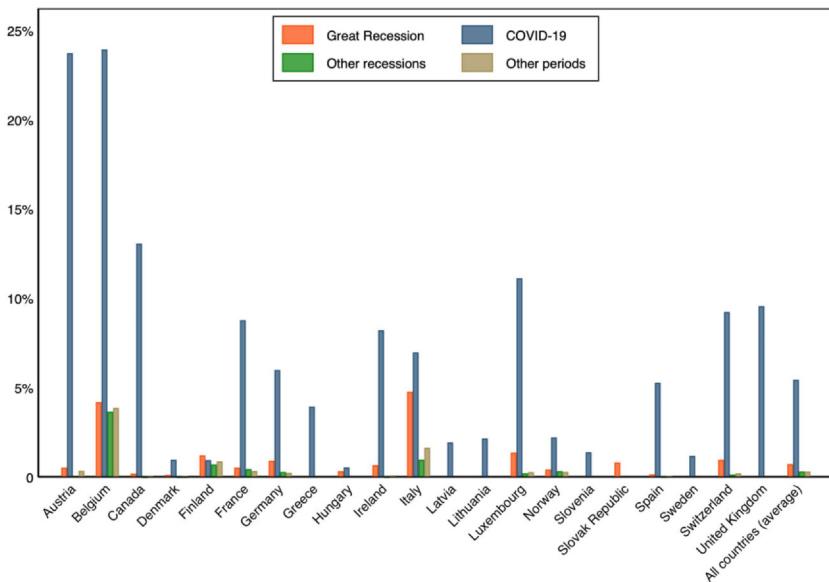


FIG. 5 STW take-up in large recessions and in other periods. Notes: This figure reports the average annual take-up rates over the period 2000–2022, depending on data availability for each country, according to four periods: (i) Great Recession, (ii) COVID-19, (iii) other crises, and (iv) other periods. The annual take-up rate is calculated by dividing the number of employees who have used STW at least once in the year by the average salaried employment for the year. The definition of each period for each country is based on the following sources: the Behavioral Finance and Financial Stability Global Crises Data by Country developed by the Harvard Business School is used for the period 2000–2017, Contraction and Expansion Indicators from Federal Reserve Bank of St. Louis for the period 2017–2019, information of stay-at-home restrictions from the Oxford COVID-19 Government Response Tracker ([Hale et al., 2021](#)) for the COVID-19 period. Country specific different sources are used in the case of lack of availability of information from the above-mentioned sources. The COVID-19 period is defined by years 2020 and 2021 for certain cases and years 2020, 2021, and 2022 for countries where restrictions were lifted later on. The Great Recession period also varies according to each country setting. Countries such as Estonia, Greece, Latvia, Lithuania, and Sweden have had a STW program only during the COVID-19 period. The Netherlands and Turkey are not included because of lack of information. Poland is not included given that the scheme, only available during the Great Recession, presents a take-up rate lower than 1%.

France, following the Great Recession of 2009, intense lobbying by sectors that greatly benefited from STW prompted the government to increase its contributions in 2012. Consequently, this reform primarily favored a small number of large firms that rely heavily on the scheme to manage their seasonal business fluctuations. The impact of corporate lobbying on increasing government contributions to STW following recession periods is also documented, such as in Germany, where STW was phased out at the end of the Great Recession.

3.4 Take-up and labor market regulation

The use of STW is influenced by labor market regulations at both the micro and macroeconomic levels. At the microeconomic level, companies are more incentivized to use STW when layoff costs are high and when renegotiations that adjust wages downward are less frequent (Lydon et al., 2019). show that companies facing higher layoff costs more often opt for an adjustment in hours by resorting to STW when their activity decreases. They also find that firms with a higher proportion of long-tenured workers, whose dismissal costs are typically high, are more likely to resort to STW: a ten percentage point increase in the share of workers with five or more years working in the firm increases the likelihood of STW take-up by 0.6 to 0.9 % points. The presence of unions in the company also promotes the use of STW. This can be related to the employer's ability to better share information with workers thanks to the union, especially for filing a request for access to STW with the administration. But it can also be due to collective agreements negotiated by unions, which limit wage adjustment and external flexibility (Biancardi et al., 2022).

At the macroeconomic level, STW schemes are often more prevalent in countries with rigorous employment protection regulations, as indicated by the OECD employment protection index — see Fig. 6 — and higher coverage of collective bargaining — see Fig. 7.

The correlation between STW, on one hand and job protection and collective bargaining coverage on the other hand, underscores a regulatory balance between internal and external flexibility. Countries that prioritize internal flexibility tend to have robust employment protection measures and high collective bargaining coverage paired with comprehensive STW schemes. In contrast, those emphasizing external flexibility typically have lax employment protection, lower collective bargaining coverage and less STW utilization. Internal flexibility is particularly beneficial for employees on permanent contracts, who are protected by employment protection legislation. In this regard, STW may contribute to increasing labor market segmentation. Like employment protection, it can partly reflects the ability of insiders to protect their jobs through regulation.

STW does not only have the effect of reinforcing the impacts of employment protection on the job stability of permanent workers. The adoption of rules favoring STW can also limit company failures in countries where employment protection is strict. Indeed, high dismissal costs that make downward employment adjustments expensive can lead companies facing a reduction in their activity to bankruptcy. This mechanism was analyzed by (Samaniego, 2006) and (Koeniger and Prat, 2007), who find that higher firing costs increase firm destruction and workforce turnover more if exiting firms default on firing costs. In this context, resorting to STW, by temporarily reducing costs, can promote company survival. This reason for resorting to STW can lead public authorities to favor STW if many companies can be in

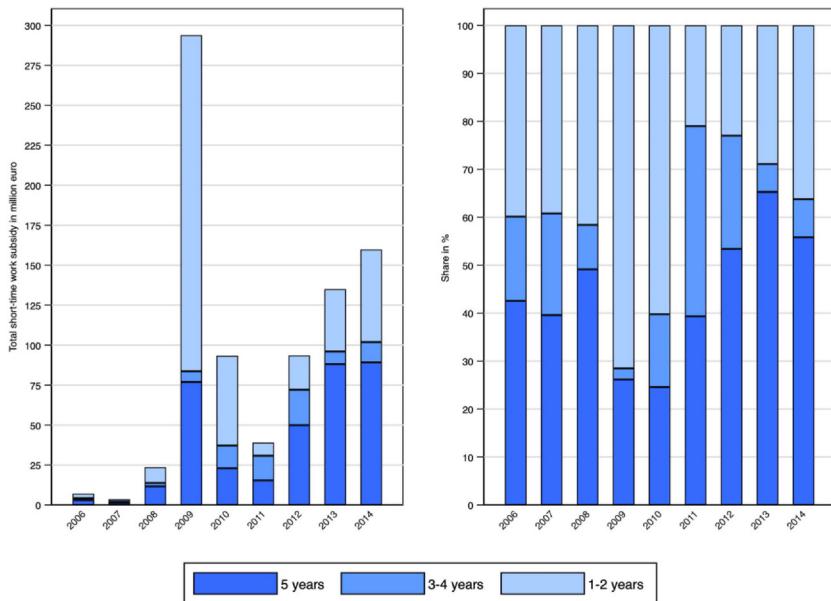


FIG. 6 Decomposition of public expenditure for STW according to the extent of regular use in France. **Notes:** This figure breaks down the public expenditure for STW according to the type of user company from 2006 to 2014 in France. “5 years” represents systematic users using short-time work every year over the last 5 years; “3 to 4 years” represents repeated users using short-time work 3 to 4 years over the last 5 years; “1 to 2 years” represents occasional users using shorttime work 1 to 2 years over the last 5 years. Source: [Cahuc and Nevoux \(2018\)](#).

default due to large-scale negative macroeconomic shocks, such as the Covid-19 pandemic.

3.5 STW design and administrative capacity

Boeri and Bruecker (2011), Cahuc and Carcillo (2011), Hijzen and Martin (2013) examine the features of STW systems correlated with its usage rate during the Great Recession using quarterly OECD data collected by (Hijzen and Venn, 2011). This dataset encompasses indicators highlighting the primary attributes of STW schemes across four essential dimensions: work-sharing, eligibility, conditionality, and generosity. The work-sharing dimension delineates the allowable reductions in weekly hours for short-time workers. Eligibility criteria define the prerequisites that both employers and workers need to fulfill to engage in STW programs. The conditionality aspect outlines behavioral expectations for all participants in STW schemes. Lastly, a programme’s generosity aspect gauges the participation cost for both companies and employees, as well as the maximum duration of their involvement.

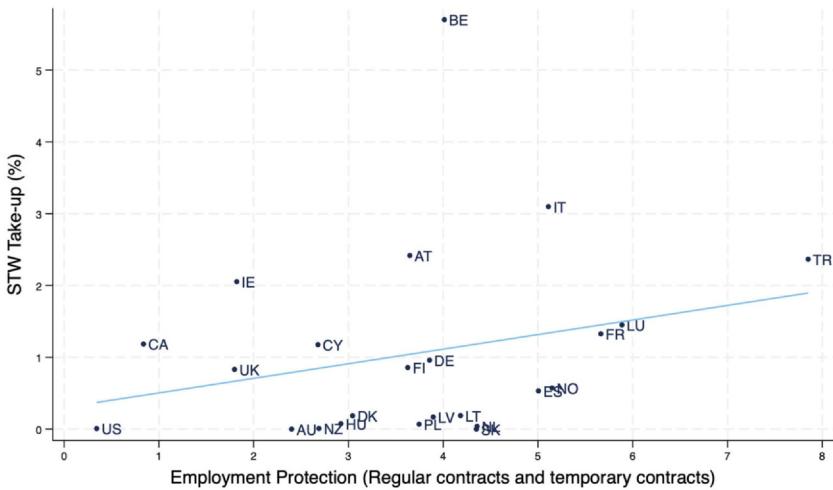


FIG. 7 STW take-up and EPL for regular and temporary contracts. **Notes:** This figure displays the relation between the index of Employment Protection and the take-up rates of STW programs for selected countries. The index of Employment Protection, sourced from OECD statistics (version 3), is equal to the sum of the indexes for regular and temporary contracts. For OECD statistics and take-up rates, values are averages over the years 2000–2022. Source: OECD for EPL and various sources for STW take-up.

The take-up is positively correlated with the permissible reductions in weekly working hours that can be compensated, with the scheme's maximum duration, and with the proportion of labor cost for reduced hours that is subsidized. There is evidence suggesting that when firms are required to share in the STW cost, the take-up's responsiveness to output shocks diminishes. Take-up rates do not seem to be influenced by stringent conditions to qualify for STW compensation, such as a commitment to retain employees for a specified period post-compensation, job search mandates, the creation of a recovery plan, or employee training. It is possible that these conditions, despite their intentions, do not significantly impact outcomes due to challenges in enforcement.

The uptake rate of STW is primarily affected by an administration's capacity to process requests from companies experiencing temporary shocks. This capability influences several factors: the processing time of company requests, the delay in reimbursement for wages paid by companies and later reimbursed by the administration—as is typical in most countries—and the timing of direct wage payments by the administration for unworked hours, which is observed in Belgium, Finland, Hungary, Norway, Spain, Turkey, and the United States (OECD, 2021).

The consequence of the administration's functioning on the approval and take-up rate is documented by (Kopp and Siegenthaler, 2021) and (Cahuc et al., 2021). They highlight the extent of the heterogeneity in the approval rate of

companies' STW applications between local administrations that are supposed to apply the same national rules during the Great Recession of 2008–2009 ([Kopp and Siegenthaler, 2021](#)). argue that the vague formulation of the eligibility criteria and the absence of clear instructions on their implementation left substantial discretionary scope for the cantonal employment agencies in Switzerland. This scope is reflected in cantonal differences in approval rates, ranging from 55 % to 100 % ([Kopp and Siegenthaler, 2021](#)). show that these differences are persistent and cannot be entirely explained by the characteristics of the establishments that apply ([Cahuc et al., 2021](#)). demonstrate that the decentralized administration of short-time work leads to significant differences in approval rates across French départements, with rates varying from 45 to 100 % in 2008. They assert that a low approval rate is indicative of poor administrative management. To gauge the quality of this management, they use an indicator of administrative inconsistency from 2007–2008. Specifically, some firms initially had their applications denied but were approved upon reapplication. The approval rate is negatively correlated with both the average number of short-time work application refusals that are later overturned and with the proportion of applications where the response time exceeds 14 workdays. Put simply, longer administrative delays and more frequent reapplications (due to initial rejections) suggest a more burdensome and less efficient administrative process ([Cahuc et al., 2021](#)). show that poor management quality decreases the STW take-up.

4 The theoretical models of STW

Early theoretical models explored the consequences of STW from a normative perspective. They aimed to assess its effectiveness in promoting a better distribution of job loss risks when job destruction is inefficient. These reflections led to mixed conclusions: while STW can correct inefficient separations, it may do so at the expense of excessively reducing working hours. Subsequent research analyzed the impact of STW on firms and workers behaviors to better understand, from a positive perspective, its effects on employment and working hours.

4.1 Normative approach

Part of the literature analyzes the effectiveness of STW from a normative standpoint within the framework of implicit contract models in which firms offer employment contracts that insure risk-averse workers against random variations in productivity ([Rosen, 1985](#)). In this context, the presence of unemployment compensation combined with the absence of contractual severance pay can lead to inefficient separations ([Burdett and Wright, 1989](#)). showed that STW can reduce these inefficient separations. However, while STW schemes correct some of the inefficiencies of the traditional unemployment insurance systems by bringing back efficient levels of employment, they are also often likely to lead to inefficient levels of working time. Indeed, the subsidy for unworked hours encourages substituting leisure for consumption,

which reduces working hours. Thus, in theory, STW has an undetermined impact on the total volume of working hours and production.

Another part of the literature analyzes the effectiveness of STW within the framework of models of optimal taxation. One potential advantage of STW for public finances is that it can help preserve jobs that are at risk of being destroyed. By conditioning subsidies on reduced working hours, STW supports firms with jobs the most at risk of destruction but with recovery potential. For this reason, STW schemes often forego a strict economic justification. When justification is provided, it tends to be informal, not anchored to a quantitative threshold, or it may sometimes mandate a minimum hours reduction (OECD, 2021; Teichgräber et al., 2022). develop a model of job retention policies in which the social planner cannot observe which jobs are truly at risk of destruction because the productivity of job is not verifiable. In this context, assuming that hours of work are verifiable, hour reductions of short-time workers act as a screening mechanism to mitigate the adverse selection problem, which provides a rationale for STW. Compared to alternatives like wage or hiring subsidies, short-time work stands out as an efficient tool to support employment. This conclusion, however, is based on the assumption that the hours worked can be perfectly verified by the administration. In reality, administrative checks are very costly, especially since the hours worked are generally reported only ex-post, which may increase moral hazard problems as employers can use STW as a wage subsidy without implementing any hour reduction (Bossler et al., 2023).

4.2 Positive approach

The literature that analyzes the impact of partial unemployment schemes on employment and working hours from a positive perspective generally fits into a dynamic context in which STW aims to preserve employment for companies facing temporary shocks (Albertini et al., 2022; Balleer et al., 2016; Cooper et al., 2017; Dengler and Gehrke, 2022; Gehrke and Hochmuth, 2021; Gehrke et al., 2019; Giupponi and Landais, 2023; Lydon et al., 2019; Tilly and Niedermayer, 2017). analyze this problem within the framework of job search and matching models à la (Mortensen and Pissarides, 1994).

The assumptions about the determination of hours and wages vary according to the various contributions (Tilly and Niedermayer, 2017). assume that firms unilaterally choose working hours and offer non-negotiable contracts that stipulate a fixed hourly wage to workers who can search for work on-the-job (Balleer et al., 2016; Dengler and Gehrke, 2022; Gehrke et al., 2019; Lydon et al., 2019). assume that firms choose working hours while wages are negotiated (Cahuc et al., 2021). assume that wages and working hours are negotiated (Albertini et al., 2022). assume that wages depend on productivity with a lower bound equal to the minimum wage, while working hours are set by employers

(Giupponi and Landais, 2023). assume that wages and hours are functions of productivity and that employers choose the number of hours worked that benefit from STW when they use it.

The main predictions of these approaches can be summarized in a simple model that represents the evolution of the value of a job, assuming that wages and working hours are negotiated. The model considers an infinite horizon and discrete time, where each working hour produces a quantity A . It is assumed that A is observed by both the firm and the worker, but not by the government, which only observes the hours of work. At the beginning of each period, A is known, but its value changes randomly between periods. The worker's preferences are represented by a function equal to the sum of the wage and the disutility of working hours, h , equal to $\phi(h)$, where ϕ is an increasing and convex function. If the worker loses his job, his expected discounted utility is equal to U . U depends notably on unemployment compensation and the probability of finding another job. The firm must pay a dismissal cost to separate from the worker, and the value of employment for the firm is null after the separation, because the job is vacant or is destroyed. Therefore, in the event of separation, the worker obtains U and the firm an expected gain equal to $-F$ where F denotes the dismissal cost, assumed to be a red-tape cost. The STW scheme allows benefiting from a subsidy σ for each hour not worked below the threshold H . This threshold often corresponds, in fact, to the usual working duration stipulated in the employment contract.

In this context, the value for the firm of a job with h working hours and productivity A , denoted by $\Pi(A, h)$, is defined by the following Bellman equation

$$\Pi(A, h) = Ah - wh + \beta \mathbb{E}[\max[\Pi(A', h), -F]]$$

where $\beta > 0$ is the discount factor, \mathbb{E} the expectation operator, and A' the value of A in the following period.

The value of this job for the worker, denoted by $W(A, h)$, satisfies:

$$W(A, h) = wh - \phi(h) + \sigma \max(H - h, 0) + \beta \mathbb{E}[\max[W(A', h), U]]$$

By definition, the surplus of this job

$$S(A, h) = W(A, h) - U + \Pi(A, h) + F$$

Using the definitions of $W(A, h)$ and $\Pi(A, h)$, the surplus of the job is defined by the following equation

$$S(A) = \max_h Ah - \phi(h) + \sigma \max(H - h, 0) - U + F + \beta \mathbb{E}[\max[S(A'), 0]] \quad (1)$$

When working hours and wages are negotiated, working hours maximize the surplus and the wage results from a surplus-sharing rule, which generally corresponds to the Nash solution to the negotiation problem. Employment is destroyed when the surplus is negative. Assuming, for illustrative purposes,

that the function ϕ is quadratic: $\phi(h) = h^2/2$, the working hours that maximize the surplus satisfy

$$h^* = \begin{cases} A & \text{if } A \geq H \\ \max(A - \sigma, 0) & \text{if } A < H \end{cases} \quad (2)$$

This equation illustrates a first result. It shows that firms and workers use STW if labor productivity is low. More precisely, they use STW for jobs where labor productivity is below a threshold that depends on H , the threshold of hours below which non-working hours are compensated by STW. In our example, Eq. (2) indicates that this productivity threshold below which it may be of interest to use STW is simply equal to H . This result is illustrated on the left panel of Fig. 9 which shows that the hours of work, h^* , increase with the productivity parameter A and are lower by the amount σ when STW is used.

The STW also has an impact on job destruction. The surplus for the optimal value of h defined by Eq. (2) is expressed as:

$$S(A) = \frac{1}{2}A^2 + \sigma \max(H - A, 0) - U + F + \beta \mathbb{E}[\max[S(A'), 0]] \quad (3)$$

This expression shows that the surplus grows with the productivity parameter A , when the subsidy σ is sufficiently small, which we will assume for the sake of realism. This implies the existence of a reservation productivity, for which the value of the surplus is equal to zero, below which jobs are destroyed.

The reservation productivity is reduced in the case of recourse to STW, since the surplus (3) is increased by the term $\sigma(H - A)$ when the unworked hours below H are subsidized, as illustrated by the left panel of Fig. 9. The following results can be deduced:

First, STW, by increasing the value of jobs, reduces their destruction rate. Destruction is reduced due to the effective use of STW in the current period for subsidized employment. It is also reduced thanks to the positive impact of STW on anticipated future gains, $\mathbb{E}[\max[S(A'), 0]]$, even in periods when STW is not used, since the reservation productivity depends on these anticipated gains. For the same reason, STW has a positive effect on job creation.

Secondly, STW reduces the working hours for all jobs whose productivity is between the reservation productivity in the case of recourse to STW – denoted R_σ in Fig. 9 –, and H . However, some of these jobs are not destroyed in the absence of STW. Indeed, jobs whose productivity is between the reservation productivity in the absence of recourse to STW (i.e. when $\sigma = 0$), denoted R , and H are not destroyed without the use of STW, because the job surplus with STW is positive for these values – see Fig. 9. Since STW can only be used if $R < H$ (otherwise it is preferable to destroy jobs whose productivity is less than H), this shows that STW can only save jobs if there is

a portion of jobs using STW that would not have been destroyed if they had not used it. Workers occupying these jobs work fewer hours than if they had not used STW. This consequence of STW has been highlighted by the first contributions dedicated to STW from a normative perspective (Burdett and Wright, 1989).

Until now, we have considered the case where H , the threshold number of hours of work below which non-worked hours are subsidized, ensures that there exists values of productivity for which it is worth using STW. However, the existence of this case depends on the values of the STW parameters, σ and H , and of the other parameters. The right panel of Fig. 8 represents the situation where it is preferable to eliminate the job rather than resort to STW when productivity is too low. In this situation, the employment surplus is negative for values of productivity A greater than H .

The expression (3) of the surplus provides information on the circumstances in which this situation can occur.

First, the reservation productivity increases with U , the worker's expected utility in the event of separation. Thus, when workers have good external opportunities, it is preferable to destroy the job rather than reduce working hours. This model therefore predicts that the use of STW is less likely when unemployment is low, as workers then have better opportunities. It also predicts that the use of STW increases during a recession not only because productivity decreases, but also because workers' external opportunities are degraded.

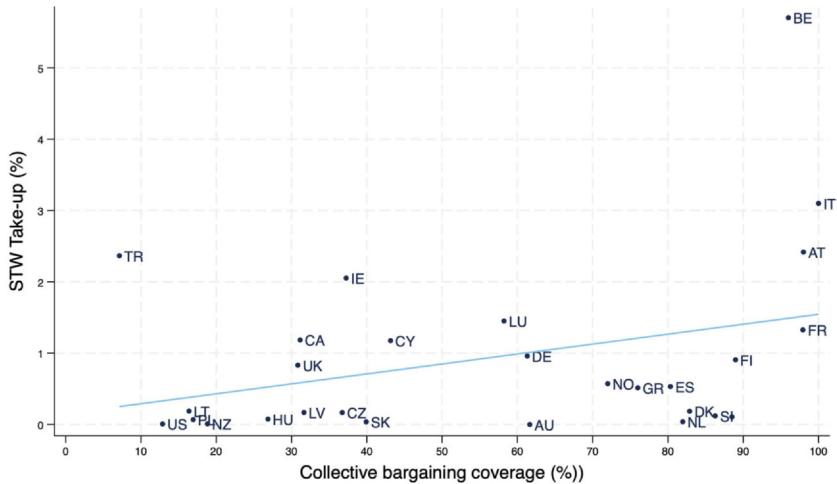


FIG. 8 STW take-up and collective bargaining coverage. **Notes:** This figure displays the relation between the collective bargaining coverage and the take-up rates of STW programs for selected countries. Values are averages over the years 2000–2022. Source: OECD for collective bargaining coverage and various sources for STW take-up.

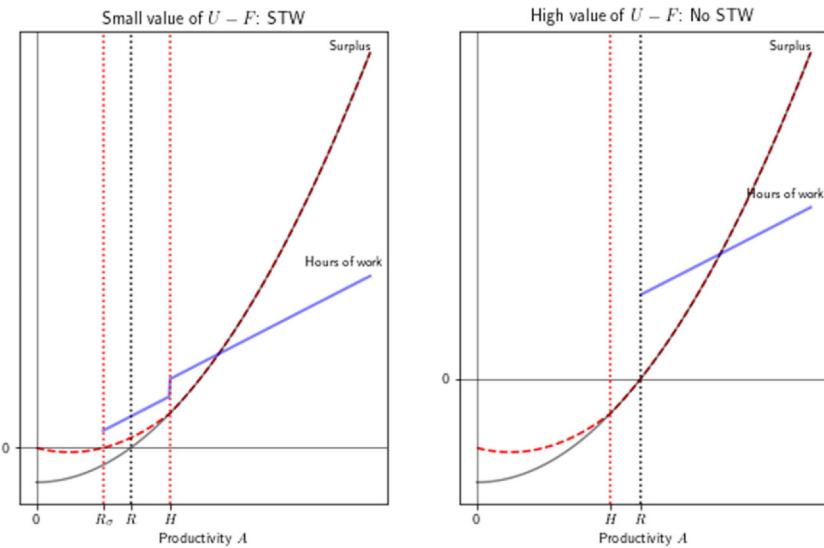


FIG. 9 Hours of work and value of jobs with and without STW use. **Notes:** This figure displays the hours of work and the value of the job surplus with and without STW use. The horizontal axis is the level of labor productivity A . The thick continuous blue line displays the hours of work. The continuous black curve represents the value of the surplus when the firm does not use STW in the current period, and the dotted red curve represents the value of the surplus when the firm can use STW. When productivity is above the threshold H , it is optimal not to use STW, and the value of both surpluses is identical. When productivity is below H , the left graph shows that the value of the surplus is higher when STW is used. This implies that the reservation productivity, below which employment is destroyed, is smaller with STW (equal to R_σ) than without STW (equal to R_0). The situation described in the left graph can exist if $U - F$ is small enough for R to be less than H , meaning if the worker's external opportunities are weak or if the dismissal costs are high. Otherwise, as represented by the right graph, it is optimal to destroy the job rather than resort to STW when A is small. This figure is constructed with $\phi(h) = h^2/2$, $\sigma=0.5$, $H=2$, $U - F - \beta\mathbb{E}[S(A')] = 1$ for the left hand side panel and equal to 3 for the right hand side panel.

Second, the reservation productivity R decreases with $\mathbb{E}[\max[S(A'), 0]]$, the expectation of future gains. This expectation of future gains is higher when negative shocks are less persistent. Consequently, this model predicts that the use of STW is more likely when negative shocks are less persistent, because in this case the reservation productivity is lower and therefore more likely to be less than H . This can be an advantage of STW, which encourages firms to retain jobs whose present value is high enough, and to destroy those whose prospects are too degraded. As we will see below (Section 5.2.2), this property of STW is often cited to justify its effectiveness relative to wage subsidies independent of working hours.

Third, the reservation productivity decreases with the dismissal costs F , which implies that the use of STW is more likely when the dismissal costs are higher. This prediction is consistent with the empirical facts, since it is

observed that firms facing higher labor adjustment costs use STW more ([Lydon et al., 2019](#)).

Search and matching models are also used to analyze the consequences of STW on job reallocation. To the extent that STW encourages companies to save jobs with low productivity, STW can slow down the reallocation of workers and capital towards more productive jobs. In search and matching models, this phenomenon arises from the fact that retaining workers in firms increases labor market tightness, which discourages job creation ([Albertini et al., 2022; Balleer et al., 2016; Cooper et al., 2017; Diaz et al., 2023; Giupponi and Landais, 2023](#)). The magnitude of this phenomenon is, however, difficult to empirically evaluate from this type of approach, as it is very sensitive to the modeling choices of the production function, preferences, and the matching function between vacant jobs and job seekers.

From the review of theoretical literature on STW, it is clear that existing knowledge is inadequate to pinpoint the key parameters that would define an optimal STW scheme coordinated with unemployment insurance. However, insights gleaned from the theoretical models, though informal, can be useful in assessing the effectiveness of STW in varied contexts.

5 The efficiency of STW

The justification of STW from an economic analysis standpoint remains an open question, with responses that are contingent upon the context. One primary reason is that it is very difficult to assess the gains induced by job destructions avoided by STW for society as a whole. Another reason is that the market failures or the inefficiencies of public interventions that STW can address can be corrected by other tools, which can be more effective. For instance, STW can be justified by the inefficiency of employment termination decisions tied to the fiscal and social externalities highlighted by ([Feldstein, 1976](#)). However, an experience rating system for unemployment insurance financing is generally more effective in achieving this goal, as it does not lead to reduced working hours ([Blanchard and Tirole, 2008; Burdett and Wright, 1989; Cahuc and Zylberberg, 2008](#)). Furthermore, the downward adjustments in working hours facilitated by STW can also be achieved through partial unemployment benefits, found in many unemployment insurance systems ([Boeri and Cahuc, 2023](#)). Company liquidity issues combined with downward wage rigidity can lead to socially inefficient job losses that could justify the use of STW ([Giupponi and Landais, 2023](#)). However, a system of state-guaranteed business loans can better target companies facing liquidity shortages. STW can be justified by its positive effects on aggregate demand ([Balleer et al., 2016](#)). But it is more efficient to transfer resources to the unemployed, who have lower incomes and whose marginal propensity to consume is higher than that of salaried workers. One potential advantage of STW for public finances is that, by conditioning subsidies on reduced working hours, STW supports firms

with jobs the most at risk of destruction but with recovery potential. STW can thus help target subsidies towards the companies facing the most temporary difficulties. However, the moral hazard arising from the possibility of declaring unworked hours, which is costly for the administration to verify, limits the effectiveness of STW (Bossler et al., 2023). Subsidies conditioned on companies' performance indicators might be more suitable for targeting those facing temporary negative shocks.

These observations indicate that the consequences of STW must be assessed based on the context, that is, the set of institutions and the other policy tools of the country in question. To understand how STW programs integrate into this analytical framework, it is beneficial to explore the determinants of the Marginal Value of Public Funds (MVPF) in the context of STW. It is an analytical framework that allows for the consideration of STW's impact on its direct beneficiaries as well as on society as a whole and is also suitable for comparing the effectiveness of different policies. The Marginal Value of Public Funds (see (Hendren, 2016) for an extensive presentation and (Hendren and Sprung-Keyser, 2020), for applications) measures the impact of expenditures dedicated to a policy on social well-being W defined as the weighted sum of individual utility functions

$$W = \sum_i \psi_i U_i$$

where ψ_i represents the social weight of individual i for policy j . The public expenditure, denoted by E_j , is a function of the policy parameters. In this context, the Marginal Value of Public Funds is equal to the ratio between the weighted sum of individual willingness to pay⁷ for policy j divided by the net public cost of the policy:

$$dW = \frac{\text{Weighted Sum of Willingness to Pay}_j}{\text{Net public cost}_j} dE_j \quad (4)$$

The net public cost is equal to the sum of the direct cost at unchanged behaviors, called the mechanical cost, and the indirect cost, called the behavioral cost, resulting from the modification of behaviors induced by the policy. In the case of STW, the behavioral effects notably include the reaction of work hours and employment. The behavioral cost can be positive if it increases spending, but also negative, if it reduces it. For example, if STW preserves many jobs, it may allow for savings thanks to the reduction of unemployment compensation expenses and the additional tax paid on the wages of the preserved jobs.

The expression (Albertini et al., 2022) of the impact of public policy on social welfare is useful for analyzing the effects of STW. It provides a measure to compare the effectiveness of different policies while distinguishing their

⁷ The willingness to pay of individual i for policy j is equal to her marginal utility associated with policy j divided by her marginal utility of income.

consequences on the well-being of individuals and public spending. The numerator illustrates the impact of STW by considering the direct effects, on those who use short-time work, and indirect effects on non-users. This assessment hinges on individual preferences, but also on social preferences, as represented by the allocation of weight to each person within the social welfare function. The denominator accounts for both the direct and indirect costs associated with STW, notably how it influences behavioral changes. Ideally, the Marginal Value of Public Funds compares the effectiveness of STW expenditure against other strategies like employment subsidies or enhanced unemployment benefits. Empirical research on STW offers limited comparisons but sheds light on its impact on individual well-being and public spending.

5.1 The social willingness to pay for STW

The social willingness to pay for STW depends on several factors which affect its impact on short-time workers but also those who do not use STW.

5.1.1 *The impact on short-time users*

People who retain their jobs thanks to STW generally have fewer income losses than those who lose them. The main reason is that STW allows workers to continue working whereas layoffs are most of the time associated with large drop in hours of work since it takes time to find a job. Job loss is also associated with a decrease in well-being, independent of the loss of income, which is reflected in a deterioration of physical and mental health. Companies that retain their employees thanks to STW avoid losing skills that contribute to improving their performance ([Tilly and Niedermayer, 2017](#)). The social willingness to pay for STW, therefore, crucially depends on the number of jobs preserved, as highlighted by the seminal contribution of ([Burdett and Wright, 1989](#)). In this regard, it is essential to consider the jobs preserved for the short-time workers. The findings from empirical studies on this subject are examined below.

It is also essential to account for the institutional context. In countries where employment protection regulation is strict, significant income shocks to companies can lead them to bankruptcy if adjusting employment is too costly ([Koeniger and Prat, 2007; Samaniego, 2006](#)). This could ultimately result in a chain of bankruptcies and significant job losses. Under these circumstances, STW can be particularly effective in cushioning the impact of major shocks on employment and working hours. The strictness of employment protection increases the willingness to pay for STW even more, as it extends the duration of unemployment spells, thus making job losses more costly for employees. A labor contract regulation that restricts the possibility of renegotiating wages and working conditions can increase the social willingness to pay for STW, just

as a financial system whose situation exacerbates companies' liquidity problems (Giroud and Mueller, 2017).

Conversely, certain contexts limit the willingness to pay because insurance mechanisms provide services similar to those of STW. For example, the combination of temporary layoffs with an experience rating system that encourages companies to rehire workers reduces the unemployment duration and the income drop following job loss (Albertini et al., 2023). The Covid-19 pandemic in the United States provides a very good illustration of this example. Temporary layoffs constituted about two-thirds of the decline in paid employment at the onset of the pandemic (Cajner et al., 2020). In this context, (Hall and Kudlyak, 2022) argue that understanding unemployment during the pandemic requires differentiating between those without jobs — 'jobless unemployment' — and those who, while still technically employed, are temporarily laid off and therefore not working — 'temporary-layoff unemployment'. The recovery rate for temporary-layoff unemployment is much swifter than for jobless unemployment. The majority of those temporarily laid off are often recalled to their jobs, bypassing the lengthy search and matching process that typically slows recovery. According to (Bell et al., 2021), among the claimants who entered the California UI system during the second quarter of 2020 and were fully separated from their employer, 51 % of those who initially expected to be recalled and 30 % of those who did not report they expected to be recalled were recalled to their prior employer by the end of 2020 (Ganong et al., 2021). find that about 75 % of unemployment exits reflected recall to a prior employer in May 2020.

The high rehiring rate for those on temporary layoff reduces the willingness to pay for STW. Moreover, the ability to combine partial unemployment benefits with income while being rehired by the same company and working fewer hours than before implies that an unemployment insurance system covering temporary layoffs and offering partial unemployment benefits provides coverage very similar to that of STW. The way unemployment insurance is financed also influences employers' incentives to dismiss their employees and to recall them when they are unemployed: (Albertini et al., 2023) estimate that the experience rating system for financing unemployment insurance in the United States significantly encourages employers to recall workers they have laid off. Therefore, the willingness to pay for STW must be assessed in light of the institutional context.

5.1.2 *The impact on short-time non-users*

It is also crucial to consider the situation of those who do not use STW, who can be affected. STW can have both positive and negative effects on the well-being of STW non-users. On one hand, by safeguarding jobs, STW can improve the well-being of non-users: the preservation of jobs for short-time users can benefit their family members and acquaintances (Britto et al., 2022; Hilger, 2016); it can reduce criminality (Draca and Machin, 2015; Fougère et

al., 2009); STW can reduce future job loss risk for non-users who foresee benefiting from it in the future as shown in the analysis of job surplus in Section 4; STW can sustain aggregate demand which can benefit to STW non-users. On the other hand, although STW is essentially a mechanism that promotes job sharing, employees with stable employment generally benefit more than other workers. STW can then increase the difficulty for workers in unstable employment and the unemployed to find a job if it results in a reduction in hiring. The impact of STW on workforce turnover can also have consequences on productivity if it promotes the retention of the workforce in low performing firms. In these cases, the potentially positive impact on the well-being of short-time users on social welfare may be offset, in the medium to long term, by the reduction in job creation and productivity (Albertini et al., 2022; Cooper et al., 2017; Tilly and Niedermayer, 2017). If such effects emerge, the social willingness to pay for STW may strongly depend on the weight of each individual in the social welfare function.

5.2 The Net public cost of STW

Examining the impact of STW on the Net public cost provides additional insight beyond its effects on the social willingness to pay, in particular on the fiscal externalities induced by the behavioral response to STW and its effectiveness at targeting jobs at risk of destruction.

5.2.1 Costs and benefits for public expenditure

First and foremost, job losses lead to social and fiscal repercussions that employers do not consider when they lay off workers. These externalities suggest that job losses are socially inefficient (Blanchard and Tirole, 2008; Cahuc and Zylberberg, 2008; Feldstein, 1976). These externalities are multifaceted, encompassing unemployment benefits, social transfers provided to unemployed workers, and the decline in tax revenue and social contributions resulting from job losses. Additionally, there are indirect costs such as heightened healthcare expenses and spikes in crime associated with rising unemployment (Britto et al., 2022; Draca and Machin, 2015; Fougerre et al., 2009). Fiscal externalities are especially significant in contemporary welfare states with strong social protection.

For these reasons, STW can reduce fiscal expenditures if it preserves jobs. In principle, STW should reduce the Net public cost if it results in a decrease in job losses and an increase in total hours worked thanks to the preservation of jobs. However, STW can also encourage a reduction in working hours, since it subsidizes non-working hours (Burdett and Wright, 1989). This behavioral effect increases the Net public cost. Therefore, in theory, STW can have both a positive and negative impact on public expenditure depending on its consequences on employment and hours worked.

For now, available studies do not provide consistent results regarding the sign of the fiscal externality due to behavioral effects, associated with ST. They

likely depend on the design of STW ([Giupponi and Landais, 2023](#)). found a positive fiscal externality in Italy, suggesting that for every euro spent on short-time work, the total cost to the government, due to behavioral responses, is around Euro 1.38. Conversely, ([Kopp and Siegenthaler, 2021](#)) identified a negative fiscal externality in Switzerland, indicating that the policy is essentially self-financing, meaning that the behavioral response does not deteriorate public finances.

5.2.2 Effectiveness at targeting vulnerable jobs

Several empirical studies on the Covid-19 recession emphasize that subsidies to firms and wage subsidies, when not accurately targeted, are less effective than STW, given that STW serves as a screening mechanism – see [Section 4](#) ([Autor et al., 2022a](#)) and ([Autor et al., 2022b](#)) analyze the consequences of the Paycheck Protection Program in the US, which distributed vast amounts of aid to small businesses to preserve jobs and provide liquidity. The estimated cost-per-job-saved by this program varies from \$169,000 to \$258,000, or to put it differently, from 3.4 to 5 times the median annual salary. They argue that this excessively high cost stems from poor targeting of the aid, and that preserving jobs and providing liquidity should be better served by addressing these issues directly and separately, thus enabling better targeting and a more progressive incidence. According to ([Autor et al., 2022b](#)), the job retention goal could be better achieved by relying on wage subsidy programs explicitly targeted to firms that had experienced declines in revenue and by STW. Other contributions share this point of view ([Smart et al., 2023](#)). evaluate the effects on workers and firms of the Canada Emergency Wage Subsidy (CEWS), which offered subsidies up to 85 % of eligible payroll costs to virtually all private-sector employers, at a total fiscal cost of five percent of GDP in the first year of the pandemic. Their findings confirm those found by ([Autor et al., 2022b](#)) in the US: much of the spending supported inframarginal jobs that would have existed in the absence of the subsidies. The estimated net wage elasticity of employment was -0.11 , implying a small aggregate employment effect of the program and an estimated fiscal cost per job saved of nearly \$200,000 per year ([Borland and Hunt, 2023](#)). study the impact of the JobKeeper program on job losses and unemployment in Australia during the Covid-19 pandemic. The JobKeeper program was a government subsidy that was paid to employers to keep employees on their payroll during the pandemic. It was available to eligible businesses that experienced a decline in revenue of at least 30 %. The subsidy covered 50 % of an employee's wages up to a maximum of \$1500 per fortnight ([Borland and Hunt, 2023](#)). find that the program was effective in preventing job losses and reducing unemployment. The cost per saved job, estimated at approximately \$100,000 annually, was less than that of the US Paycheck Protection Program. However, it is still relatively high ([Borland and](#)

Hunt, 2023). contend that with improved targeting of jobs at risk of destruction, STW could reduce this cost.

Overall, the Covid-19 pandemic has spotlighted the pressing need for effective job preservation strategies. The merits of STW as a screening tool, underscores its potential in reducing windfall effects and ensuring targeted support for the most vulnerable jobs. Yet, the effectiveness of job preservation strategies hinges on their execution, particularly the accuracy of administrative checks and the reduction of potential adverse selection and moral hazard.

The contribution of (Bossler et al., 2023) who conducted a survey to evaluate the extent of free riding behaviors associated with STW in the Covid-19 pandemic in Germany, provides interesting information on this issue. They consider three specific types of misconduct reported by employees. *i.)* Exceeding STW allowance hours: some employees on STW can work more hours than declared in their STW allowance. This means that employers can exploit unemployment insurance by having employees work hours that are technically covered by STW payments. *ii.)* Unchanged workload: employees can report no reduction in their workload despite being on STW. This contradicts the fundamental premise of STW, which is meant for situations where there is a substantial reduction in work due to unforeseen circumstances. *iii.)* Dismissal announcements during STW: workers can be informed of their impending dismissal while still on STW. Although it is legal to dismiss employees post-STW, the scheme's primary objective is job preservation. It is not intended to merely extend the duration of unemployment benefits (Bossler et al., 2023). find that 17.6 % of respondents reported working more hours than stated in their STW allowance, with estimates ranging from 14.3 % to 22.1 % depending on the methods and groups of STW workers surveyed. 38.3 % indicated an unchanged workload while on STW, with estimates fluctuating between 36.1 % and 42.2 %. 4.3 % stated that they had been informed of their job termination before beginning their STW period, with a range of 2.0 % to 10.3 %.

These results highlight the difficulty in verifying the accuracy of statements regarding working hours, which limits the effectiveness of STW, even in countries like Germany, where this scheme has been in place for nearly a century. They underscore the importance of rigorous monitoring and regulatory framework to ensure the integrity and effectiveness of STW schemes. They also suggest that STW is not always more effective than subsidies that are conditional on companies' performance indicators, especially when it comes to directing public aid towards companies at a higher risk of job losses. This likely depends on the context.

The analysis of the factors that determine the social willingness to pay and the net public cost of STW shows that the impact of STW on employment, working hours, and job reallocation plays a crucial role. The following sections present the results of empirical studies on these topics, first at the macroeconomic level, then from the firms' perspective, and finally from the workers' standpoint.

6 Effects of STW at the macroeconomic level

Studies that take a macroeconomic approach, utilizing data from various countries (as seen in works by (Abraham and Houseman, 1994); Boeri and Bruecker, 2011; Brey and Hertweck, 2020; Cahuc and Carcillo, 2011; Hijzen and Martin, 2013; Van Audenrode, 1994) or from different states within the United States (Abraham and Houseman, 2014), have generally pointed towards a favorable influence of STW on employment by allowing for adjusting hours of work instead of employment.

6.1 STW before the Great recession of 2008-2009

The seminal contribution of (Abraham and Houseman, 1994) provided the first systematic, cross-country insights into the effects of short-time schemes. They questioned the prevailing notion that the increasingly stringent job security regulations of the 1970s and 1980s in European countries were significantly impeding the adjustment of total work hours in response to unforeseen shocks. They proposed that stringent job security regulations were typically complemented by measures designed to promote alternatives to layoffs, such as work-sharing. They sought to determine the degree to which variations in working hours could provide employers with a feasible alternative to layoffs. To this end, they examined patterns of aggregate adjustment in employment and hours worked across different countries and time periods, using quarterly time-series data from Belgium, France, Germany, and the United States.

Their findings revealed that the pace of employment adjustment in response to output changes was considerably slower in the manufacturing sectors of Germany, France, and Belgium compared to the United States. However, the adjustment of total hours worked (i.e., hours multiplied by employment) seemed to be similar in these countries. The adjustment of weekly hours was quicker in Belgium, France, and Germany, where STW programs were in place.

Van Audenrode (1994) corroborates these findings through his analysis of the adjustment of hours and employment across ten OECD countries from 1969 to 1988. His study reveals that five countries, namely the United States, Belgium, Denmark, Italy, and Sweden, exhibit comparably rapid adjustments in total hours. Interestingly, in the four European nations, this swift adjustment in total hours occurs despite the pace of employment adjustments being significantly slower than in the United States (Van Audenrode, 1994). attributes this to the more generous STW schemes in these European countries compared to the United States. As a result, he deduces that generous STW schemes contribute to work flexibility and promote rapid adjustment in total hours, even in the face of firing restrictions. However, he also posits that in countries with less generous or no STW schemes, working time is not flexible enough to offset the slower employment adjustments caused by firing restrictions.

Despite this, he observes that in countries with robust job protection, overall labor adjustments manage to match the flexibility seen in the United States, as adjustments in working time compensate for firing restrictions.

6.2 STW in the Great recession of 2008-2009

Several empirical studies have confirmed the positive impact of STW on employment through an adjustment of working hours, found in previous contributions, during the Great recession of 2008–2009. Some of these studies obtain results from reduced-form econometric analyses while others are based on calibrated structural models or structural vector autoregressions. They provide interesting insights into the impact of STW depending on the economic situation, the nature of jobs saved thanks to STW, and potential issues posed by its implementation.

To overcome the selection issue associated with STW use, the studies relying on reduced-form analyses generally use indicators of past use of STW as instrumental variables for STW. However, these studies are often based on a limited set of observations, which restricts their capacity to establish a causal link between STW and employment. That said, it has been observed that STW played a role in stabilizing employment and curbing unemployment during the economic recession of 2008–2009 (as noted by (Boeri and Bruecker, 2011); Cahuc and Carcillo, 2011; Hijzen and Martin, 2013). A rise of one percentage point in STW take-up rates approximately correlates with a one percentage point drop in unemployment and a similar increase in employment. Overall, these assessments imply that STW compensation schemes played a significant role in safeguarding jobs during the economic downturn. The most substantial impacts were observed in Germany and Japan, where STW helped preserve 0.7–0.8 % of jobs during the recession.

These studies also reveal that the beneficial impact of STW on employment was confined to workers with permanent contracts (Cahuc and Carcillo, 2011; Hijzen and Martin, 2013). This, in turn, amplified the labor market segmentation, widening the divide between workers in regular employment and those in temporary positions.

Furthermore, the employment effects of STW depend on the amplitude of the recession, as shown by (Brey and Hertweck, 2020), who find that STW saves more jobs when GDP growth is deeply negative. The timing of STW is also important: (Hijzen and Martin, 2013) estimate that the persistent application of STW during the recovery phase had an adverse effect on the job-content of the recovery. Consequently, they estimate that the net effect on employment significantly declined during the recovery, and in some countries, it even turned negative. By the last quarter of 2010, the cumulative employment impact of STW since the onset of the crisis was marginally negative in Germany (0.7 %) and Italy (0.1 %), and markedly negative in Japan (1.5 %).

The results of reduced-form analyses are corroborated by structural models that take into account various factors, other than STW, that could influence employment and unemployment during the recession. This approach is particularly relevant in the case of Germany, where unemployment increased very little during the recession. The Hartz reforms and wage moderation that preceded the Great Recession could indeed have helped to mitigate the impact of the recession on employment (Dustmann et al., 2014; Rinne and Zimmermann, 2013). Moreover, (Burda and Hunt, 2011) have demonstrated that the adjustment of working hours to maintain employment primarily occurs through institutional mechanisms in Germany, such as STW, but also includes working time accounts, overtime, or regular part-time work (Gehrke et al., 2019). estimate a general equilibrium model which disentangles the role of institutions (short-time work, government spending rules) and shocks (aggregate, labor market, and policy shocks) (Balleer et al., 2016). identify the same types of factors with a structural vector autoregressive model. Both contributions confirm that STW played a significant role in preserving numerous jobs in Germany. However, the low-response of unemployment during the Great Recession cannot be solely attributed to STW. Other factors, particularly working time accounts and wage moderation, must have also been instrumental.

6.3 STW in the Covid-19 crisis

Many OECD countries used STW schemes as an instrument for securing jobs at unprecedented levels during the Covid-19 crisis. The impact of STW on employment in businesses closed for administrative reasons to halt the spread of the virus is evidently positive. From this perspective, the Covid-19 crisis is very peculiar. Nevertheless, the currently available studies that address the Covid-19 crisis from a macroeconomic perspective shed light on two aspects: the stabilizing effect of STW on aggregate demand and the extent to which STW may have excessively reduced the number of hours worked.

Dengler and Gehrke (2022) construct a New Keynesian model, incorporating incomplete asset markets and labor market frictions. This model features both an endogenous firing decision and a short-time work decision, providing a framework to analyze the impact of STW on precautionary savings. They posit that during recessions, STW diminishes the unemployment risk for workers, thereby alleviating their motive for precautionary savings and causing less of a decline in aggregate demand. Their findings suggest that this mechanism can enhance the stabilization potential of short-time work over the business cycle by up to 55 %, and even more so when monetary policy is constrained by the zero lower bound. Furthermore, they argue that an increase in the STW replacement rate can be more effective compared to an increase in the unemployment benefit replacement rate.

[Albertini et al., \(2022\)](#) build a dynamic model incorporating incomplete markets, search frictions, human capital, and both aggregate and idiosyncratic productivity shocks. This model is designed to capture the specific supply and demand effects of STW. Their study focuses on France, where STW covered up to 40 % of wage earners and where the net replacement ratio was increased to 100 % at the minimum wage and 84 % for higher wages, up to a maximum of 4.5 times the minimum wage, covering more than 95 % of wage earners. Their findings show that this highly generous STW scheme led to an increase in saving and wealth dispersion. In this context, in line with the predictions of the theoretical models discussed in [Section 4](#), they also find that, while STW stabilized employment, it induced an excessive reduction in hours worked.

6.4 The role of the timing of STW regulation and eligibility criteria

In relation to the implementation of STW, ([Brey and Hertweck, 2020](#)) emphasize that the influence of STW on the unemployment rate wanes at higher take-up rates. This highlights the importance of eligibility criteria in preventing significant deadweight losses—i.e., subsidies paid for jobs that employers would have retained in the absence of the subsidy—during economic downturns ([Brey and Hertweck, 2020](#)). also observe that the impact of STW is markedly less potent in countries with newly instituted STW schemes. The reduced effectiveness of new STW schemes may stem from the difficulties encountered by the administration in managing complex rules when it is not adequately prepared. This also suggests that workers and employers may require a period of adjustment to understand how to utilize a fresh STW scheme, or that the roll-out of certain new STW schemes may have been insufficiently timely to avert the majority of layoffs during the Great Recession.

From a related perspective, ([Balleer et al., 2016](#)) contend that while the rule-based component of STW serves as a cost-efficient safeguard for jobs, its discretionary component falls short of effectiveness to the extent that a discretionary relaxation of the STW eligibility criterion merely subsidizes jobs that would have survived even without the intervention. Their dynamic general equilibrium model, calibrated with data from Germany, provides a plausible explanation focusing on the forward-looking dimension of job value. Indeed, if discretionary interventions are temporary, firms' future expectations regarding the availability of STW support remain unchanged (i.e., the expected maximum value $\mathbb{E}[\max[S(A'), 0]]$ in the surplus [Eq. \(3\)](#) does not alter). As a result, the beneficial impact of STW on employment levels diminishes. Conversely, rules exert a direct influence on employment by impacting firms' hiring and firing decisions through future expectations ([Gehrke and Hochmuth, 2021](#)). delve deeper into this matter, examining the effects of STW across the business cycle. They employ vector autoregressive models, estimated using German data. Their findings reveal that the impacts of discretionary STW policy exhibit significant fluctuations across the business cycle and that discretionary STW is

not always ineffective. When implemented during recessions, discretionary STW bolsters employment, whereas its effect during expansions is negligible and could potentially become negative in the long term. When examining extreme events, particularly the Great Recession, the estimated effects are more pronounced and enduring. The number of jobs preserved per employee on STW due to discretionary policy reached its highest value during the Great Recession, with 0.87 jobs saved per discretionary short-time worker, but could potentially become negative during expansions, in line with the findings of (Hijzen and Martin, 2013).

7 Effects on firms

Contributions that scrutinize the impact of STW utilizing firm-level data provide a valuable supplement to macroeconomic analyses, given their access to a significantly larger pool of observations. Nonetheless, they also encounter a selection issue when attempting to identify a causal impact of STW, because STW users may inherently differ from non-users. Early studies endeavored to address this selection issue through propensity score matching, yielding results that were ambiguous and lacked robustness (Kruppe and Scholz, 2014), utilizing German data, and (Kato and Kodama, 2019), utilizing Japanese data, found no discernible effects of STW on employment. In France, (Calavrezo et al., 2010) found that establishments authorized to use short-time work were more likely to declare bankruptcy. The remainder of this section presents the results of contributions that rely on other identification strategies.

7.1 Employment and hours of work

Boeri and Bruecker (2011), using prior experience of the firm with the program in Germany as an instrument for participation in STW, found positive effects of STW on employment. This empirical strategy represents an advancement over previous studies. However, as (Bellmann et al., 2015) highlight, this identification strategy is potentially flawed since empirical evidence indicates that firms utilizing short-time work tend to adjust employment more drastically in response to output falls than firms not using STW. This behavior of STW users may be a result of technical constraints: firms are more incentivized to use short-time work if their production process implies higher costs to store production or more difficulties to find productive activities for incumbent employees when demand drops. Consequently, instrumenting program use with prior experience is likely to bias estimation of the impact of STW on employment.

Contributions that have leveraged variations in the approval rates of firms' STW applications (Cahuc et al., 2021; Kopp and Siegenthaler, 2021) or in the eligibility of firms (Biancardi et al., 2022; Giupponi and Landais, 2023) employ strategies better suited to address selection issues. They also lead to

more convergent results, highlighting the positive impact of STW on the preservation of employment during the Great Recession of 2008–2009.

Kopp and Siegenthaler (2021) employ a difference-in-differences strategy, comparing changes in outcomes for establishments in Switzerland that successfully applied for STW with those whose applications were denied. Establishments with denied applications and untreated establishments are matched to treated establishments based on nearest-neighbor propensity score matching. They document that changes in outcomes for establishments whose STW applications were denied provide a valid counterfactual for changes in outcomes for establishments whose STW applications were approved. Their findings suggest that the approval of STW leads to a cumulative reduction in permanent layoffs into unemployment of at least 10 % of an establishment's workforce three years post-application. Notably, the effect of STW on dismissals extends beyond the period during which treated establishments receive STW benefits, indicating that STW permanently prevents dismissals rather than merely postponing them. STW increased full-time equivalent employment by 9 %–17 % 4.5 years post-application. The STW program primarily preserves jobs of workers with lower educational attainment. Overall, their findings suggest that the Swiss STW scheme preserved 0.19–0.36 full-time jobs for every worker in the program. They also estimate that the program's direct fiscal benefits, which arise in the form of a reduction in spending on unemployment benefits, may have been almost large enough to offset the total fiscal spending on STW benefits.

Giupponi and Landais (2023) employ an identification strategy that leverages the variation in eligibility rules related to the industry and size of Italian firms. They find that STW has substantial and significant effects on firms' employment and working hours. Compared to their counterfactual counterparts, firms treated with STW experience a 40 % reduction in hours worked per employee, and a similar magnitude increase in the number of employees in the firm, with no apparent effect on wage rates. The employment effects are driven by a modest positive effect on inflows and a substantial negative effect on outflows. The positive employment effects are primarily driven by an increase in the number of employees on open-ended contracts. Conversely, the number of employees on fixed-term contracts experiences a negative impact. This supports the notion that STW treatment interacts with labor market duality, thereby shifting the structure of employment towards open-ended contracts. Furthermore, in contrast to the findings of (Kopp and Siegenthaler, 2021) in Switzerland, Giupponi and Landais do not observe a sustained employment impact of STW in Italy once its usage stops. This might be the consequence of the longer recession in Italy.

The beneficial impact on employment is primarily attributed to a decrease in layoffs according to (Kopp and Siegenthaler, 2021) and (Giupponi and Landais, 2023). However, Kopp and Siegenthaler find that STW also reduces hiring, while (Giupponi and Landais, 2023) find a small increase in hiring.

[Cahuc et al., \(2021\)](#) introduce a model that elucidates the conditions under which STW programs can preserve employment. Specifically, their analysis reveals that STW is effective in safeguarding jobs within firms subjected to significant negative revenue shocks. Conversely, in firms with less severe impacts, the model demonstrates that STW leads to a reduction in hours worked without necessarily preserving employment. This concept is detailed in the model discussed previously in [Section 4.2](#), highlighting that STW's ability to reduce work hours does not guarantee job preservation, especially when the firm's productivity parameter exceeds the threshold of reservation productivity in the absence of STW use. Their identification strategy, applied with French administrative data during the Great Recession, hinges on variations in local approval rate interacted with a local measure of the size of the revenue shock impacting each firm. They uncover no statistically significant positive impact of STW on total employment in 2009, with only minor positive effects appearing by 2011. STW notably curtails the total number of work hours. However, for firms grappling with the most significant revenue declines, STW exerts a positive influence on both employment and working hours. This effect was observed in 2009, during the recession, and lasted at least until 2011. These firms affected by large negative revenue shocks have been able to recover rapidly in the aftermath of the Recession, thanks to STW. STW does not preserve employment in other firms and reduces their working hours. From a related perspective, ([Tracey and Polacheck, 2020](#)), find that cyclically sensitive firms have about 14 % lower layoff rates when they use STW, but find no difference for more cyclically stable firms.

[Biancardi et al., \(2022\)](#) point to another source of heterogeneity of the impact of STW on firms, by showing that the impact of STW on Italian companies is influenced by the presence of unions. They find a greater downward sensitivity of working hours per employee to STW in highly unionized firms compared to those with low unionization. If local union density is interpreted as a proxy for union power, these results align with the notion of strong unions advocating for the use of STW as a work-sharing device to safeguard employment of incumbent workers, who are predominantly union members. These effects, however, are temporary and dissipate within two years, once the legal duration of STW schemes is reached.

Overall, these studies converge to observe that STW preserves employment, particularly in companies facing significant negative shocks. However, the results are more uncertain regarding the duration of these effects. It seems that these effects persist several years after the end of the scheme in France and Switzerland, while they disappear quickly in Italy. The same observation applies to working hours, whose evolution seems strongly linked to the context, and more specifically to the magnitude of the revenue shock ([Cahuc et al., 2021](#)) and the bargaining power of unions ([Biancardi et al., 2022](#)).

7.2 Firm productivity, profitability and firm survival

By reducing the average number of hours worked per employee within the firm, STW clearly decreases the average productivity of workers. Beyond this effect, the impact of STW on the productivity and profitability of firms depends on the balance of power between employee representatives and employers, given that collective negotiations are generally required prior to the implementation of STW. The influence of workers' representatives extends beyond the mere adoption of STW, potentially impacting firm performance as well. They may advocate for a more substantial reduction in per-capita working hours as a strategy to prevent significant employment losses. Furthermore, they may engage in negotiations over wages, the selection of short-time workers, and working conditions. This includes factors such as working hours, task organization, and other forms of internal flexibility. Although these numerous factors may influence labor productivity during periods when companies adopt STW, the available results, in Italy, do not show any impact of STW on hourly productivity ([Biancardi et al., 2022](#); [Giupponi and Landais, 2023](#)).

For the reasons we have just mentioned, the impact of STW on firm profitability depends on the institutional context. In Italy, ([Biancardi et al., 2022](#)) find that STW has a negative impact on the return on assets (ROA). They also observe that a higher union density is associated with a greater reduction in working hours and a smaller decrease in wages in companies that use STW, which should have a negative impact on their profitability. In return, higher union density can facilitate cooperation within the company, which can improve the profitability of companies that use STW. These two "faces of unions" - monopoly power and voice - highlighted by ([Freeman and Medoff, 1984](#)), may explain why ([Biancardi et al., 2022](#)) find that higher union density is not associated with a greater decrease in the profitability of companies that use STW, even though it results in higher wages and fewer working hours.

The importance of the institutional context is also highlighted in ([Kato and Kodama, 2019](#)) who argue that sharing jobs thanks to STW, by distributing the burden among employees, helps them collectively overcome adversity, thus fostering supportive interactions among colleagues. This, in turn, can strengthen the alignment of goals between workers and the firm. Goal alignment facilitates the implementation of strategic changes aimed at enhancing performance. This viewpoint suggesting that STW is more equitable because it operates as a work-sharing scheme that spreads the burden of adjustment across a larger group of workers by reducing their work hours, as opposed to situations where certain workers are abruptly let go has also been put forward, among others, by ([Abraham and Houseman, 1994; Vroman and Brusentsev, 2009; Walsh et al., 1997](#)). This holds especially true when STW is put into effect as part of solidarity agreements' that aim to prevent layoffs. The findings of ([Kato and Kodama, 2019](#)) obtained by combining propensity score matching with a difference-in-differences estimation for Japanese firms over the period

2008–2014 are consistent with this interpretation. They find that STW has a positive impact on sales and Return on Assets for companies that have implemented STW.

One can expect the effects of STW on firms' survival to be positive, given that they benefit from subsidies to adjust working hours and save on dismissal costs. However, the impact may be influenced by the quality of labor relations and the power balance between employers and employees. Indeed, STW has a positive impact on firm survival in Switzerland ([Kopp and Siegenthaler, 2021](#)) and Italy ([Giupponi and Landais, 2023; Cahuc et al., 2021](#)). do not find a significant impact in France, which may be a consequence of their sample selection that excludes all companies with less than 5 employees.

7.3 Job reallocation and productivity

Given that STW is a form of job protection that leads to workforce retention and can reduce worker reallocation from low to high productive firms, it is possible that STW decreases productivity. First, there is ample empirical evidence indicating that job protection affects the allocation of resources among industries, existing firms, and different groups of workers, as well as the intensity of firm creation and destruction. The impact of job protection on job reallocation has implications for productivity, as it affects the movement of jobs from low-productivity to high-productivity industries and firms. Empirical studies have found that the negative impact of job protection on job reallocation decreases total factor productivity ([Cahuc and Palladino, 2024](#)). Secondly, companies that adopt STW practices are generally less productive and less profitable. Therefore, STW is likely to confine workers within less productive firms.

The study conducted by ([Giupponi and Landais, 2023](#)) on Italian firms provides valuable insights into the impact of STW on the process of reallocation. Their findings indicate that STW tends to subsidize persistently low productivity matches, as less productive firms exhibit a propensity to opt for STW arrangements. Additionally, they utilized variations across local labor markets to assess the association between (exogenously) higher exposure to STW and the employment growth of high productivity firms. Notably, they observed a significant negative correlation, suggesting that in labor markets with greater accessibility to STW for low productivity firms, high productivity firms encounter greater challenges in achieving employment growth. This finding supports the notion that STW can impede the process of reallocation. However, the magnitude of the estimated effects remains relatively small despite the clear evidence of STW's influence on reallocation dynamics.

[Cooper et al., \(2017\)](#) examine the impact of STW during the Great Recession using a search and matching model that incorporates heterogeneous firms, calibrated on German data. The recession is represented by an aggregate shock that induces the least productive firms to intensify the use of STW, in

line with empirical findings. The labor retention associated with STW significantly dampens the impact of the recession on unemployment. However, it exacerbates labor market tightness, leading to recruitment difficulties and reduced job creation for firms that, on average, are more productive than those that utilized STW. The negative impact of STW on job creation persists for several years ([Cooper et al., 2017](#)). estimate that STW has reduced GDP, but this is mainly attributed to the decrease in working hours. The decrease in labor reallocation associated with STW is quantitatively marginal.

8 Effects on workers trajectories

The limited number of studies examining the impact of STW on workers' trajectories grapple with the challenge of identifying a suitable counterfactual for STW beneficiaries. The difficulty in finding a relevant counterfactual stems from the fact that generally not all workers in an establishment using STW are short-time users. Typically, the selection of workers to be included in the STW scheme is a decision made collaboratively by employers and workers' representatives. Thus, workers are subject to STW following a dual selection process: the selection of the establishment and the selection of workers within the establishment.

[Tilly and Niedermayer \(2017\)](#) exploit German administrative data over the period 2009–2011 and show that STW take-up is increasing in experience and tenure after accounting for a large number of other observables. They examine the trajectories of individuals employed full-time in January 2009 with at least six months of tenure in their current job and compare how full-time employment and earnings evolve in response to an initial transition into either STW or unemployment. The vast majority of short-time workers return to full-time work with their current employer. Short-time workers do not experience long-term effects on earnings or employment. In contrast, laid off workers experience a long-term loss in earnings and this loss is largest for workers who are experienced and have high tenure at the time of the layoff. These results are interesting. However, comparing short-time workers with those who have been laid off yields limited insights into the effects of STW, given that not all short-time workers would have faced layoffs in the absence of STW.

[Pavlopoulos and Chkalova \(2022\)](#) delve into the employment impact of STW in the Netherlands during the Great Recession (2009–2011). They compare short-time workers from firms that utilized STW with workers from firms that did not engage with STW. The selection of firms that use STW is accounted for by incorporating covariates such as the firm's revenue change, economic sector, and firm size. Their findings suggest that STW mitigated the risk of unemployment and job separation. The effectiveness of STW in safeguarding workers from unemployment was most pronounced in firms that extended the use of the program to a large number of workers for a relatively small number of hours.

[Giupponi and Landais \(2023\)](#) compare the trajectories of Italian short-time workers from firms that utilized STW, with two types of workers: i) workers from firms that were not eligible to STW, ii) laid off workers from firms that were not eligible to STW. Workers from firms not eligible for STW exhibit a similar employment probability to that of short-time workers in the first year, but higher work hours and income in the subsequent four years. This arises because firms utilizing STW are, on average, less productive than the entirety of firms not eligible for STW. The employment probability, income, and work hours of laid-off workers are lower the first year. The employment probability of short-time users and laid-off workers converges to become identical after four years. But the income and work hours of laid-off workers remain slightly lower at this horizon.

Unlike previous contributions, ([Arranz et al., 2018](#)) do not study the impact of transitioning to STW, but rather the consequences of an increase in the generosity of STW for short-time workers. In March 2009, about one year after the onset of the Great Recession, the Spanish government increased the financial incentives provided to employers and employees for the reduction of working time. The impact of this reform is analyzed using a difference-in-differences estimation, comparing the trajectories of short-time workers and non-short-time users before and after the reform. The impact on employment probability is estimated to be null in the short run (after one year) for the group of participants and negative in the medium run (after two to three years), suggesting that increasing the generosity of STW may delay workers' transition to more productive companies.

9 Conclusion

To conclude this chapter, let us reflect on the policy implications discussed earlier and tie them back to the experiences of the U.S. and Germany during the Great Recession of 2008–2009. The U.S. approach, characterized by greater labor market flexibility, led to significant job losses, while Germany's reliance on strong employment protection and the implementation of the STW scheme resulted in a more resilient labor market response. The period during the COVID-19 pandemic further highlights this difference: on the one hand, several continental European countries widely used STW, while on the other hand, countries like the United States, Canada, and Australia made little or no use of STW, opting instead to deploy wage subsidies.

This contrast underscores how different labor market policies can shape the impact of economic crises on employment, offering important lessons for future policy design.

The findings in this chapter highlight the complexity of designing effective labor market policies during economic downturns. While the U.S. model emphasizes flexibility and mobility, leading to rapid labor market adjustments, countries that rely on the STW scheme demonstrate the potential benefits of

maintaining employment relationships during periods of economic distress. Both approaches have their merits, and the right balance may depend on the specific institutional context and the severity of the economic shock.

From these observations, policymakers should recognize that the effectiveness of interventions like STW depends not only on the immediate economic context but also on broader institutional frameworks such as labor protections, wage-setting mechanisms, and the capacity of firms to endure downturns. Expanding the role of STW or similar policies could provide more stability in the face of future crises, though careful consideration is needed to avoid locking workers into low-productivity jobs.

Ultimately, this chapter shows that policy decisions must balance the trade-offs between short-term labor market stabilization and long-term economic flexibility. The analysis suggests that while different countries' experiences offer valuable lessons, neither model presents a one-size-fits-all solution.

In this context, it becomes evident that the literature's insights remain constrained in critical areas that require further research.

The interplay between STW and unemployment insurance is conceptually nebulous. The integration of STW into an optimal unemployment insurance framework *à la* (Baily, 1978), to examine the optimal distribution between unemployment benefits and STW, if possible with sufficient statistics to leverage the data, would be very relevant. However, this integration introduces substantial theoretical challenges, especially in incorporating work hours and contract termination effects. Despite these hurdles, such integration is pivotal for formulating effective policies.

Empirically, the effects of STW on employment and work hours are underexplored. Research is scant on firms and workers responses to STW scheme features. A key inquiry is the consequences of financial contributions by firms and workers. A system that extensively remunerates employees for idle hours without corporate financial input may strain public finances and subsidize otherwise sustainable jobs. On the contrary, a less generous system might have negligible impact on job loss. Enhancing our comprehension of these effects is key. Additionally, future studies should elucidate STW's influence on workforce reallocation.

References

- Abraham, K., Houseman, S., 1994. Does employment protection inhibit labor market flexibility? Lessons from Germany, France, and Belgium, in: Social Protection versus Economic Flexibility: Is There a Trade-Off? National Bureau of Economic Research, Inc, pp. 59–94.
- Abraham, K.G., Houseman, S.N., 2014. Short-time compensation as a tool to mitigate job loss? Evidence on the U.S. experience during the recent recession. Ind. Relations: J. Econ. Soc. 53 (4), 543–567.
- Adams-Prassl, A., Boneva, T., Golin, M., Rauh, C., 2020. Furloughing. Fisc. Stud. 41 (3), 591–622.

- Albertini, J., Fairise, X., Poirier, A., Terriau, A., 2022. Short-time work policies during the covid-19 pandemic. *Ann. Econ. Stat.* 146, 123–172.
- Albertini, J., Fairise, X., Terriau, A., 2023. Unemployment insurance, recalls, and experience rating. *J. Macroecon* 75, 103482.
- Arranz, J.M., García-Serrano, C., Hernanz, V., 2018. Short-time work and employment stability: evidence from a policy change. *Br. J. Ind. Relat.* 56, 189–222.
- Van Audenrode, M.A., 1994. Short-time compensation, job security, and employment contracts: evidence from selected OECD countries. *J. Political Econ.* 102 (1), 76–102.
- Autor, D., Cho, D., Crane, L.D., Goldar, M., Lutz, B., Montes, J., et al., 2022a. The \$800 billion paycheck protection program: where did the money go and why did it go there? *J. Econ. Perspect.* 36 (2), 55–80.
- Autor, D., Cho, D., Crane, L.D., Goldar, M., Lutz, B., Montes, J., et al., 2022b. An evaluation of the paycheck protection program using administrative payroll microdata. *J. Public Econ.* 211, 104664.
- Baily, M., 1978. Some aspects of optimal unemployment insurance. *J. Public Econ.* 10 (3), 379–402.
- Balleer, A., Gehrke, B., Lechthaler, W., Merkl, C., 2016. Does short-time work save jobs? A business cycle analysis. *Eur. Econ. Rev.* 84, 99–122 (European Labor Market Issues).
- Bell, A., Hedin, T.J., Schnorr, G., von Wachter, T., 2021. An analysis of unemployment insurance claims in California during the covid-19 pandemic, Policy Report 2021–19, California Policy Lab.
- Bellmann, L., Gerner, H.-D., Upward, R., 2015. The response of German establishments to the 2008–2009 economic crisis. In: Commendatore, P., Kayam, S., Kubin, I. (Eds.), *Complexity and Geographical Economics: Topics and Tools*, vol. 19. Springer, pp. 165–207.
- Bermudez, N., Dejemeppe, M., Tarullo, G., 2023. Theory and empirics of short-time work: a review. LIDAM Discussion Paper IRES, 2023(18). Funding information: We acknowledge funding from the Belgian Federal Administration Science Policy within the program “BRAIN-be 2.0 (2020–2025)” (contract No. B2/202/P3) and from the Research Foundation - Flanders (FWO) (FWO project number: G010421N).
- Biancardi, D., Lucifora, C., Origo, F., 2022. Short-time work and unionization. *Labour Econ.* 78, 102188.
- Blanchard, O.J., Tirole, J., 2008. The joint design of unemployment insurance and employment protection: a first pass. *J. Eur. Econ. Assoc.* 6 (1), 45–77.
- Boeri, T., Bruecker, H., 2011. Short-time work benefits revisited: some lessons from the Great Recession. *Econ. Policy* 26 (68), 699–765.
- Boeri, T., Cahuc, P., 2023. Labor market insurance policies in the twenty-first century. *Annu. Rev. Econ.* 15 (10), 2003–2019.
- Borland, J., Hunt, J., 2023. JobKeeper: an initial assessment. *Aust. Econ. Rev.* 56 (1), 109–123.
- Bossler, M., Osiander, C., Schmidke, J., Trappmann, M., 2023. Free riding on short-time work allowances? Results from an experimental survey design. *Kyklos* 1–20.
- Brey, B., Hertweck, M.S., 2020. The extension of short-time work schemes during the Great Recession: a story of success? *Macroeconomic Dyn.* 24 (2), 360–402.
- Britto, D.G.C., Pinotti, P., Sampaio, B., 2022. The effect of job loss and unemployment insurance on crime in Brazil. *Econometrica* 90 (4), 1393–1423.
- Burda, M., Hunt, J., 2011. What explains the German labor market miracle in the great recession? *Brook Pap. Econ. Act* 273–319.
- Burdett, K., Wright, R., 1989. Unemployment insurance and short-time compensation: the effects on layoffs, hours per worker, and wages. *J. Political Econ.* 97 (6), 1479–1496.

- Cahuc, P., Carcillo, S., 2011. Is short-time work a good method to keep unemployment down. *Nordic Econ. Policy Rev.* 1 (1), 133–165.
- Cahuc, P., Kramarz, F., Nevoux, S., 2021. The heterogeneous impact of short-time work: from saved jobs to windfall effects, Discussion paper 16168, IZA.
- Cahuc, P., Nevoux, S., 2018. The inefficiency of regular reliance on short-time work, Policy Brief 33, Institut des Politiques Publiques, Paris, France.
- Cahuc, P., Palladino, M.G., 2024. Employment protection legislation and job reallocation across sectors, firms, and workers: a survey, IZA Discussion Paper 16747, IZA—Institute of Labor Economics.
- Cahuc, P., Zylberberg, A., 2008. Optimum income taxation and layoff taxes. *J. Public Econ.* 92 (10), 2003–2019.
- Cajner, T., Crane, L.D., Decker, R.A., Hamins-Puertolas, A., Kurz, C., 2020. Tracking labor market developments during the covid-19 pandemic: a preliminary assessment, Working paper, FEDS.
- Calavrezo, O., Duhautois, R., Walkowiak, E., 2010. Short-time compensation and establishment exit: an empirical analysis with French data, IZA Discussion Papers 268, IZA Institute of Labor Economics.
- Cooper, R., Meyer, M., Schott, I., 2017. The employment and output effects of short-time work in Germany, NBER Working Paper 23688, National Bureau of Economic Research.
- Dengler, T., Gehrke, B., 2022. Short-time work and precautionary savings, CESifo Working Paper 9873, Center for Economic Studies and ifo Institute (CESifo).
- Díaz, A., Dolado, J.J., Jáñez, Á., Wellschmied, F., 2023. Labour market reallocation effects of Covid-19 policies in Spain: a tale of two recessions, IZA Discussion Papers 16095, IZA Institute of Labor Economics.
- Draca, M., Machin, S., 2015. Crime and economic incentives. *Annu. Rev. Econ.* 7 (2015), 389–408.
- Dustmann, C., Fitzenberger, B., Schönberg, U., Spitz-Oener, A., 2014. From sick man of Europe to economic superstar: Germany's resurgent economy. *J. Econ. Perspect.* 28 (1), 167–188.
- Feldstein, M., 1976. Temporary layoffs in the theory of unemployment. *J. Political Econ.* 84 (5), 937–957.
- Fougère, D., Pouget, J., Kramarz, F., 2009. Youth unemployment and crime in France. *J. Eur. Econ. Assoc.* 7 (5), 909–938.
- Freeman, R.B., Medoff, J.L., 1984. *What Do Unions Do?* Basic Books, New York, NY.
- Ganong, P., Greig, F., Noel, P., Sullivan, D.M., Vavra, J.S., 2021. Spending and job search impacts of expanded unemployment benefits: evidence from administrative micro data, Working Paper 2021–19, Becker Friedman Institute for Research In Economics.
- Gehrke, B., Hochmuth, B., 2021. Counteracting unemployment in crises: non-linear effects of short-time work policy. *Scand. J. Econ.* 123 (1), 144–183.
- Gehrke, B., Lechthaler, W., Merkl, C., 2019. The German labor market during the Great Recession: shocks and institutions. *Econ. Model* 78, 192–208.
- Giroud, X., Mueller, H.M., 2017. Firm leverage, consumer demand, and employment losses during the great recession*. *Q. J. Econ.* 132 (1), 271–316.
- Giupponi, G., Landais, C., 2023. Subsidizing labour hoarding in recessions: the employment and welfare effects of short-time work. *Rev. Econ. Stud.* 90 (4), 1963–2005.
- Giupponi, G., Landais, C., Lapeyre, A., 2022. Should we insure workers or jobs during recessions? *J. Econ. Perspect.* 36 (2), 29–54.
- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., et al., 2021. A global panel database of pandemic policies (oxford Covid-19 government response tracker). *Nat. Hum. Behav.* 5, 529–538.

- Hall, R.E., Kudlyak, M., 2022. The unemployed with jobs and without jobs. *Labour Econ.* 79, 102244.
- Hendren, N., 2016. The policy elasticity. *Tax Policy Econ.* 30.
- Hendren, N., Sprung-Keyser, B., 2020. A unified welfare analysis of government policies. *Q. J. Econ.* 135 (3), 1209–1318.
- Hijzen, A., Martin, S., 2013. The role of short-time work schemes during the global financial crisis and early recovery: a cross-country analysis. *IZA J. Labor Policy* 2 (5).
- Hijzen, A., Salvatori, A., 2022. The impact of the Covid-19 crisis across different socio-economic groups and the role of job retention schemes – the case of Switzerland, OECD Working Papers 268, OECD Social, Employment and Migration Working Papers.
- Hijzen, A., Venn, D., 2011. The role of short-time work schemes during the 2008–09 recession, Working Paper 115, OECD Social, Employment Migration Working Paper.
- Hilger, N.G., 2016. Parental job loss and children's long-term outcomes: evidence from 7 million fathers' layoffs. *Am. Econ. J.: Appl. Econ.* 8 (3), 247–283.
- Kato, T., Kodama, N., 2019. The consequences of short-time compensation: evidence from Japan, Technical report, IZA Discussion Paper, n°12596.
- Koeniger, W., Prat, J., 2007. Employment protection, product market regulation and firm selection*. *Econ. J.* 117 (521), F302–F332.
- Kopp, D., Siegenthaler, M., 2021. Short-time work and unemployment in and after the great recession. *J. Eur. Econ. Assoc.* 19 (4), 2283–2321.
- Kruppe, T., Scholz, T., 2014. Labour hoarding in Germany: employment effects of short-time work during the crises, Technical report, Discussion Paper 17, Institute for Employment Research (IAB).
- Lydon, R., Mathä, T.Y., Millard, S., 2019. Short-time work in the Great Recession: firm-level evidence from 20 EU countries. *IZA J. Labor Policy* 8 (2).
- Mortensen, D.T., Pissarides, C.A., 1994. Job creation and job destruction in the theory of unemployment. *Rev. Econ. Stud.* 61 (3), 397–415.
- OECD, 2021. Job retention schemes during the Covid-19 crisis: promoting job retention while supporting job creation. *OECD Employment Outlook 2021: Navigating the COVID-19 Crisis and Recovery*, OECD Publishing, Paris, pp. 100–150.
- Pavlopoulos, D., Chkalova, K., 2022. Short-time work: a bridge to employment security or a springboard to unemployment? *Econ. Ind. Democracy* 43 (1), 168–197.
- Rinne, U., Zimmermann, K., 2013. Is Germany the North Star of labor market policy? *IMF Econ. Rev.* 61 (4), 702–729.
- Rosen, S., 1985. Implicit contracts: a survey. *J. Econ. Lit.* 23 (3), 1144–1175.
- Samaniego, R.M., 2006. Do firing costs affect the incidence of firm bankruptcy? *Macroeconomic Dyn.* 10 (4), 467–501.
- Smart, M., Kronberg, M., Lesica, J., Leung, D., Liu, H., 2023. The employment effects of a pandemic wage subsidy, Working Paper 10218, CESifo.
- Teichgräber, J., Žužek, S., Hensel, J., 2022. Optimal short-time work: screening for jobs at risk, Working Paper 402, University of Zurich Department of Economics.
- Tilly, J., Niedermayer, K., 2017. Employment and welfare effects of short-time work, Unpublished.
- Tracey, M.R., Polacheck, S.W., 2020. Heterogeneous layoff effects of the us short-time compensation program. *Labour* 34 (4), 399–426.
- Vroman, W., Brusentsev, V., 2009. Short-time compensation as a policy to stabilize employment. University of Delaware, November 2009, manuscript.
- Walsh, S., London, R., McCanne, D., Needels, K., Nicholson, W., Kerachsky, S., 1997. Evaluation of short-time compensation programs, Technical report, Final Report for the U.S. Department of Labor, March 1997, Berkeley Planning Associates/Mathematica Policy Research Inc.

This page intentionally left blank

Chapter 6

Job search, unemployment insurance, and active labor market policies

Thomas Le Barbanchon^a, Johannes Schmieder^b, and Andrea Weber^{c,*}

^aBocconi University, CEPR, IGIER, IZA, J-PAL, ^bBoston University, NBER, CEPR, IZA,

^cCentral European University, CEPR, IZA, RF, Berlin

*Corresponding author. e-mail address: WeberA@ceu.edu

Chapter Outline

1 Introduction and background on UI	436	3.6 UI effects beyond unemployment duration	523
1.1 The origin of unemployment insurance and active labor market policy	437	3.7 Other UI design questions	530
1.2 Unemployment insurance today	439	3.8 Micro and macro effects of UI programs	534
		3.9 Discussion	539
2 Micro-foundations of job search among the unemployed	441	4 Active labor market policies	541
2.1 A brief history of job search theory	441	4.1 Meta analysis studies	542
2.2 The basic job search model	442	4.2 Lesson 1: New insights in the role of caseworkers	544
2.3 Evidence from job finding rates and re-employment wages	451	4.3 Lesson 2: More focus on programs for special groups	546
2.4 New empirical moments	459	4.4 Lesson 3: Program design takes demand side into account	550
2.5 Refining the search model	473	4.5 Lesson 4: Internationalization of ALMP use and evaluations	552
2.6 Discussion	496	4.6 Lesson 5: Advances in labor market design on online search platforms	555
3 Design of UI policy	496	4.7 Lesson 6: Growing awareness of spillover or displacement effects	558
3.1 The structure of unemployment insurance policies	497	4.8 Lesson 7: Discussion of cost effectiveness	560
3.2 The welfare effects of unemployment insurance	497	4.9 Lesson 8: Wide range of outcome variables	563
3.3 Quantification of behavioral costs	504		
3.4 Quantification of the social value of UI changes	507		
3.5 The marginal value of public funds	518		

4.10 Lesson 9: What are the mechanisms explaining program effects	568	Acknowledgments	570
4.11 Lesson 10: Novel identification strategies	568	Appendix A Supporting information	570
		References	570

1 Introduction and background on UI

Unemployment is a fundamental feature of modern labor markets. It is widely acknowledged that unemployment is caused by search frictions, because it takes time for workers to find jobs and for employers to fill vacancies, as well as structural imbalances in the labor market driven by mismatch of skills or seasonal demand, among others. Fig. 1 shows unemployment rates in 2022 among OECD countries. In this year the average unemployment rate in the OECD was 5 % but there was large variation in unemployment rates across countries. Even if the unemployment rate was just above 2 % in some countries, there is no country with zero unemployment.

In this chapter we present a comprehensive overview of how labor economists understand job search among the unemployed and how job search is shaped by unemployment insurance (UI) and active labor market policies (ALMP). The chapter is laid out in three parts: Section 2 presents the basic search model that has been the foundation of most of the literature on unemployment and job search. The section discusses key features and implications of this model and how they match several stylized facts from the empirical literature. The section focuses in particular on the recent empirical literature shedding new light on the micro-foundations of

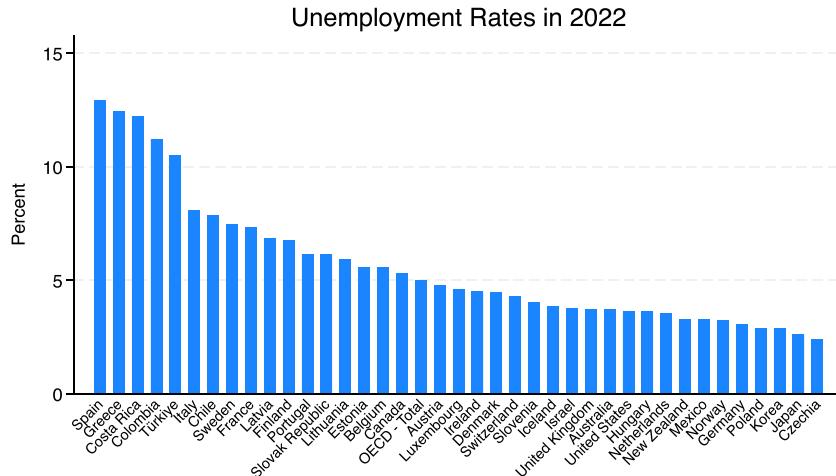


FIG. 1 Unemployment rates in 2022. Notes: The figure shows unemployment rates in 2022 for OECD countries. Source: OECD (2024), Unemployment rate (indicator). doi: 10.1787/52570002-en (Accessed on 03 July 2024).

job search and how insights from this literature have sharpened our understanding of job search and led to refinements of the search model.

We then turn to labor market policies that aim at reducing the negative consequences of unemployment. In [Section 3](#), we discuss how unemployment insurance provides benefits to unemployed workers that cover part of the loss in income. The design of unemployment insurance systems has been the subject of intensive research with many important theoretical and empirical results. We first introduce the Baily-Chetty approach to derive sufficient statistics formulas for welfare effects of altering the generosity of UI benefits, their level and potential duration. We then discuss the key results from the empirical literature that serve as the relevant inputs to these welfare formulas, with a particular emphasis on alternative ways to estimating the social value of UI. Next, we relate the Baily-Chetty formula to the concept of the Marginal Value of Public Funds. Last, we discuss how to incorporate in the design of UI effects on outcomes beyond unemployment duration (wages and separations), and macro effects. We consider a range of other policy-relevant questions related to the design of UI, such as its cyclicalities, and its time path.

In many countries, passive unemployment insurance benefit policies are complemented by active labor market policies. These policies aim at activating unemployed workers by making their job search more effective, training them to upgrade their skills, or providing subsidized employment opportunities that integrate unemployed workers in the labor market. [Section 4](#) of the chapter reviews the recent literature evaluating the effectiveness of ALMPs. We highlight exciting new developments in terms of program design, country coverage, target populations, evaluation strategies, and expansions in the comprehensiveness of evaluation studies such as cost benefit analyses and discussions of displacement effects. Overall, we summarize the new findings in this lively literature in 10 main lessons.

This chapter offers a detailed exploration of the three main topics: job search models, unemployment insurance, and active labor market policies. Each section is designed to stand independently, allowing readers to focus on any one topic in isolation. For a more comprehensive overview across all the three areas, readers may choose to focus on specific subsections in each part: for example, [Sections 2.2–2.4](#) in the job search section, [Sections 3.2–3.5](#) in the unemployment insurance section, and [Sections 4.3, 4.5–4.7](#) in the active labor market policies section.

Before turning to the main topics of this chapter, we provide some context by giving a short historical overview of the development of labor market policies and a cross-country comparison of labor market spending in the next subsection.

1.1 The origin of unemployment insurance and active labor market policy

The first systems that insured workers against job loss were established with the industrialization in the 18th and 19th century. Insurance was organized at

the city level by local guilds or trade unions and coverage was strictly restricted to members. The small scale made the financing of these early insurance systems vulnerable to large crises. In addition, coverage was very limited. Thus, municipalities, provinces, or even national governments increasingly stepped up to subsidize and organize the local systems. The adoption of national unemployment insurance systems was often triggered by large national or international shocks. The United States launched their UI system in 1935 in response to the Great Depression when the unemployment rate was above 20 %. Canada followed in 1940 and many national systems in European countries go back to the early 20th century as well.

Eventually, the variety of historical insurances developed into specific types of national insurance systems. One important type is a purely tax-financed national UI system; an example is the UK system which was established in 1911. In another type, unemployment insurance features as a component of the contribution-based social insurance system. Within this system Germany first established health insurance in 1883, followed by pension and invalidity insurance in 1889. In 1927 a national unemployment insurance was added to the system. Compared to tax-financed systems, the philosophy of social insurance systems features a stronger connection between contributions and benefits and leaves less room for redistribution. A third group of countries operate a combination of tax and contribution-financed systems, such as the US system. But there are also national systems which are not centrally organized. In four Nordic countries, Denmark, Finland, Iceland, and Sweden, unions operate unemployment insurance through government subsidized UI funds ([Holmlund and Lundborg, 1999](#)). In these countries UI membership is voluntary, while centrally organized systems typically mandate membership at least for private sector workers. In some developing countries the insurance component of UI is replaced by a savings based system.

Today unemployment insurance is available in all OECD countries and in a range of other countries. Unemployment insurance is often complemented by a severance pay (SP) system, which provides lump sum payments to displaced workers. [Gerard et al. \(2024\)](#) review UI and SP systems worldwide and report that while higher income countries are more likely to have UI, severance pay systems are more universally available across all types of countries. Savings based systems, in comparison, are rare. Comparisons by program generosity show that UI is more generous in terms of benefit levels and potential benefit durations in high income countries, while SP offers a broader coverage in low income countries. In countries with high levels of informality, coverage with UI is naturally low and accordingly the re-distributive purpose of the programs is limited.

The earliest ALMPs were developed in response to severe market failures. During periods of high and persistent unemployment positions in the public sector were created to employ workers who could not find jobs in the private sector. In a push to modernize the economy, workers with obsolete skills were

trained to acquire new skills that were in high demand. Up until the early 1970's most ALMP programs were either training programs or public sector employment programs. At this time the first job search assistance programs emerged as a low-cost alternative with the main aim of moving benefit recipients back to work. In the US public opinion changed from the view that benefit recipients needed to get jobs with the help of public sector employment programs because none were available on the private market, or that they needed to get training because their skills were obsolete to the view that benefit recipients needed to quit stalling and get to work. In the US, this development led to the 1996 Welfare reform along with a major shift in the focus of ALMPs towards job search assistance. European countries followed with some delay. By the early 2000's job search assistance programs were strongly expanded at the cost of training and especially public sector employment programs which were greatly reduced.

1.2 Unemployment insurance today

To finance labor market policies governments incur considerable fiscal costs, which are summarized in Fig. 2. The figure compares spending on passive and active labor market policies as a percentage of GDP across OECD countries in 2018, the last year with pre-pandemic data. Total spendings on labor market policies range from zero in Mexico to almost 3 % of GDP in Denmark and the OECD average is 1.1 %. The majority of countries spend a higher share of the budget on passive UI benefit policies than on active labor market policies. But there are some notable exceptions. Denmark, Sweden and Finland, countries

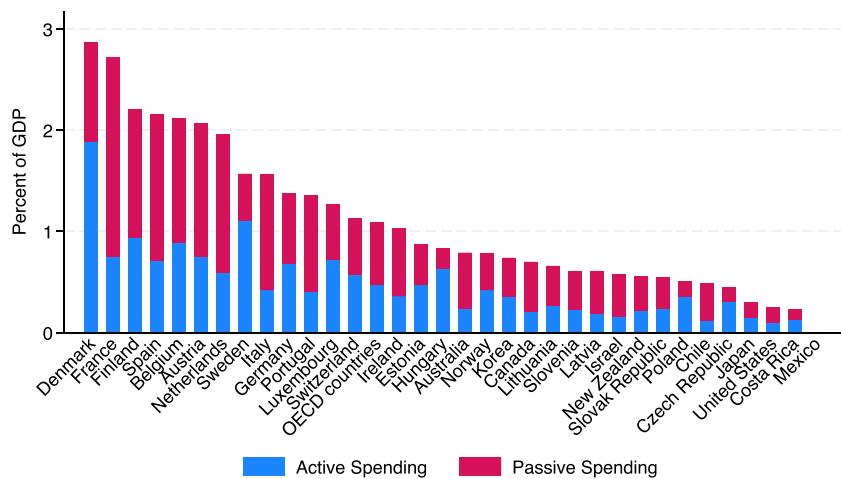


FIG. 2 Labor market policy spending in 2018. Notes: The figure shows spending on active and passive labor market policies as a share of GDP in 2018 for OECD countries. Source: OECD 2023.

with voluntary unemployment insurance systems also have the highest shares of active labor market expenditures, spending about 1 % or more of GDP or more on active programs. The expenditures by unemployed worker also vary widely across countries, which can be seen from a comparison of Figs. 1 and 2; countries with the highest unemployment rates are not necessarily those with the highest expenditures on labor market policies.

Fig. 3 shows the development of active and passive labor market spending as percentage of GDP for selected OECD countries over the last 30 years, split by region. Spending on UI benefits, shown in panels (A) and (B), tend to follow the business cycle, a pattern which is most pronounced in the US where the system features large benefit extensions during times of high unemployment. The spending paths of other countries also indicate policy shifts, for example, passive spending in Denmark, Sweden, and Germany show declines over the 30 year period. Across European country regions, spending on passive labor market policy seems to have converged to lower levels over time.

Interestingly, spending on active labor market policies shows less cyclicality or variation across countries and over time; it seems to be mostly driven by policy regimes. Generally, spending on active labor market policies is substantially lower in Anglican countries than in Europe, especially Northern and Western Europe. Sweden used to be leading in active spending in the

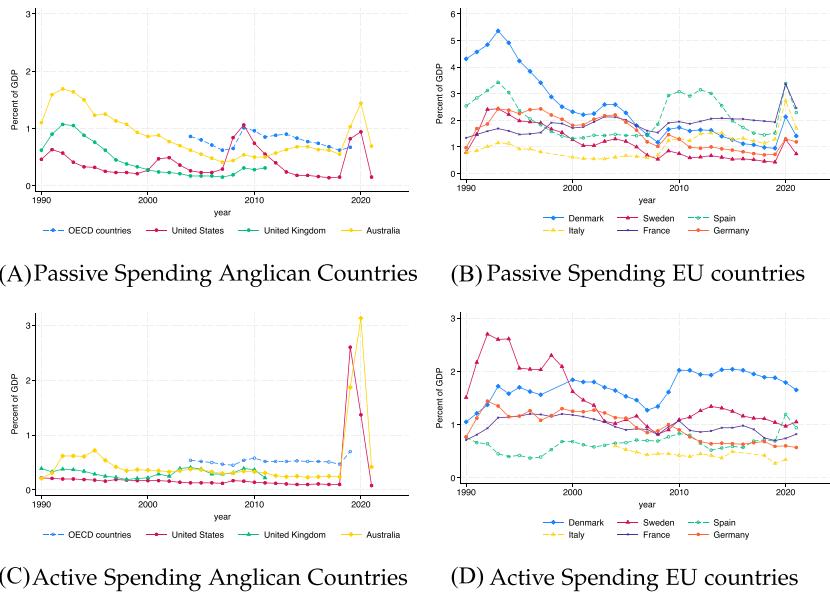


FIG. 3 Spending for passive and active labor market policy. (A) Passive spending anglican countries. (B) Passive spending EU countries. (C) Active spending anglican countries. (D) Active spending EU countries. Notes: The figure shows spending on active labor market policies for selected countries. Source OECD 2023.

1990s but has cut spending tremendously over the 2000s. Denmark, on the other hand, has slightly increased their high level of spending and is leading in Europe since the 2010s. The COVID-19 pandemic led to major disruptions in spending patterns which, however, will be temporary.

2 Micro-foundations of job search among the unemployed

This section explores recent developments seeking to understand the micro-foundations of job search. We first lay out the basic partial equilibrium job search model featuring search effort and reservation wages and describe its core predictions. We then present several stylized facts from administrative data on UI, job-finding rates, and reemployment wages that have emerged from the literature. Next, we describe new types of empirical evidence that have emerged in recent years. Based on this new evidence, we then discuss implications for our understanding of job search. We will show what this evidence reveals about the different channels in the basic search model, as well as discuss various refinements to the job search model that have been proposed in the literature and the degree to which the existing evidence supports these refinements.

2.1 A brief history of job search theory

Early analyses of the labor market relied on static demand and supply models. Labor demand in such models was derived from profit-maximizing firms, while labor supply was derived from individual utility maximization in the presence of a, often nonlinear, budget set. This modeling framework proved powerful for understanding many labor market phenomena. In particular, it provides many insights for understanding the impact of various tax and transfer programs on labor supply. However, these models also feature efficient labor markets and no involuntary unemployment, which seems at odds with obvious frictions in the labor market as well as pervasive unemployment in the actual world.

George Stigler was the first economist to develop a theory for understanding frictions in the search process in 2 seminal papers (Stigler, 1961, 1962). In Stigler's model workers have imperfect knowledge about the available jobs and have to shop around to find the best job. The question then is how many possible jobs the worker should sample given some costly search process. Since workers compete for job offers, some workers may not receive an offer and remain unemployed.

Building on this static formulation of job search several papers developed the first dynamic search models (McCall, 1970; Mortensen, 1970; Gronau, 1971). In these papers, workers receive job offers randomly drawn from a wage offer distribution. The arrival rate of offers is exogenous, so the only choice for the worker is whether or not to accept a job. These models give rise to a reservation wage, which is defined as the lowest wage a worker is willing to accept. Thus the optimal search strategy of a worker is fully described by her reservation wage at any given point in time.

The 1970s saw many refinements of this basic job search / reservation wage model. A particularly important extension was proposed by [Lippman and McCall \(1976\)](#), who extended the reservation wage model by letting workers choose how much job search effort to exert.

By the time Dale Mortensen wrote the chapter on job search for the first edition of the Handbook of Labor Economics ([Mortensen, 1986](#)), the theory of job search was quite developed and well on its way to conquering the hearts and minds of labor and macro-economists.¹

Indeed, in addition to further developments of the theory, the 1980s and 1990s job search became an important area for empirical analysis. Economists used rich new datasets to study the key predictions of the job search models. Empirically estimating job search models raised many difficult challenges. Economists had to deal with omitted variable bias, incomplete data,² and identification challenges.³

The 1980s and 1990s also saw increased interest in analyzing UI policy and some of the very first papers using administrative data in labor economics ([Katz and Meyer, 1990; Meyer, 1990; Moffitt, 1985](#)). See [Devine and Kiefer \(1993\)](#) for a good overview of this earlier empirical literature.

While in 1986, Mortensen still stated that “it is too soon for either an ‘Oscar’ or knighthood”, no one was surprised when, in 2010, Dale Mortensen and Peter Diamond were awarded the Nobel Prize in Economics for their contributions to the development of job search theory and its remarkable success in shaping modern economic thinking.

2.2 The basic job search model

In this subsection, we lay out a general version of the partial equilibrium job search model that has become the workhorse model in the study of unemployment and UI. The model focuses on the job search decision of an unemployed worker, who chooses search effort as well as whether or not to accept a job offer paying a certain wage.⁴ The model is quite flexible and allows for a changing environment throughout the unemployment spell. For example, unemployment insurance (UI) benefits may be paid only for a finite period of time or the wage offer distribution may change with unemployment duration, for instance, due to skill depreciation.

The model is set in discrete time and starts when a worker enters unemployment in period $t=0$. In each period t the worker chooses search effort τ_t .

¹ For a review of the huge success of job search theory in macro economics see [Rogerson and Shimer \(2011\)](#).

² Data typically includes incomplete unemployment spells, i.e. observations where the start of the unemployment spell is observed but the end date is censored.

³ Such as how to identify the wage offer distribution given that offers below the reservation wage are typically not observed [Flinn and Heckman \(1982\)](#).

⁴ The earliest model featuring endogenous search intensity and reservation wages is [Lippman and McCall \(1976\)](#).

The level of effort determines the probability of receiving a job offer π_t via the search production function $\phi(\cdot)$, such that: $\pi_t = \phi(\epsilon_t)$. In any given period a worker can receive at most one job offer.⁵ The cost of job search is given as $\psi_t(\epsilon_t)$. We assume that $\psi_0 = 0 = \psi'(\epsilon_0) = 0'$ and that ψ_t is convex. If a worker receives a job offer, it comes with a wage ω_t , which is drawn from a wage offer distribution with CDF: $F_t(\cdot)$. When not working, workers receive income (such as UI benefits or home production) η_t . The future is discounted at the discount factor δ .

Flow utility from consumption when unemployed is given as $u_t(\cdot)$, when employed the flow utility is given as $u_t(\cdot)$. Using different utility functions for employment and unemployment allows for differences in consumption patterns as well as potential psychological costs of unemployment. This notation is particularly common in the optimal UI literature. Both u_t and u_{t+1} are assumed to be increasing, differentiable and concave.

Once a job is accepted, we assume an individual will keep it forever. The value of accepting a job in period t that pays a wage ω_t is therefore given as:

$$v_t(\omega_t) = u_t(\omega_t) + \delta v_{t+1}(\omega_t)$$

Since the environment is constant at that point $v_t(\omega_t) = v_t(\omega_t)$ and we can simplify the value of employment to:

$$v_t(\omega_t) = \frac{u_t(\omega_t)}{1 - \delta} \quad (1)$$

The value of unemployment is given as the flow utility from UI benefits minus the cost of search effort, plus the discounted expected value of receiving a job offer:

$$u_t(\omega_t) = u_t(\omega_t) - \psi_t(\epsilon_t) - \delta \left(\int_{\epsilon_t}^{\infty} \left(u_t(\omega_t) - \psi_t(\epsilon_t) \right) d\epsilon_t \right) \quad (2)$$

Given that $v_t(\omega_t) = v_t(\omega_t)$ is monotonic, we can rewrite this value function in terms of ϵ_t as:

$$v_t(\omega_t) = u_t(\omega_t) - \psi_t(\epsilon_t) - \delta \left(\int_{\epsilon_t}^{\infty} \left(u_t(\omega_t) - \psi_t(\epsilon_t) \right) d\epsilon_t \right) \quad (3)$$

where $\psi_t(\epsilon_t)$ is the composite of the actual cost of effort and the inverse of the search production function: $\psi_t(\epsilon_t) = \phi^{-1}(\pi_t)$. This reformulation implies that the problem can be solved as if the optimization is with respect to the probability of exiting unemployment π_t . Most of the literature does not rely on data on actual search effort and therefore normalizing search effort to be equal to the job finding probability comes without loss and simplifies the

⁵ It is straightforward to model the possibility of multiple job offers per period (Mortensen, 1986) but this comes at the cost of more cumbersome notation.

notation. We will use this normalization here as well to derive the key implications of the model, but return to the more general formulation when discussing evidence on actual search effort ϕ_t in Sections 2.4 and 2.5.

The problem satisfies the so-called reservation wage property, which simply means that the value of employment is increasing in ϕ_t and therefore there is a unique wage such that all offers above it are accepted. We call this unique wage the reservation wage ϕ_t^* for jobs that start in period t . The value of unemployment can then be written as:

$$V_t = \phi_t^* - \phi_t - \delta \left(\int_{\phi_t}^{\infty} V_t dt - \int_{\phi_t^*}^{\infty} V_t dt \right) \quad (4)$$

Any wage such that $\phi_t^* \geq \phi_t$ is accepted, therefore the reservation wage ϕ_t^* is the lowest such wage and has to satisfy $\phi_t^* = \phi_t$. Using Eq. (1) we get:

$$\phi_t^* = -\delta \quad (5)$$

First Order Conditions: Given the reservation wage, the first-order condition determining optimal search effort is:

$$\phi_t' = \delta \left(\int_{\phi_t}^{\infty} V_t dt - \int_{\phi_t^*}^{\infty} V_t dt \right)$$

or

$$\phi_t' = - \left(\delta \left(\int_{\phi_t}^{\infty} V_t dt - \int_{\phi_t^*}^{\infty} V_t dt \right) \right) \quad (6)$$

Using the fact that $V_t = -\delta \phi_t$ and $\phi_t^* = -\delta$ we can write:

$$\phi_t' = - \left(\frac{\delta}{-\delta} \left(\int_{\phi_t}^{\infty} -\delta \phi_t dt - \left(\phi_t^* \right)_t \right) \right) \quad (7)$$

Given the optimal level of search effort in period t this will pin down the reservation wage in t .

Combining (5) and (4) we get an expression for the reservation wage in period t , ϕ_t^* given optimal search ϕ_t in that period and the reservation wage ϕ_t :

$$\phi_t^* = -\delta \left(\phi_t - \phi_t - \delta \left(\int_{\phi_t}^{\infty} V_t dt - \int_{\phi_t^*}^{\infty} V_t dt \right) \right) \quad (8)$$

Using Eqs. (5) and (1), we get:

$$\phi_t = -\delta \phi_t + \delta \left(\int_{\phi_t}^{\infty} \phi_t \right) \quad (9)$$

2.2.1 Steady state

Suppose that after some period we reach a stationary environment, where the wage offer distribution and the benefit level remain constant for all future periods $t \geq \tau$. In this stationary environment it has to be the case that optimal search and the reservation wage $\phi = \phi_t = \phi_\tau$ is constant. Using this, we can write the first order conditions in the steady state as:

$$= - \left(-\frac{\delta}{-\delta} \left(\int_{\phi}^{\infty} \phi \right) \right) \quad (10)$$

and

$$\phi = -\delta \phi + \delta \left(\int_{\phi}^{\infty} \phi \right)$$

We can rearrange this to:

$$\phi = -\delta \left(\int_{\phi}^{\infty} \phi \right) \quad (11)$$

Note that Eqs. (10) and (11) form a system of equations which, given the model parameters, has 2 unknowns: ϕ and ϕ_τ .

To fully solve the system, one first solves the steady state system for ϕ and ϕ_τ and then uses backwards induction to find ϕ_t and ϕ_τ for all prior periods.

2.2.2 Empirical moments: hazard rate and reemployment wage

The hazard rate λ_t in period t (that is the number of unemployment spells ending in period t conditional on being unemployed for at least t periods) is given as:

$$\lambda_t = \lambda \left(-\phi_t \right) \quad (12)$$

Denote the expected log reemployment wage of individuals who leave unemployment at the end of period t as $\bar{\phi}_t \equiv \mathbb{E}[\phi_t] \geq \phi_t$. Given the model, this is given as:

$$\bar{\phi}_t = \mathbb{E}[\phi_t] = \frac{\int_{\phi_t}^{\infty} \phi_t}{-\phi_t} \quad (13)$$

Since the hazard rate λ_t and the reemployment wage ϕ_t are empirically observable, we can compare those to direct estimates from the data. The expected unemployment duration $\bar{\tau}$ is

$$\bar{\tau} = \sum_t t$$

Where π_t is the survival function which is related to the hazard as: $\pi_t = \prod_{t=0}^T \pi_t = 1 - \lambda_t$.

2.2.3 Search effort and reservation wages throughout the unemployment spell

Single Type Consider the first-order condition for search effort in Eq. (6). How search effort evolves throughout the unemployment spell will depend on how the value of unemployment and the value of employment evolve. As an example, let's consider a case where the only source of non-stationarity is that workers exhaust UI benefits in period T . In this case, λ_t will fall throughout the unemployment spell until the UI exhaustion point. Eq. (6) implies that search effort will therefore increase until benefits are exhausted. Similarly, Eq. (8) implies that reservation wages decrease throughout the unemployment spell. A falling reservation wage and increasing search effort, both contribute to an exit hazard that is increasing throughout the unemployment spell. Finally, due to falling reservation wages, reemployment wages will fall throughout the unemployment spell.

Multiple Types These predictions of the model are for a single individual, or a homogeneous group (a ‘type’) of individuals. With heterogeneous individuals (‘multiple types’), the aggregate search effort, reservation wage and exit hazard at time t is the average value for those individuals who are still unemployed at time t . Suppose for example that there are n different types of individuals who have different hazard rates λ_{it} and reservation wages ϕ_{it} , and where the share of type i among all workers entering UI is π_{it} . In this case, the aggregate hazard is the weighted average of the type-specific hazards where the weights are the survival function multiplied with the type share:

$$\lambda_t = \sum_i \pi_{it} \frac{\lambda_{it}}{\sum_i \pi_{it}}$$

Similarly, the aggregate reemployment wage is the average reemployment wage among the people who find a job in that period and is given as:

$$\phi_t = \sum_i \pi_{it} \frac{\phi_{it}}{\sum_i \pi_{it}}$$

The aggregate hazard/reemployment wage will change throughout the spell, both because the hazard/wage of each type will change with unemployment duration, but also because the fraction of who is still unemployed will change differentially by type.

To illustrate how the aggregation of multiple types can lead to aggregate job finding hazards and reemployment wage paths that are starkly different from the within person paths, we simulate a version of the basic job search model with 2 types. The types differ by 1 parameter: a scaling parameter for the cost of job search.⁶ Type 1 has a lower cost of job search than type 2.

Fig. 4A shows the simulated search effort for the two types in an environment with potential benefit duration of $\tau = 12$ months. For both types, effort increases up to the exhaustion point and then becomes flat. Type 1 has a lower cost of job search and thus both a higher baseline effort as well as a sharper increase. Panel (B) shows the log reservation wage for the two types. For both types, reservation wages fall for the first 12 months and then stay flat after UI is exhausted. Panel (C) shows the corresponding exit hazard. In this calibration the mean of the wage offer distribution is close to the reservation wage for Type 1, so that type 1 workers reject about half of all job offers and the exit hazard is about half the search effort. By contrast, type 2's reservation wage is much lower and type 2 workers accept almost all job offers. Panel (D) shows the evolution of the log reemployment wage for the 2 types. Since type 1 has a higher reservation wage than type 2, the average accepted wage is also substantially higher. Panel (E) shows the corresponding survival rates for both types. Since type 1 workers have lower search cost, and a higher hazard, they exit the pool of unemployed quickly and the survival rate drops rapidly. Type 2 workers have a much lower exit hazard and thus stay unemployed longer. Correspondingly, Panel (F) shows how the type shares among the still-unemployed change throughout the spell. While both types initially make up 50 % of the unemployed, the share of type 1 workers falls fast, while type 2 workers make up an ever lasting share of the remaining unemployed. These changes in the type shares are the key driver for the aggregate hazard rate shown in Panel (G): Initially the aggregate hazard rate is pulled up by the high job finding rates of type 1 workers. As these low-cost workers find jobs, the high cost type 2 workers remain and the aggregate hazard approaches the low hazard of type 2. The aggregate hazard also exhibits a small spike at the UI exhaustion point ($t = 12$). The spike is mostly driven by type 1 workers who still make up about 20 % of the unemployed at the exhaustion point and show a sharply increasing exit hazard leading up to UI exhaustion. Panel (F) shows the aggregate reemployment wage. While reemployment wages fall within both types, the aggregate reemployment wage falls much faster as the type 1 workers, who generate many job offers and only take high paying jobs, exit earlier and the pool of unemployed consists increasingly of the high cost type 2 workers who accept even very low paying jobs.

Note that the aggregate hazard thus shows a very different time path than the type-specific hazard rates. This shows how dynamic selection, that is

⁶ We assume a log utility function $U(c) = \ln(c)$, a log normal wage offer distribution $f(w) = \frac{1}{w\sigma}\exp\left(-\frac{(w-\mu)^2}{2\sigma^2}\right)$ and a cost of job search function $C(s) = \frac{\gamma s}{2}$.

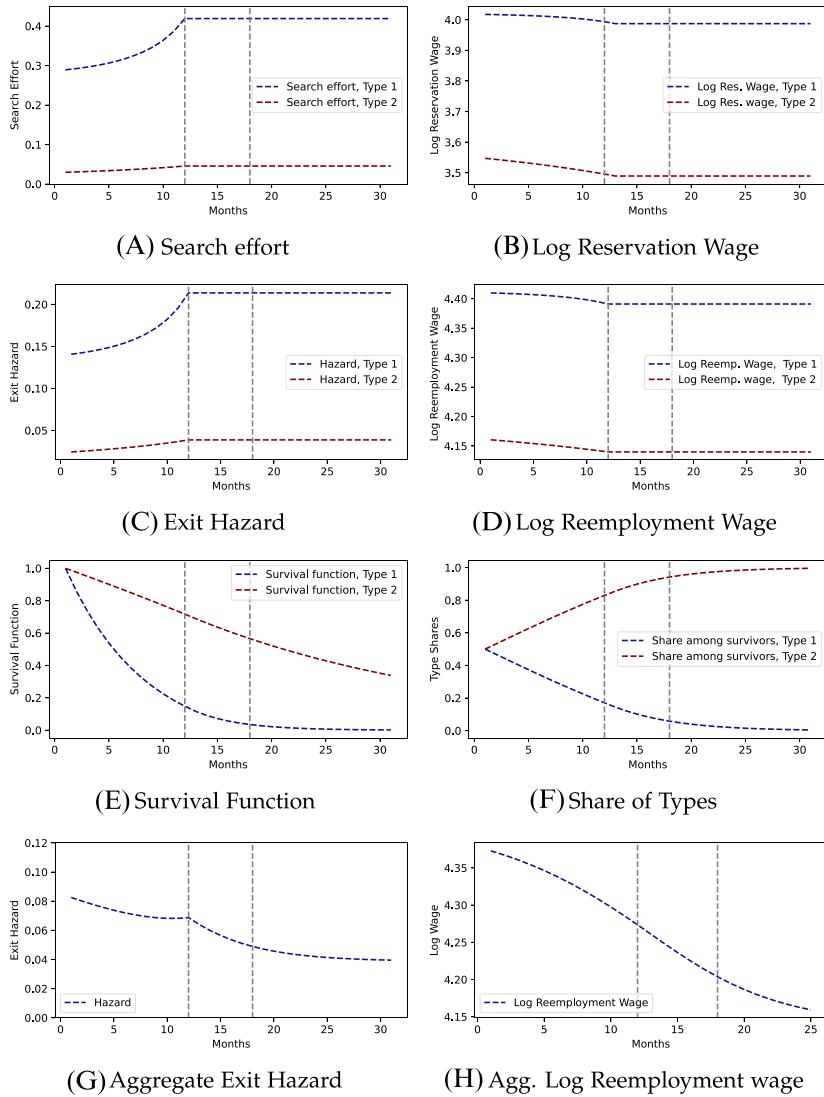


FIG. 4 Simulation of basic job search model with 2 types. (A) Search effort. (B) Log reservation wage. (C) Exit hazard. (D) Log reemployment wage. (E) Survival function. (F) Share of types. (G) Aggregate exit hazard. (H) Agg. log reemployment wage. Notes: The figure shows simulations of the basic model with 2 types. The types differ by the cost of job search.

changes in the composition of workers who remain unemployed throughout the spell, can have a first-order effect on observed hazard rates. Similarly, the expected reemployment wage path of individual worker types may be very different than the aggregate reemployment wage path.

2.2.4 The effects of UI on job finding and reemployment wages

Next, we consider what the model predicts for the effects of UI on job search outcomes.

Static Environment: The static search effort and reservation wage are solutions of the system of two Eqs. (10) and (11). First, Eq. (10) defines search effort as a function of the reservation wage without any dependence on the UI benefit . Recall that Eq. (10) writes (omitting the subscript):

$$\phi' = \frac{\delta}{-\delta} \left(\int_{\phi}^{\infty} - \phi \right) \quad (14)$$

Standard differentiation of the above equation (using Leibniz rule) yields:

$$\begin{aligned} \frac{\partial}{\partial \phi} &= \frac{\delta}{-\delta} \int_{\phi}^{\infty} \frac{\partial}{\partial \phi} - \phi \\ &= \frac{-\phi'}{\frac{\delta}{-\delta}} - \phi \\ &= 0 \end{aligned} \quad (15)$$

where the last inequality stems from the convexity of the composite cost function and from positive marginal utility.

We can now differentiate the second Eq. (11) to obtain the effect of UI benefits on reservation wage. Recall that Eq. (11) writes:

$$\phi = - \frac{\delta}{-\delta} \left(\int_{\phi}^{\infty} - \phi \right) \quad (16)$$

The differentiation is simplified as the contribution of changes in search effort () disappears because of the first order condition on search effort. We obtain:

$$\phi' \phi = \phi' - \frac{\delta}{-\delta} \phi' - \phi \phi$$

After some manipulation, it writes:

$$\frac{\partial \phi}{\partial} = - \frac{\phi'}{\phi' - \frac{\delta}{-\delta} \phi' - \phi} = 0$$

Putting the results together we get:

$$\frac{\partial \phi}{\partial \phi} = \frac{\partial \phi}{\partial} = 0$$

Thus higher UI benefits increase the reservation wage and decrease search effort. Combined with Eq. (12) we get that

$$= 0 \quad (17)$$

In a static environment, the expected duration of an unemployment spell can be written as: $\frac{1}{\phi} = \frac{1}{\lambda} - \frac{\mu}{\lambda}$ and therefore $\frac{1}{\phi} - \frac{\mu}{\lambda} > 0$, i.e. more generous UI benefits lead to longer unemployment durations.

Non-stationary Environment: To derive the comparative statics in the non-stationary environment, consider Eq. (6). Consider an increase in UI benefits ΔP . Clearly, this does not affect the value of employment but raises the value of unemployment $\frac{\partial u_t}{\partial P}$. Furthermore for all t , an increase in $\frac{\partial u_t}{\partial P}$ will decrease search effort so that $\frac{\partial s_t}{\partial P} < 0$. Similarly Eq. (4) implies that $\frac{\partial r_t}{\partial P} < 0$. Since both the decrease in search effort and increase in the reservation wage lead to longer unemployment durations we have $\frac{\partial d_t}{\partial P} > 0$. Similarly one can derive $\frac{\partial \lambda_t}{\partial P} > 0$.

To illustrate the model predictions for the effects of UI, we simulate an extension of potential benefit durations in the model above. Fig. 5 shows how job search responds when PBD is extended from 12 to 18 months. Panel (A)

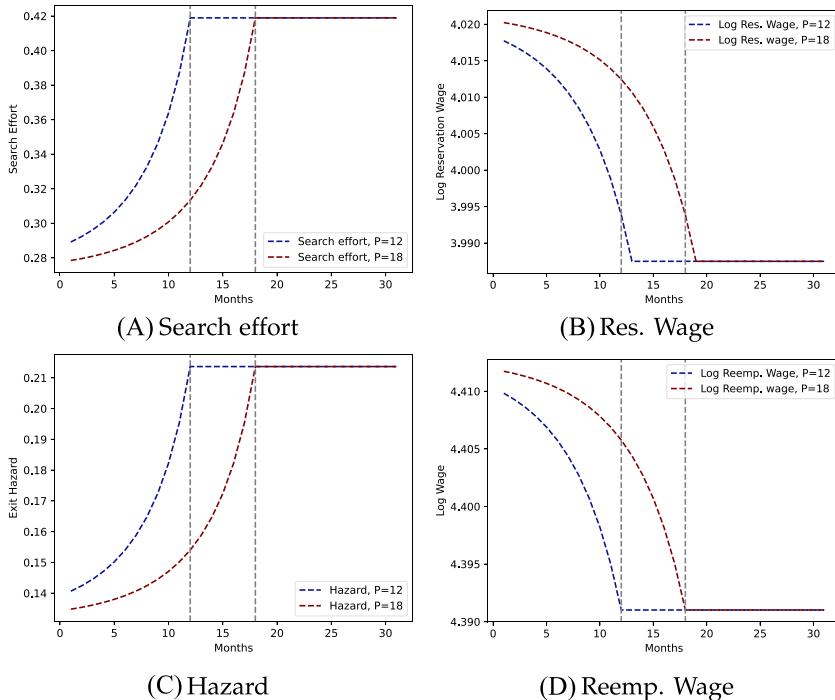


FIG. 5 Simulating the effects of a UI extension. (A) Search effort. (B) Res. wage. (C) Hazard. (D) Reemp. wage. Notes: The figure shows simulations of the model in 8 when changing the potential UI benefit duration from 12 to 18 months. Panels (A) to (D) show how search effort, the reservation wage, the exit hazard and the reemployment wage change for Type 1 workers. Panels (E) and (F) show how the aggregate exit hazard and reemployment wage respond to the PBD increase. Panels (E) and (F) also overlay the empirical exit hazards and reemployment wages for Germany from Schmieder et al. (2016), which were used to calibrate the model here.

shows how search effort evolves for a single type of worker. Increasing PBD from 12 to 18 months shifts the search effort path to the right, so that search effort is lower at every duration less than 18 months. On the flip side, the reservation wage increases at all durations until 18th months. Both the change in the reservation wage and the search effort lead to a lower hazard rate under the more generous UI regime until benefit exhaustion (Panel C), while the reemployment wage path is shifted upwards (Panel D).

To summarize, the search model makes a few key predictions. Within homogenous types, an increase in the generosity of UI benefits decreases the job finding rate and weakly increases reservation and reemployment wages conditional on unemployment duration.

Importantly while the effect of UI on nonemployment durations is clearly positive, the effect on average reemployment wages is less clear. On the one hand more generous UI leads to higher reservation wages, but on the other hand it leads to, on average, longer unemployment durations and thus workers finding jobs later in the spell, when reemployment wages are lower on average, either due to lower reservation wages or changes in the wage offer distribution throughout the unemployment spell. In the appendix section A we show that if the exhaustion of UI benefits is the only source of non-stationarity, then the positive effect of higher reservation wages dominates the effect of longer unemployment durations and average reemployment wages are strictly increasing in PBD. However, suppose there are other sources of non-stationarity, such as skill depreciation that leads to lower wage offers in later periods. In that case, this latter channel may dominate and lead to a negative effect of PBD on reemployment wages (see, for example, Schmieder et al., 2016).

When aggregating over multiple types, the model predicts that more generous UI increases nonemployment duration, but dynamic selection means that the job finding hazard (reemployment wages) do not necessarily decrease (increase) for all unemployment durations since the type composition is changing.

2.3 Evidence from job finding rates and re-employment wages

Over the two decades, the literature has documented a set of stylized findings with respect to job search outcomes and how they relate to UI benefits. These findings rely on a combination of high quality administrative datasets, with clean and transparent empirical strategies to estimate the effects of UI on these outcomes. Here we will lay out these stylized findings, highlighting a few selected examples from the literature. The examples focus on regression discontinuity designs, which are straightforward to understand and provide transparent visual evidence, but by no means do they represent and exhaustive list.

We focus here on highlighting some stylized qualitative findings, without discussing the magnitudes of the effects. We will return to the magnitudes and provide a more systematic overview in [Section 3](#), where we discuss the effects of UI policy and their implications for UI policy design.

2.3.1 The effects of UI on unemployment duration

As discussed above, the one robust prediction of the basic job search model, even with heterogeneous types is that an increase in UI generosity leads to longer unemployment durations. While this had been tested and confirmed in empirical work since the 1970s, modern evidence based on administrative data and regression discontinuity designs made this point extremely convincingly and clear. Fig. 6 shows four examples of papers that estimated the effect of a UI benefit extension on nonemployment durations. Panel (A) and (B) provide evidence from Austria and represent, to our knowledge, the 2 first RD designs in the UI literature. Card et al. (2007) shows the effect of a PBD increase from 20 to 30 weeks at a work history cutoff (number of months employed in previous 5 years) and documents a clear jump in nonemployment duration by about 7 days. Lalive (2008) shows the effect of a particularly large increase in PBD from 39 to 209 weeks at an age cutoff (age 50) and a corresponding doubling of unemployment durations. Panel (C) shows the effect of increasing PBD from 12 to 18 months at an age 42 cutoff and from 18 to 22 months at an

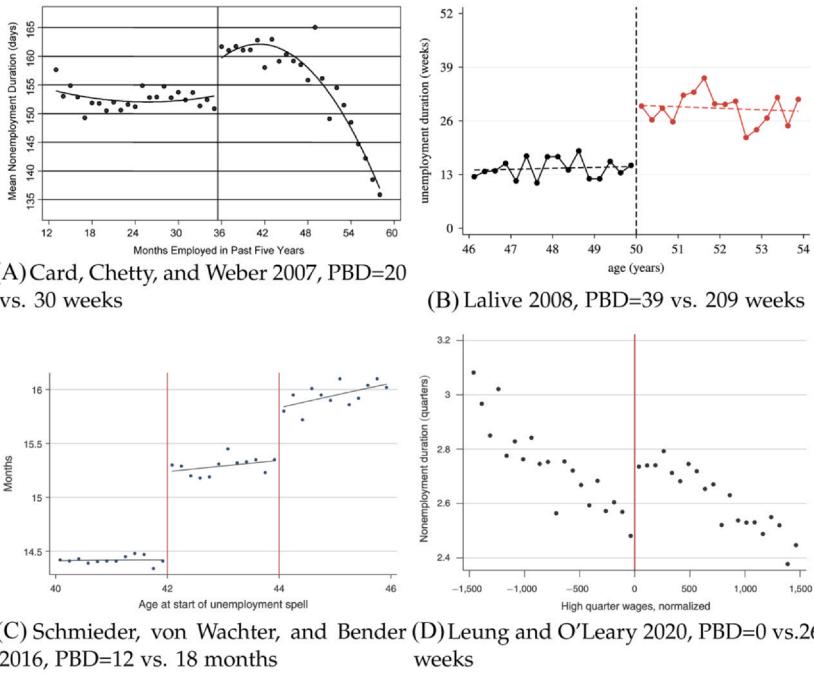


FIG. 6 The effects of UI on nonemployment duration: examples (A): Card, Chetty, and Weber 2007, PBD= 20 vs. 30 weeks (B): Lalive 2008, PBD= 39 vs. 209 weeks (C): Schmieder, von Wachter, and Bender 2016, PBD= 12 vs. 18 months (D): Leung and O'Leary 2020, PBD= 0 vs. 26 weeks. Notes: Panel (A) replicates Fig. 8a from Card et al. (2007), Panel (B) replicates Fig. 2 from Lalive (2008), Panel (C) replicates Fig. 6 from Leung and O'Leary (2020), and Panel (D) replicates Fig. 2b from Schmieder et al. (2016).

age 44 cutoff in Germany (Schmieder et al., 2016) with clear upward jumps in nonemployment duration. Panel (D) shows the effect of being ineligible for UI (i.e. PBD = 0 weeks) to the left of the cutoff vs. being eligible to a PBD of 26 weeks to the right of the cutoff, also showing a clear upward jump in nonemployment durations.

Overall, the evidence that increases in PBDs lead to longer nonemployment durations is extremely strong. A large number of high quality papers across a wide range of countries have produced similar estimates using RD designs (e.g., Centeno and Novo (2009) for Portugal, Huang and Yang (2021) for Taiwan, or Gerard and Naritomi (2021) for Brazil). There are also many papers providing clean evidence using Difference-in-Differences designs or, more recently, Regression Kink Designs (especially for the effects of UI benefit levels).

2.3.2 *The effects of UI the hazard rate*

The advent of high frequency administrative data, has allowed economists to obtain non-parametric estimates of the job finding hazards among unemployed individuals. The earliest estimates along those lines were Katz and Meyer (1990) and Meyer (1990). More recently, papers have provided estimates of how hazard rates shift in response to UI PBD changes.

Fig. 7 shows several such examples. Panel (A), from Card et al. (2007), presents the first figure in the literature that shows how the weekly job finding hazard shifts when PBD is extended (here from 20 to 30 weeks). The weekly hazard shows spikes every 4 weeks, likely because many jobs end at the end of the month and start on the 1st of the month. Furthermore, the figure clearly shows that for most of the spell the hazard rate is declining with unemployment duration until it increases again leading up to and right after UI exhaustion. Exactly at the exhaustion point for the PBD= 20 weeks group the hazard shows a spike in job finding rates. The figure also shows how the PBD extension leads to a lower job finding hazard for most of the spell roughly until the new exhaustion point (PBD=30 weeks). Panel (B), taken from Marinescu and Skandilis (2021), shows a similar figure for France plotting the monthly job finding hazards for 5 groups (PBD = 0, 6, 12, 24 and 36 months). The findings are qualitatively identical: the job finding hazard decreases throughout the spell except for a spike around UI exhaustion. Extending PBD moves the spike and reduces job finding rates up to the new exhaustion point. Panel (C) shows qualitatively the same result for Taiwan (Huang and Yang, 2021) and Panel (D) for Germany (Schmieder et al., 2016).

Overall, the decline in the job finding hazard with unemployment duration in conjunction with a spike in the hazard at UI exhaustion is now well documented.⁷

⁷ Other examples include Le Barbanchon (2016) for France, DellaVigna et al. (2017) for Hungary, DellaVigna and Paserman (2005) and Ganong and Noel (2019) for the US, Gerard and Gonzaga (2021) for Brazil, Uusitalo and Verho (2010) for Finland, and many more.

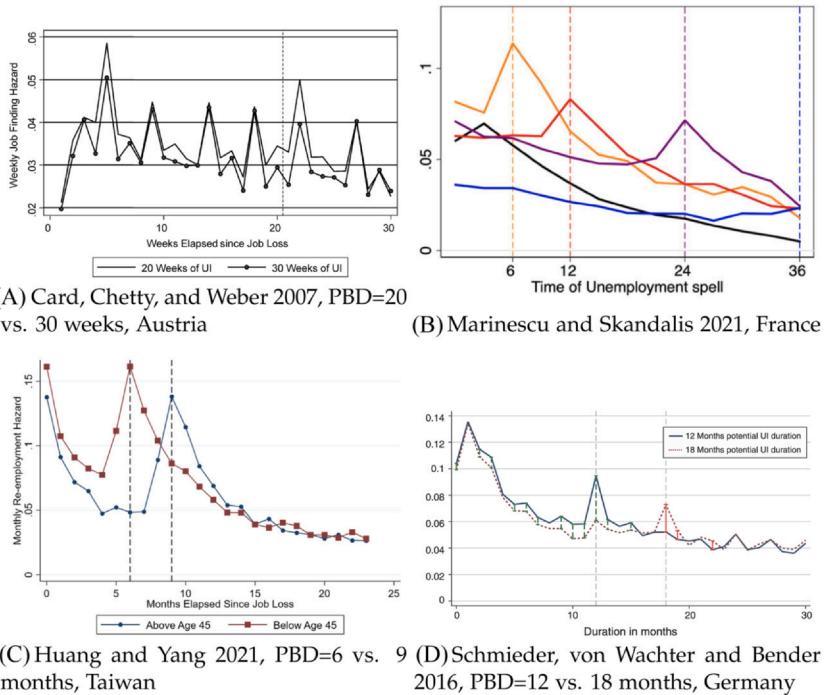


FIG. 7 The effects of UI on the job finding hazard: examples (A): Card, Chetty, and Weber 2007, PBD=20 vs. 30 weeks, Austria (B): Marinescu and Skandulis, (2021), France (C): Huang and Yang 2021, PBD=6 vs. 9 months, Taiwan (D): Schmieder, von Wachter and Bender 2016, PBD=12 vs. 18 months, Germany. Notes: The figure shows estimates of the unemployment exit hazard from different papers in the literature. Panel (A) replicates Fig. 9 from Card et al. (2007), Panel (B) replicates Fig. 2a from Marinescu and Skandulis (2021), Panel (C) replicates Fig. 7a from Huang and Yang (2021), and Panel (D) replicates Fig. 6b from Schmieder et al. (2016).

2.3.3 Reemployment wages

As discussed above, the basic job search model is ambiguous with respect to how UI extensions affect average reemployment wages if the wage offer distribution changes with unemployment duration (e.g., due to skill depreciation). Thus it is perhaps not surprising that the literature has found mixed results when estimating this effect. Figure 8 shows several examples from the recent literature. Panel (A), taken from Schmieder et al. (2016) shows that a 6 (and 4) month extension of UI benefits in Germany reduces average reemployment wages by about 0.8 log points. By contrast, Panel (C), from Nekoei and Weber (2017) shows that a 9 week extension in Austria slightly increases average reemployment wages and Panel (E), from Huang and Yang (2021) finds essentially no impact of an extension.

Several papers have analyzed how reemployment wages develop throughout the unemployment spell and how the reemployment wage path shifts in response

to a UI extension. Panel (B), from Schmieder et al. (2016) shows that reemployment wages are declining throughout the unemployment spell, by about 25 log points over 1 year. The figure also shows that extending PBD from 12 to 18 months has virtually no impact on the reemployment wage path throughout the unemployment spell. The only exception is that at the UI exhaustion points, the reemployment wage dips down relative to the other group. Panel (D) shows as similar figure from Lalive et al. (2015) for Austria. The finding is qualitatively very similar, with a declining reemployment wage path that is virtually unaffected by a large extension in UI benefits. Finally, Panel (F) shows the reemployment wage path among UI recipients with 6 and 9 months of PBD in Taiwan (based on Huang and Yang (2021)) also finding no shift in the reemployment wage path.⁸

Overall, there seems to be strong evidence that reemployment wages decline with unemployment duration. This is true whether the dependent variable is simply the post-unemployment wage or the difference between post- and pre-unemployment wages. The evidence on the effect of UI on reemployment wages is quite mixed. Many estimates in the literature are close to 0 and when they are statistically significant they are still estimated with sizable standard errors.

2.3.4 *Can the basic search model rationalize the evidence on job finding rates and reemployment wages?*

The empirical evidence that we laid out above highlights several stylized facts that appear to hold consistently across time and space: increasing PBD leads to larger nonemployment duration, the job finding hazard is decreasing for most of the spell and exhibit a spike at the exhaustion point, and reemployment wages are decreasing and respond only moderately to changes in PBD.

These broad findings are very consistent with the basic job search model that we developed in this chapter. To illustrate this, we calibrate the model to match the hazard rates and the reemployment wage path in Schmieder et al. (2016) using 4 different types of job seekers that differ by the cost of job search and the mean of the wage offer distribution.

Fig. 9 shows the aggregate hazard and the aggregate reemployment wage path for this calibrated model, simulated for $\gamma = 1$ and $\alpha = 1$ months. The figure also shows the corresponding empirical hazard rates from Schmieder et al. (2016). Panel (A) shows that the simulated aggregate hazard matches the main empirical pattern of a declining hazard rate and a spike at the exhaustion point very well. The model also captures the effect of the UI extension to $\gamma = 1$ months: First, the spike in the hazard moves to the new exhaustion point. Second, the hazard rate for $\gamma = 1$ is higher than for $\gamma = 0$ up to the new exhaustion point. Third, the hazard rates eventually cross and subsequently, the hazard rate is slightly higher for the $\gamma = 1$ group. From Fig. 8 and the previous

⁸ Similar declines of the reemployment wage path have been provided by Fallick et al. (2021), though without a comparison group.

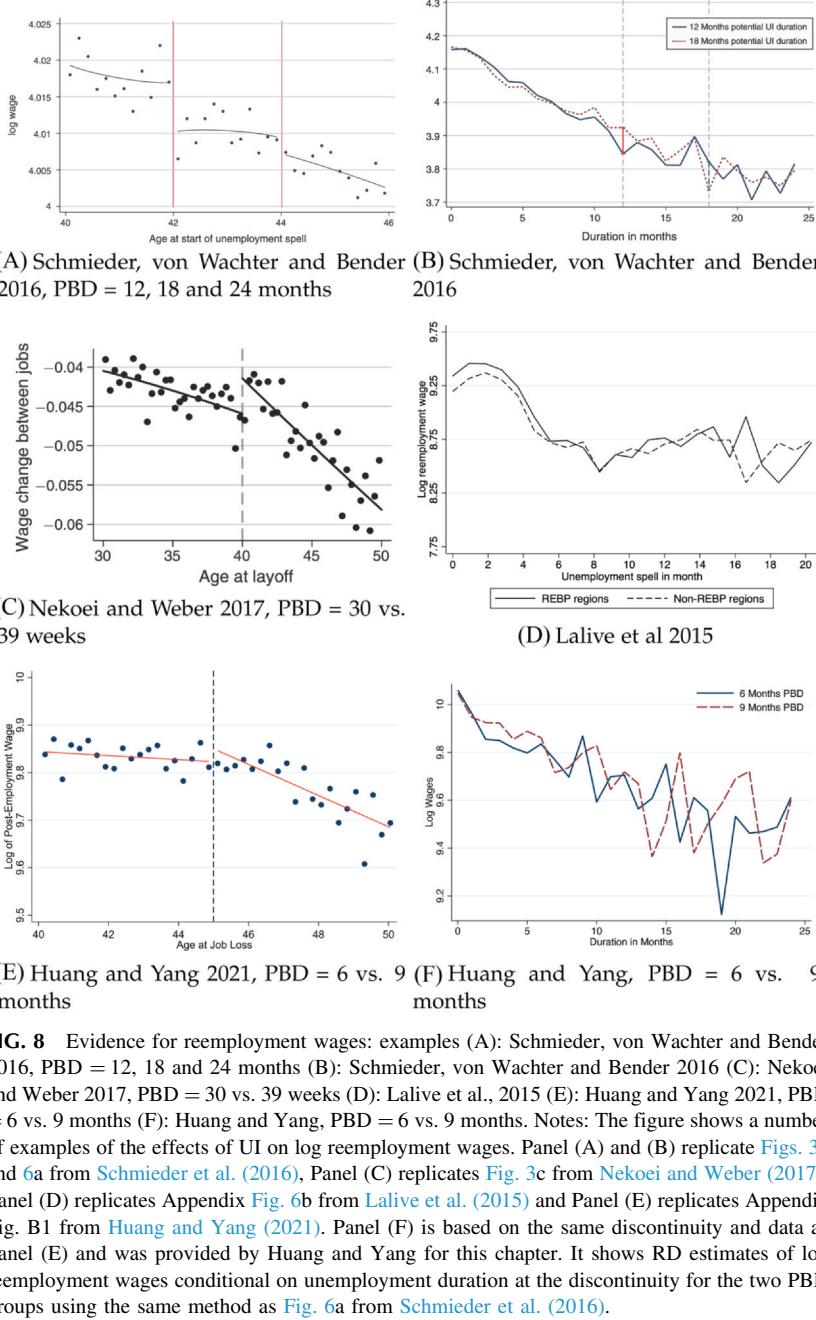


FIG. 8 Evidence for reemployment wages: examples (A): Schmieder, von Wachter and Bender 2016, PBD = 12, 18 and 24 months (B): Schmieder, von Wachter and Bender 2016 (C): Nekoei and Weber 2017, PBD = 30 vs. 39 weeks (D): Lalive et al., 2015 (E): Huang and Yang 2021, PBD = 6 vs. 9 months (F): Huang and Yang, PBD = 6 vs. 9 months. Notes: The figure shows a number of examples of the effects of UI on log reemployment wages. Panel (A) and (B) replicate Figs. 3a and 6a from Schmieder et al. (2016), Panel (C) replicates Fig. 3c from Nekoei and Weber (2017), Panel (D) replicates Appendix Fig. 6b from Lalive et al. (2015) and Panel (E) replicates Appendix Fig. B1 from Huang and Yang (2021). Panel (F) is based on the same discontinuity and data as Panel (E) and was provided by Huang and Yang for this chapter. It shows RD estimates of log reemployment wages conditional on unemployment duration at the discontinuity for the two PBD groups using the same method as Fig. 6a from Schmieder et al. (2016).

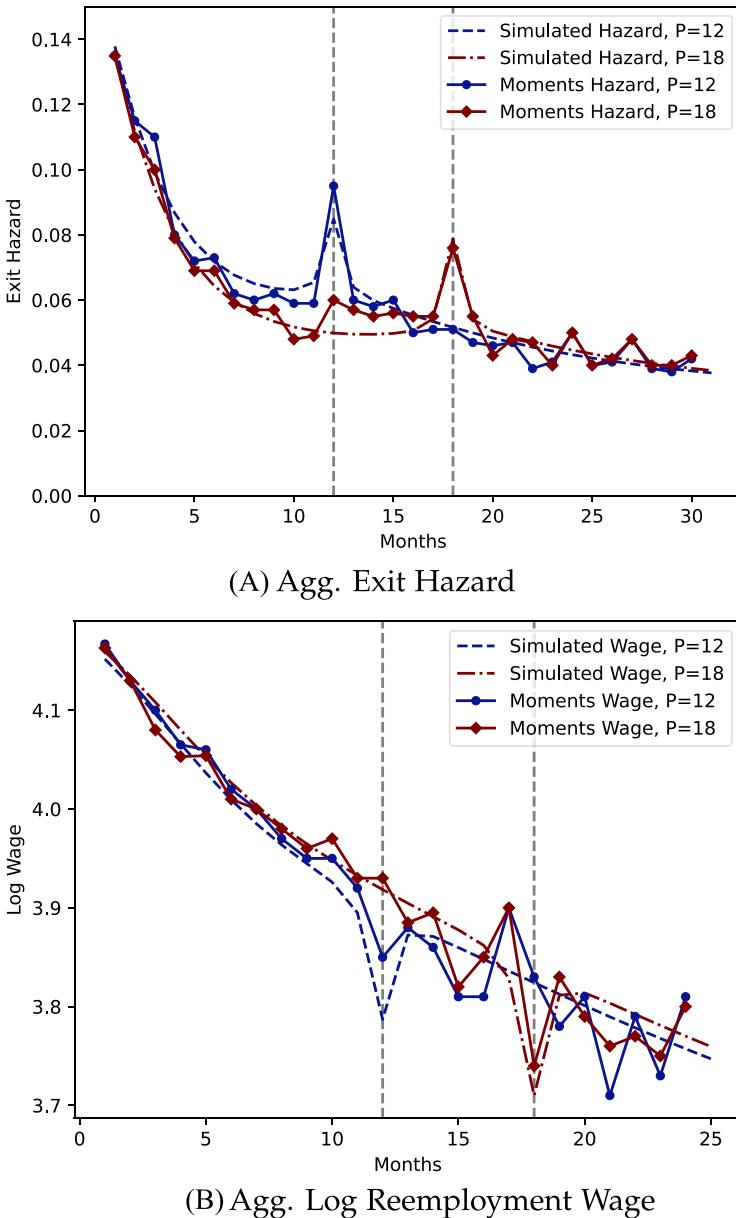


FIG. 9 Simulating the effects of a UI extension (A): Agg. exit hazard (B): Agg. log reemployment wage. Notes: The figure shows simulations of the model in 4 when changing the potential UI benefit duration from 12 to 18 months. Panels (A) to (D) show how search effort, the reservation wage, the exit hazard and the reemployment wage change for Type 1 workers.

discussion, we know that within the individual types, the hazard rate is increasing, there is no spike, and the hazard rates for γ weakly dominate the δ hazard rate. Thus the changes in type composition, i.e. dynamic selection, are the key driver of the empirical hazard rates. In the appendix (section A) we show the corresponding simulations for the 4 individual types, which highlight how the dynamic selection of individual types generates the spike in the exit hazard as well as the decline in hazard and reemployment wages.

Panel (B) shows the simulated reemployment wage and contrasts it with the model. Here too, the model captures the broad empirical pattern fairly well. The aggregate reemployment wage declines throughout the unemployment spell. There is a small upward shift in the reemployment wage path (driven by the increase in reservation wages).

This highlights that, when allowing for multiple types, the model is very flexible and can fit patterns that are very different from what a single type (homogenous) model would predict. The basic search model with heterogeneity has thus proven a very powerful tool to understand and analyze UI policy, both in papers relying on reduced form empirical methods (e.g., by providing the theoretical basis for welfare analysis, following [Baily \(1978\)](#) and [Chetty \(2008\)](#)), as well as papers that structurally estimate the search model, e.g., for policy predictions. However, this flexibility also has a downside. In practice there are many possibilities for specifying functional forms and/or the exact nature of heterogeneity and the model is often underidentified to distinguish between such choices. For example, it may *ex ante* not be obvious whether heterogeneity should be in the cost function, the discount parameter, or the utility function and the model may well fit the data quite similarly.⁹ Thus one has to be careful not to be lured into blindly trusting an estimated model with good in-sample fit, as the out of sample predictions or welfare implications may not be very robust to alternative model specifications. We argue below that many recent papers in this literature have been aware of these challenges and address them very thoughtfully, by carefully justifying the model choices and showing robustness checks to alternative specifications.

The challenge of under-identification is also particularly acute when trying to distinguish between different microfoundations of the search model. For example, in the model calibration above, the hazard rate falls throughout the unemployment spell entirely due to compositional changes (dynamic selection), while within types the hazard is increasing. However, a small modification of the model would be to assume that the cost of job search increases with unemployment duration. This would lead to declining search effort and job finding rate within a single job seeker type. Since we cannot observe the

⁹ Examples of this can be found in [DellaVigna et al. \(2017\)](#) or [Gerard and Naritomi \(2021\)](#), where both papers rely on criteria separate from the empirical moments to distinguish between different forms of heterogeneity

hazard rate within individuals over time (since each job seeker is only observed exiting the spell once), it is difficult distinguish these two explanations for the declining hazard empirically from each other, if the only data that is available is the typical administrative data with UI receipt, job start and end dates as well as wage information. This has inspired an active literature focused on additional information on job search to complement the evidence from administrative employment and UI records.

2.4 New empirical moments

Administrative data and the advent of credible research designs with clean, transparent identification has provided a strong basis to better understand job search and how job seekers respond to changes in UI policy. In particular admin data has established several key facts such as a declining exit hazard and reemployment wages throughout the spell as well as that increasing UI generosity leads to longer unemployment durations. As shown above these key findings can be rationalized very well with the basic job search model with multiple types of job seekers. However, administrative data also has important shortcomings. In particular the while the search model can rationalize the main findings the evidence on the model is only indirect. Indeed the two key control variables of the job seeker: search effort and the reservation wage are not observed in typical administrative datasets. Furthermore, since for each worker there is only a single job acceptance event with a single reemployment wage per unemployment spell, it is impossible to estimate empirical hazard rates or reemployment wage paths within individual. This in turn makes it very difficult to differentiate whether the time path in the aggregate hazard and reemployment wage are due to dynamic selection or individual behavior changing.

Motivated by these shortcomings of administrative data, recent years saw a plethora of high quality papers that seek to expand our understanding of job search by bringing new data to the table: high frequency panel survey data on job search, data from online job platforms, consumption data and more.

Here we lay out this new evidence on the key ingredients of the job search model: job search effort, reservation (or target) wages, and consumption. These data have shed new light on the micro foundations of job search. Afterwards in [Section 2.5](#) we will discuss various refinements and extensions of the basic job search model have been proposed, at least in part, to explain these findings.

2.4.1 Search effort

2.4.1.1 Survey data

One, and perhaps the most straightforward, way to obtain information on search behavior is to ask the unemployed. Such a measure could then be used to test the predictions from the standard job search model regarding the evolution of search through the spell and how it is affected by UI. There was a small literature in the 70 and 80 s that relied on small, cross-sectional surveys

on job search behavior (see for example [Devine \(1991\)](#), [Devine and Kiefer \(1993\)](#) for a discussion), but the evidence was relatively scarce. The research in the late 90s and 2000s shifted towards policy evaluations and natural experiments relying on the type of administrative data discussed in the previous Section (2.3).

The early 2010s then saw a revival of interest in shedding light on the microfoundations of job search. A first example was the use of time use diary data, pioneered by Alan Krueger and Andreas Mueller. In [Krueger and Mueller \(2010\)](#), they study job search using the American Time Use Surveys (ATUS) from 2003–2007. By asking respondents to fill out a time diary to carefully account for the time spent during a specific day, time use surveys are likely less are arguably less distorted. For example, in a survey that clearly focuses on job search, respondents may report higher search effort due to social desirability or acquiescence bias. While, in contrast, a time diary is more general and forces a certain degree of consistency on the respondent (the hours of the day have to add up), arguably reducing biases.¹⁰ [Krueger and Mueller \(2010\)](#) show that unemployed workers only spend about 41 min on job search related activities on a workday. They also show regression results showing that after controlling for other characteristics, workers who receive higher UI benefits spend less time on job search. Finally, they plot job search effort before and after jobloss and show that search increases up to the exhaustion point and then decreases, thus creating a spike in effort resembling the well known spike in the unemployment exit hazard. In a related paper, [Krueger and Mueller \(2012\)](#) analyze time use data from 16 countries in Europe and North America. The paper finds that the average time spent on job search among the unemployed is even lower in Europe, with only about 14–16 min on a weekday compared with 38 min in Canada and 41 min in the US. The paper estimates cross-country regressions for the time spent on search and finds little relation between UI generosity and search effort, but a relative strong correlation between wage inequality and search effort, consistent with the theoretical prediction that higher inequality leads to higher returns to search.

The downside of time use datasets is that they are relatively small when conditioning on unemployment and are cross-sectional with only a single observation per unemployed worker. This makes it difficult to trace out how search behavior changes over the unemployment spell in a way that is not confounded by dynamic selection. To fully understand the nature of job search, requires data that follows individuals over time throughout their unemployment spell with repeated observations of their job search effort. To fill this gap, [Krueger and Mueller \(2011\)](#) conducted a carefully crafted survey, which combined the strengths of the administrative data process with a high

¹⁰ For example, [Chou and Shi \(2021\)](#) discuss how hours worked are overestimated in the CPS and more accurately measured in the ATUS and how this can bias labor supply estimates.

frequency online survey. The [Krueger and Mueller \(2011\)](#) Survey (KM Survey), was based on a complete list of UI recipients in NJ as of September 2009. It then generated a random sample of around 63 thousand individuals, stratified by their unemployment duration at the time. Each individual was then invited via letter to participate in the online survey for a period of 12 weeks with weekly questions related to job search activities. Thanks to the sampling design the different cohorts (a cohort referring to all individuals with a specific unemployment duration at the start of the survey), could then be lined up to trace out job search activities for almost 2 years.

The questionnaire consisted of both a time use diary for a single day, as well as questions on job search for the whole week prior to the survey (recall question). The level of search from the time use diary is around 70 min per day, higher than in [Krueger and Mueller \(2010, 2012\)](#), but this is likely explained by the fact that the KM survey focuses on UI recipients who are probably more strongly attached to the labor market than the average unemployed and perhaps because UI recipient are subject to job search requirements to maintain benefit eligibility. The KM survey reveals a curious pattern: within cohorts (or within individuals when controlling for individual fixed effects) search effort is falling rapidly (about 30 min over 12 weeks), while across cohorts search effort is essentially flat. The paper discusses this pattern carefully: first, it could be that effort is truly declining within individuals and this is masked by dynamic selection across cohorts. Second, there could have been secular time trends over the survey duration (since all cohorts were interviewed over the same period duration is colinear with calendar time). Third, there may be reporting bias, e.g., respondents may have stated less effort over time in order to avoid conditional follow-up questions in the questionnaire or because they became more honest as they realized that their responses had no negative consequences. While the paper does not fully resolve these conflicting explanations, the possibility of systematic reporting bias is at least an important caveat for interpreting this key finding.¹¹ The paper provides some analysis for search effort around UI exhaustion. After controlling for unemployment duration, Search effort appears relatively flat leading up to UI exhaustion and declining moderately afterwards. One possible confounder for the exhaustion analysis is that the survey was conducted at the height of the Great Recession with several UI extensions. While UI benefits were exhausted / lapsed for some recipients, they were subsequently reinstated.

While the analysis of search effort over the spell thus comes with important caveats, the KM survey is a remarkable achievement. It provides a plethora of other information about the job search process. For example, [Krueger and Mueller \(2011\)](#) show that job offers are rare: in a given week, only about 2–4 %

¹¹ The published article was accompanied by a thoughtful discussion by Stephen Davis ([Davis, 2011](#)), who also seems to view reporting bias as a likely explanation for at least some of the within-person decline.

of UI recipients receive a job offer. Furthermore, it seems most job offers are accepted. Krueger and Mueller (2011) also provide fascinating evidence on subjective well-being. The unemployed are quite unsatisfied overall and report having a ‘bad mood’ or experiencing ‘sadness’ most of the time. Furthermore, when measured by moment-to-moment measures, workers seem to feel worse over time the longer they are unemployed. From the time diary data, it is also striking that on a given day the unemployed are the least happy, and the most sad and stressed while searching for a job.

DellaVigna et al. (2022) build on the KM survey by collecting data on job seekers in the German UI system. Following a similar overlapping cohort design, they conduct a large, high-frequency survey among UI recipients via text messages (SMS) that follows each cohort of workers over an 18 week period. By focusing on just one or two questions on any given day, the survey collects less information than the KM survey, but has substantially less attrition and may reduce survey response bias caused by interviewee fatigue. Every respondent is asked twice a week how many hours they spend on job search activities on the previous day. The study was conducted over a 2 year horizon with new cohorts starting every month. As a result, calendar time is not perfectly collinear with unemployment duration and the paper can conduct detailed checks for the possibility of reporting bias. Overall the paper finds no evidence for reporting bias and the within and between person estimates of how search effort evolves with unemployment duration are very similar.

Compared to the KM survey, the setting in DellaVigna et al. (2022) features a very stable economic environment with low unemployment rates and a predictable UI system with a clear UI exhaustion point. An important advantage of the survey is that it samples individuals with different potential benefit durations, which are determined by the contribution history and age of the UI recipient. For these different PBD groups the administrative data shows the typical hazard path of declining hazards early in the spell and a spike in the exit rate at UI exhaustion, comparable to the results in Schmieder et al. (2012) for an earlier time period. Thus the paper can explore how search effort is related to UI exhaustion and PBD across groups and compare the search effort path with the corresponding hazard rates.

The SMS data in DellaVigna et al. (2022) reveals several key findings: First, the level of search effort is with around 83 min in a similar ballpark as the recall estimate in the KM survey. Second, within individuals, search effort is essentially flat in the first 6 months after UI entry. Third, Search effort rises by about 8 % in the 2 months leading up to UI exhaustion and falls by a similar amount afterwards, thus mirroring the spike in the job finding hazard at exhaustion. This suggests that the decline in the aggregate hazard early on in the unemployment spell is not due to declines in search effort. Furthermore, the spike in effort at exhaustion is substantially smaller (around 8 %) than the spike in the job finding hazard (around 40 %).

The fact that there is a spike in search effort at the exhaustion point suggests that PBD does impact search effort and in particular that shorter PBD will lead to higher search effort earlier. This is further supported by some RD analysis, reported in the online appendix of the paper, using an age discontinuity determining PBD at age 50. The analysis provides evidence, albeit not very precisely estimated, that PBD extensions reduce search effort earlier in the spell.

Further evidence that UI generosity affects job search effort, is offered by [Lichter and Schiprowski \(2021\)](#), who study how UI generosity affects self reported job search effort in a Difference in Difference designs. The paper exploits a reform in the German UI system in 2008 that raised PBD from 12 to 15 months for workers age 50 to 54. Relying on a survey, called the IZA evaluation dataset, they show that this reform substantially reduced search effort in the affected group.

[Table 1](#), Panel A summarizes some of the stylized findings from the literature on job search based on survey data.

2.4.1.2 Search platform and process data

Another development in the recent literature was the use of data from online search platforms and other process-generated data to study job search. As the means of job search (vacancy postings, job applications etc.) have shifted from traditional offline to online methods over the past 2 decades, there is in principle a wealth of data captured by various online platforms that facilitate job search.

A first example, relies on the fact that search engines are a simple starting point for job seekers. [Baker and Fradkin \(2017\)](#) uses Google search data to construct a metropolitan area level of job search activity. They first validate this constructed search index by comparing it with estimates of job search from the ATUS job search questions and show that it seems to capture spatial and temporal variation of job search effort very well. They then estimate the effect of PBD extensions during the Great Recession on the search index using a DiD approach as well as an eventstudy design. The DiD analysis shows small, statistically significant, effects: a 10 week extension of PBD reduces search effort by about 1–2 %. The eventstudy shows statistically insignificant effects, but is underpowered and cannot reject the DiD results.

Also focusing on the PBD extensions during the Great Recession, [Marinescu \(2017\)](#) uses data from the job board Careerbuilder.com, one of (if not the) largest job search platforms in the US at the time. The paper relies on the state and federal UI extensions for 2 identification strategies: an eventstudy design and a fuzzy RD based on UI extensions that were triggered by unemployment rate thresholds. Using the eventstudy design Marinescu finds that a 10 week PBD extension reduces applications in a state by about 4 %, somewhat larger than the impact in [Baker and Fradkin \(2017\)](#). The RD design leads to very similar results, though somewhat noisier. A key advantage of using job platform data, is that the paper can observe both the supply side (applications)

TABLE 1 Evidence on job search effort.

Paper	Country	Data	Evolution through spell			Design	Effect of UI Sign of effect
			Level	Initial evolution	Exhaust. point		
Panel A: Survey Data							
Krueger and Mueller (2010)	US	ATUS	41 min per day	Spike	OLS/N	— 0	
Krueger and Mueller (2012)	Europe	Time Use	14–16 min per day	Cross-country	OLS	— ≈ 0	
Krueger and Mueller (2011)	US	NLS Web Survey	70 min per day	Flat / Decreasing	OLS	— 0	
DellaVigna et al. (2022)	Germany	SMS Survey	80 min per day	Flat	Spike	Simple comparison	— 0
Lichter and Schiprowski (2021)	Germany	IZA Eval.			DiD	DiD	— 0
Panel B: Search Platform and Process Data							
Marinescu (2017)	US	Online Search Platform			DiD / RD	— 0	
Baker and Fradkin (2017)	US	Google Search			DiD	— 0	
Faberman and Kudlyak (2019)	US	Online Search Platform	1.8 applications per week	Decreasing			
Marinescu and Skandalis (2021)	France	Online Search Platform		Increasing	Spike	Simple comparison	— 0
Massenkovoff (2023)	US	DOL Audits	2 job contacts per week	Flat	RKD	—=0	

as well as the demand side (posted vacancies). Using the same identification strategies, she shows that there is no effect of the PBD extensions on posted vacancies, and as a result labor market tightness (the difference between log vacancies and log applications) increases in response to UI extensions suggesting that each individual sent application is more likely to be successful. This has important implications for the general equilibrium effects of UI, which we will discuss in [Section 3](#).

While [Baker and Fradkin \(2017\)](#) and [Marinescu \(2017\)](#) focus on the effects of UI on aggregate search effort, [Faberman and Kudlyak \(2019\)](#) leverage data from the SnagAJob job platform to analyze the evolution of search effort within individuals. In the paper, they can follow users of the platform and analyze how search behavior changes over time. A downside of the data is that it does not have information on unemployment. So instead of focusing on unemployment spells, the paper focuses on the period when a worker first submits an application until the first time there is a 5 week break of activity on the platform. The main result of the paper is that it shows that within individuals there is a clear decline in the number of applications sent per week. While workers apply to around 3 applications in the first week, this falls to less than 1 application in the second week and then falls by another 50 % over the next 10 weeks. The paper also shows that individuals with long spells search more throughout the spell than individuals with short spells. This correlation goes against the notion that higher search effort leads to lower spells, but may be driven by selection. The paper does not observe unemployment or UI receipt, so it cannot speak to how UI generosity or exhaustion affects search.

[Marinescu and Skandalis \(2021\)](#) fill this gap by using data from a job search platform that is administered by the French Public Employment Service and linked to administrative data from the UI system. As a result, they can observe when individuals enter unemployment, their UI eligibility, and when they exhaust UI benefits. The setting also features several groups of individuals with distinct PBDs. An important strength of the paper is that it can show job finding rates for several PBD groups and for all groups job finding falls over time but with a spike at the exhaustion point. Using the application data, the paper documents that within individuals search effort increases over time until it peaks at UI exhaustion. After individuals exhaust UI benefits, search effort falls. Thus search effort shows a spike around exhaustion that mimics the spike in the job finding rate. Overall, it appears that search effort does not explain the initial decline in the job finding rate, but may play an important role in explaining the spike at the exhaustion point. By showing the search effort is clearly affected by PBD, it also provides evidence that UI generosity affects overall search effort (higher generosity leading to lower overall effort).

The paper focuses on similar questions to [DellaVigna et al. \(2022\)](#), but with different strengths and weaknesses. Due to the administrative nature of the data [Marinescu and Skandalis \(2021\)](#) features a much larger sample (around 450,000 observations) and does not suffer from attrition. Furthermore the

applications represent actual job search effort and are not driven by reporting bias. On the other hand, the French job platform only accounts for a small share of all job search (only about 5 % of jobs are found through the platform) and the reported level is very low (less than 1 application per month), while the survey in [DellaVigna et al. \(2022\)](#) should capture all forms of job search. Overall, the results from the two papers are complementary, with both showing that search effort spikes at the UI exhaustion point.

[Faberman and Kudlyak \(2019\)](#) and [Marinescu and Skandalis \(2021\)](#) come to the opposite result, with respect to how search effort varies within individuals over time. While search effort decreases in [Faberman and Kudlyak \(2019\)](#) within individual at the beginning of the unemployment spell, it increases in [Marinescu and Skandalis \(2021\)](#). One possibility might be the different definition of a spell. A spell in [Faberman and Kudlyak \(2019\)](#) starts mechanically with being active on the platform. It may then well be that over time individuals switch to other platforms and search methods thus contributing to the observed decline. On the other hand in [Marinescu and Skandalis \(2021\)](#), a spell starts with unemployment entry and individuals may only gradually begin to use the platform. Of course other differences in sample and context may also affect this comparison.

Finally, [Massenkoff \(2023\)](#) uses unique data from UI claim audits in the US. The Department of Labor conducts random audits among UI claimants via phone, asking questions about reservation wages and job applications and cross-validating the answers with employer reports. The paper has more than a million audit reports from 1987 to 2022. Using information from individuals who were audited more than once in a single UI spell, the paper shows that within a spell job applications are essentially flat. The paper also uses caps imposed on weekly UI benefit levels to estimate the effect of UI benefits on search effort using a Regression Kink Design. The paper finds Essentially no impact of UI on search effort.

2.4.2 Reservation / target wages

Now we turn to what survey and platform data reveal about wage strategies of job seekers. The theoretical job search literature has long featured the reservation wage as a key measure for describing workers search strategies at a given point in time and a key determinant for search outcomes. This is also the approach taken by the basic search model from [Section 2.2](#) that features search effort and reservation wages. However, the concept of the reservation wage also faces complications when attempting to empirically measure it. The concept is quite abstract and is not be easy to translate into a survey question. For example, the question “what is the minimum wage at which you would accept a job offer” may be quite ambiguous without specify exactly what the job is. A job seeker may ask for a certain minimum wage for one job that offers flexible hours, a pleasant work environment, and other benefits, but asks for a

higher minimum wage for another job with different attributes. Thus asking for a reservation wage may be meaningless without holding job characteristics constant, which in turn may be hard to do within or across individuals.

Several papers measure target wages instead of reservation wages. A target wage captures the notion that a key decision of workers is which job to apply for and generally speaking workers have certain expectations what a given type of job may pay. A typical way to measure the target wage is to simply ask or observe what the last job application was and how much that type of job typically pays. We will return to the difference between reservation and target wage models in the next section.

The earliest papers that empirically studied the evolution of reservation wages relied mostly on cross-sectional data, e.g., [Feldstein and Poterba \(1984\)](#). The few papers based on panel data had very small samples. For a discussion of this earlier literature see [Devine \(1991\)](#) and the discussion in [Krueger and Mueller \(2016\)](#).

Turning to the modern empirical evidence, we will first discuss studies based on survey data before turning again to papers based on data from search platforms and process data.

2.4.2.1 Survey data

The KM Survey asked detailed questions about what type of job workers are looking for (occupation, hours per week) and what is the lowest wage or salary to accept an offer. They use the latter question as a measure of the reservation wage. Since they also ask in subsequent waves about received job offers and what those offers paid and whether the job seeker accepted them it is also possible to estimate whether the reservation wage is predictive of future job acceptance. [Krueger and Mueller \(2016\)](#) provide a detailed analysis of the reservation wage information in the KM data. They use the reservation wage ratio, i.e. the ratio of the reported reservation wage and the pre-unemployment wage as the main outcome variable. The first striking fact is that reservation wage ratio is high, with a mean of around 1. There is also a wide range of reported reservation wages, with a standard deviation of the log reservation wage ratio of around 0.37, which suggests that many job seekers report reservation wages as low as 60 % of their previous wage but similarly many report reservation wages 40 % higher than their previous wage. By relying on the nonlinearities in the benefit schedule the paper estimates the effect of weekly UI benefits on the reservation wage controlling for the previous wage and other controls. The implied elasticity of the reservation wage with respect to the benefit level is close to zero, statistically insignificant and somewhat sensitive to controls.

The paper then moves on to explore how reservation wages vary with unemployment duration. Within individuals the reservation wage does not vary with unemployment duration and is essentially flat. Furthermore, it is not

affected by UI benefit exhaustion or a UI extension that occurred during the sample period. As additional measures of the job search strategy, the paper also considers whether workers apply to jobs in lower paying occupations or would be willing to accept longer commutes. The paper finds that workers are indeed applying to slightly lower paying occupations and report being willing to accept slightly higher commutes (acceptable commuting time increases by 4.6 min over one year). The paper also shows that reservation wages do show a small but significant decline for 2 subgroups: older workers and workers with higher initial savings. Overall, it appears that workers do not change their job acceptance behavior very much over the unemployment spell.

Finally, the paper analyzes whether stated reservation wages are consistent with predictions from a reservation wage model. The basic reservation wage model would imply that all jobs paying less than the reservation wage are rejected and all jobs paying more are accepted. This does not hold in the data: many jobs paying less than the stated reservation wage are still accepted, while many jobs paying more than the reservation wage are rejected. On the other hand, the reservation wage is predictive and job acceptance probabilities rise as the offered wage rises relative to the reservation wage. When job offers are not accepted the data asks why and it seems other job attributes like hours or commuting distance are the main reasons.

[DellaVigna et al. \(2022\)](#) collected information on the target wage of job seekers, by asking participants what the estimated wage was of the last job they applied to. This target wage measure is very flat throughout the spell and then declines slightly after UI exhaustion.

Using panel data from a survey among job seekers in Belgium who are interviewed at UI entry, after 3 and after 6 months, [Deschacht and Vansteenkiste \(2021\)](#) estimate the within person change in reservation wages over time. The paper finds that reservation wages fall by about 0.4 % per month or about 5 % over the course of a year.

[Lichter and Schiprowski \(2021\)](#) use the same DiD design described above to estimate the effect of a UI extension on reservation wages. The point estimate is positive but small and statistically insignificant.

2.4.2.2 Search platform and process data

The French UI system collects information on reservation wages from all new UI entrants. [Le Barbanchon et al. \(2019\)](#) use this information to estimate how reservation wages are affected by UI. They exploit a reform in 2009 that altered how PBD is determined for UI claimants. While prior to the reform, PBD was a step function of days worked in the previous year, after the reform this became a continuous function. This leads to differential changes in PBD along the previous days worked variable which the paper uses to construct a Diff-in-Diff design. The paper finds no effect of PBD on stated reservation wages and the estimates are quite precise. The paper also finds no effect on desired hours, type of contract or willingness to commute. The paper validates these results

further by exploiting an age discontinuity at age 50, where maximum PBD increases from 24 to 36 months. Using an RD design at this discontinuity leads to very similar results. Overall, the paper finds no evidence that large changes in PBD affect job selectivity.

Using the job platform data described before, [Marinescu and Skandalis \(2021\)](#) can observe what type of jobs workers apply to. By calculating expected wages for these types of jobs, they can thus construct a measure of the target wage over the unemployment spell. The target wage is decreasing but the effect is quite small with a decline of about 1.5 % over one year. Target wages also decline slightly after UI exhaustion.

Since 2015, UI recipients in Denmark are required to document in each week around 1.5–2 jobs that they applied to in order to remain eligible for UI benefits. [Fluchtmann et al. \(2023\)](#) uses this information to create various characteristics of applied-for-jobs and thus construct target wages, target hours and other measures. A key advantage compared to survey data is the large coverage (all UI recipients in Denmark) as well as the comprehensiveness of the measure, since it is not limited to applied-for-jobs from a single platform. The paper documents that within individuals target wages decline slowly with unemployment duration. Over the duration of a year, mean target wages decrease by around 1 % when controlling for person fixed effects. There is also a small but precise decrease in the probability of applying for full-time jobs by around 3 % points over one year.

Finally leveraging the DOL audit data, [Massenkoff \(2023\)](#) analyses reservation wages. The ratio is somewhat smaller than in the KM data or the French data ([Le Barbanchon et al., 2019](#)). The paper finds a significant within-person decline in reservation wages of around 5 % over one year. Relying on the Regression Kink design to estimate the effects of UI benefit levels on reservation wages, the paper finds no effect ([Table 2](#)).

2.4.3 Evidence on consumption during unemployment

Arguably, the central concern associated with unemployment is its impact on consumption. This is clearly true from the individual perspective, as the painful loss in income and the resulting drop in consumption is a key motivating factor to engage in job search.¹² It is also the case from the perspective of policy makers, who are interested in increasing aggregate welfare through insurance against income shocks and, perhaps, redistribution. This central role of consumption is highlighted in the basic search model where flow utility is defined over consumption and the cost of job search. And similarly, consumption plays a key role in the theoretical (and empirical) analysis of the optimal design of UI policy (see [Section 3](#)).

¹² It should be noted that unemployment may be painful for reasons not related to consumption/income. For example, workers may enjoy working or obtain a sense of self-worth from their employment, and play an important social role. Indeed a sizable literature has documented negative effects of unemployment on happiness, social contacts and mental health.

TABLE 2 New evidence on the reservation/target wage.

Paper	Country	Data	Measure		Evolution Through Spell		Design	Sign of Effect	Effect of UI
			Reserv. (R)	Target (T)	Level	Initial Evolution			
Panel A: Survey Data									
Krueger and Mueller (2016)	US	NJ Web Survey	R	0.95	Flat / decreasing for some groups	Flat	OIS	— ≈ 0	
DellaVigna et al. (2022)	Germany	SMS Survey	T	1.17	Flat		Decreasing		
Deschacht and Vansteenkiste (2021)	Belgium	Survey	R	0.99	Decreasing, 5 % per year				
Lichter and Schiprowski (2021)	Germany	IZA Eval.	R				DID	—=0	
Panel B: Search Platform and Process Data									
LeBarbanchon, Rathelot, and Roulet (2019)	France	Employment Agency	R	0.93			DID and RD	—=0	
Marinescu and Skandalis (2021)	France	Online Search Platform	T		Decreasing, 1.5 % per year		Decreasing		
Fluchtmann et al. (2023)	Denmark	Employment Agency	T		Decreasing, 1 % per year				
Massenkoff (2023)	US	DOL Audits	R	0.86	Decreasing, 5 % per year		RKD	—=0	

The first systematic analysis of consumption patterns of unemployed workers was conducted by [Gruber \(1997\)](#). The Panel Study of Income Dynamics (PSID) provides longitudinal information on food consumption on an annual level. Using this data [Gruber \(1997\)](#) shows that food consumption drops on average by 6.8 % at the onset of unemployment. He also shows that the consumption drop is smaller for individuals with higher UI benefits, suggesting that UI does indeed provide some consumption smoothing.¹³

[Kolsrud et al. \(2018\)](#) find similar results using registry data from Sweden. The registry data contains detailed information on consumption on an annual level. Exploiting the timing of the onset of unemployment relative to calendar years the paper can trace out how consumption drops evolve through the unemployment spell. They find that consumption drops around 4.4 % in the first 20 weeks, but more than doubles, to 9.1 % for those who are unemployed longer. Since in their sample UI benefits are constant over the spell, this suggests that individuals' ability to smooth consumption through other means is much higher for short unemployment spells. [Landais and Spinnewijn \(2021\)](#) revisit the same data and find a consumption loss of around 13 % for the unemployed.

The first high-frequency study of consumption at the onset of unemployment as well as at UI exhaustion was conducted by [Ganong and Noel \(2019\)](#). They use banking data from JP Morgan Chase to obtain household level, monthly data on income and consumption for detailed spending categories. UI spells can be identified off of direct deposits from the UI system. The paper contrasts the household income path with the household consumption path through the UI spell. At the start of unemployment, income drops suddenly by 15 %, it continues to drop around 1.5 % per month while on UI and then drops by an additional 40 % at exhaustion. All together, household income is 66 % lower after UI exhaustion relative to the pre-unemployment level. Spending on the other hand drops by 6 % at the beginning of unemployment, then declines by slightly less than 1 % per month and then drops by another 12 % at benefit exhaustion. In total, at exhaustion spending is about 25 % below the pre-unemployment level.

The sudden consumption drop at the onset of unemployment is perhaps easier to explain than at UI exhaustion. Unemployment may come as a surprise to the individual and thus they respond to the sudden information with adjusting spending. By contrast UI exhaustion should not be a surprise. People know in advance that UI benefits are only paid for a finite period (26 weeks for most workers in this context) and that they will face a high probability of reaching the exhaustion point once they are only 2 or 3 months away from it given an average job finding hazard of around 20 %. But then why do individuals not reduce consumption more prior to exhaustion to save up and to be able to smooth consumption around exhaustion?

¹³ [Hendren \(2017\)](#) revisits the PSID data and shows that UI recipients already show a 2–3 % drop in consumption in the year prior to unemployment, a sign that individuals anticipate unemployment risk.

[Gerard and Naritomi \(2021\)](#) provide another piece of high-frequency evidence on consumption, this time from Brazil. Their setting differs from [Ganong and Noel \(2019\)](#), in that job losers are eligible for a substantial severance payment (SP). After job loss they are eligible to 5 months of UI. The severance pay is quite large at almost 5 months of income. The paper contrasts workers who are laid off (and eligible to SP), fired (and not eligible to SP), and workers who do not lose their job. The paper finds that job losers with SP have a large increase in spending right after job loss, about 30–40 % in the first and second month after job loss. Spending then falls rapidly for workers who remain unemployed and is about 20 % below pre-unemployment levels 1 year after job loss. Given that job losers know that they are experiencing a large negative income shock, it is perhaps surprising that SP is not used for more consumption smoothing. The pattern holds for many types of goods, including nondurables and food, and thus is not driven by large durable purchases (with lasting utility flow). The paper also shows that there is a relative sharp drop in consumption for UI exhaustees of around 15 %. Finally, the authors provide a weekly spending analysis showing that spending is substantially higher in the week when workers receive their monthly UI payments. Overall, spending appears very responsive to income in the short term, even in the face of what appear to be substantial incentives to smooth consumption.

2.4.4 Other types of evidence

In this section, we focus on real-world evidence that relates directly to the individual search behavior of unemployed workers: job selectivity, search intensity, and consumption. There are many other related papers with either a somewhat different focus or different types of data. For example we do not cover the literature focusing on differences by groups (e.g., gender differences as in [Le Barbanchon et al. \(2021\)](#)).

One example are audit (or correspondence) studies, where researchers send out fake resumes to potential employers, randomly varying attributes of the resume.¹⁴ By their nature these studies focus on the labor demand side of job search. The researchers measure call-back rates for the resume to see how employer interest is affected by the random attributes. While economists have studied many questions using audit studies (in particular focusing on discrimination), the most pertinent to job search among the unemployed are audit studies that compare callback rates by length of unemployment. [Kroft et al. \(2013\)](#) were the first study focusing on unemployment and found that applicants with long-term unemployment spells were substantially less likely to be called back by employers. Similar results were found by [Eriksson and Rooth \(2014\)](#) in Sweden who also find lower call back rates for the long-term unemployed. By contrast [Nunley et al. \(2017\)](#), for college graduates, and

¹⁴ See [Neumark \(2018\)](#) for a review.

[Farber et al. \(2019\)](#), for somewhat women, do not find evidence of an effect of unemployment duration on call-back rates. Overall the evidence from audit studies is somewhat mixed but at least in some contexts seems to suggest that unemployment duration may have a causal, negative impact on the type of job offers workers receive.

There is also a large experimental literature studying job search in the lab. Researchers who rely on a controlled lab environment, can use carefully designed experiments and randomization to obtain estimates of deep parameters and mechanisms that would be very hard to isolate with real world data. For example, [Brown et al. \(2011\)](#) study participants reservation wage in a lab environment and find the reservation wages decline. The paper can distinguish between a number of alternative explanations why reservation wages may decline, e.g., because individuals learn about the optimal reservation wage strategy over time. The challenge with such data is that it is not clear how well the decisions in the lab extrapolate to high stakes real world decisions. We refer to the review article by [Cooper and Kuhn \(2020\)](#) that discusses this literature in some detail.

There is also a substantial literature that uses field experiments to study how various interventions affect job search outcomes. This literature is often interested in active labor market programs (ALMP), which we discuss in [Section 4](#). However, some of this work relates to the information environment of the job search process. By altering the type of information available to job seekers these papers show that information is an important constraint for job seekers, e.g., [Belot et al. \(2019\)](#) or [Belot et al. \(2022b\)](#).

2.5 Refining the search model

The past years have seen a wealth of new information on the behavior of job seekers. Much of this work was designed to improve our understanding of the mechanisms that underlay the job search process. We now turn to what this information reveals about our theoretical understanding of job search. Starting from the basic job search model from [Section 2.2](#), we first discuss what this evidence suggests about the key mechanisms captured in that model, such as the relative importance of search effort and reservation wages. We then discuss modifications and extensions to the job search model that have been proposed in the literature and to what extent the empirical evidence supports these refinements.

We summarize this discussion in [Table 3](#). The table shows for each discussed mechanism or proposed refinement the type of evidence that exists in the literature and the, perhaps tentative, conclusions we would draw from this evidence. The table also provides our assessment for the strength of evidence for this proposed conclusion. We categorized the strength as: “Unclear” if there is either no clear evidence in any paper or conflicting evidence; “Suggestive” if there is a single paper with clear evidence in support of the conclusion (and no

TABLE 3 Modeling choices and refinements with evidence from literature.

Modeling choices and refinements	Conclusion from literature	Strength of evidence	Type of evidence
Choice variables that determine UI responses	Search intensity responds to UI generosity; job selectivity does not respond	Strong	See Table 1 and Table 2
Directed Search vs. Reservation Wages	Difficult to empirically separate	Unclear	Survey and platform data
Duration Dependence in Reemp. Wages	Some evidence for duration dependence: skill depreciation / declining reservation and target wages	Moderate	Evidence from reemployment wages, direct measures of skill
Duration Dependence in Job Finding Rates	Dynamic selection accounts for majority of decline in hazard rate, little role for search effort / res. wages	Strong	See Table 1 and Table 2
Present Bias	Clear evidence for present bias	Strong	Spike in hazard, Consumption patterns, Structural estimates
Reference Dependence			
-Search Intensity	Reference dependence partly responsible for spike in hazard	Moderate	Structural estimates / Policy variation
-Wages	Wage offers evaluated relative to previous wages	Unclear	Reduced form
Biased Beliefs			
-Level	Job seekers overestimate job finding probability	Moderate	Comparing subjective and actual job finding probabilities
-Return to search	Not clear whether job seekers over or underestimate returns to search.	Unclear	
Locus of Control	Internal locus of control associate with higher search effort and job finding	Moderate	Reduced form regressions

Employer collusion / Storable Offers			
-Timing of Job Start	Evidence against collusion over job start dates	Suggestive	Survey evidence on job offer and job start dates
-Timing of Recalls	Some evidence that recalls are timed with UI exhaustion	Moderate	
Learning / Information	Some evidence that job seekers learn about stochastic process	Unclear	Reduced form regressions / structural estimates

Notes: For discussion of the evidence see the text. Strength of evidence is subjective, but follows roughly the following key: Unclear—either no clear evidence in any paper or papers with conflicting evidence; Suggestive—a single paper with clear evidence in support; Moderate—2 to 3 papers with evidence in support; Strong—4 or more papers with evidence in support.

clear evidence against); “Moderate” if there are 2–3 papers with evidence in support and “Strong” if there are 4 or more papers with evidence in support of the conclusion. This assessment is of course subjective but the intent is to give the reader a sense of conclusions about the search model that have a lot of empirical support vs. other areas that are more speculative and where additional research may be particularly valuable.

2.5.1 Channels that determine UI response

We saw that in the basic job search model UI generosity, say a PBD extension, affects both search intensity and the reservation wage ϕ which in turn affect the job finding rate: $\rightarrow 0$ and $\frac{\partial}{\partial \phi} \rightarrow 0$.

As we saw in [Section 2.4 Table 1](#), there are three papers that have analyzed the effect of UI extensions on search effort using Diff-in-Diff designs: [Marinescu \(2017\)](#), [Baker and Fradkin \(2017\)](#), and [Lichter and Schiprowski \(2021\)](#). Based on different policy variation (US and Germany, recession and boom) and different measures of search intensity (survey and platform), all three papers find clear evidence that an increase in PBD reduces search effort.

Further evidence comes from studies focusing how search effort evolves around UI exhaustion. Both [Marinescu and Skandalis \(2021\)](#) and [DellaVigna et al. \(2022\)](#) show that search effort increases prior to UI exhaustion and falls afterwards, mimicking the spike in the job finding rate at exhaustion. While the papers do not explicitly estimate the effect of a change in PBD on search effort, they provide comparisons across groups of workers with different PBD and show that the effort spike closely tracks the UI exhaustion point. This also provides strong evidence that PBD has a negative effect on search effort.

Overall, the evidence that $\rightarrow 0$ appears quite strong. By contrast, it is less clear whether UI benefit levels affect search effort. [Krueger and Mueller \(2010\)](#) and [Krueger and Mueller \(2011\)](#) find a negative effect of benefit levels on reported time spent on job search, but this is based on cross-sectional

regressions of effort on benefit replacement rates with extensive controls and without a design based identification strategy. [Kreuger and Mueller \(2012\)](#) find no relationship between benefit generosity and effort in cross country regressions. Finally, [Massenkoff \(2023\)](#) finds no effect of UI benefit levels on search effort in what is arguably the only evidence based on a clear causal identification strategy (RKD).

One important gap in the literature is that there is very little direct evidence on the returns to search effort. One piece of evidence is [Arni and Schiprowski \(2019\)](#) who find that an additional monthly application induced by a higher job search requirement reduces non-employment duration by 4 %. Another is the recent working paper by [Field et al. \(2023\)](#), who run an RCT in Pakistan that reduces the psychological cost of job applications to nudge job seekers to send job applications. They find that an additional application increases the probability of a job interview by about 6 %. They also find evidence for constant returns to search effort (i.e. an elasticity of 1). More indirect evidence on the returns to search, is the structural model in [DellaVigna et al. \(2022\)](#), which involves estimating the search production function π_t . Consistent with [Field et al. \(2023\)](#), they find an elasticity of 1 in the model with the best fit again suggesting constant returns to search. However, an important caveat is that the elasticity estimate is sensitive to alternative specifications and the setting is probably not ideal to identify this elasticity.

Turning to reservation wages, [Table 2](#) lists 4 studies that estimated the effect of UI. Two papers find no effect of PBD extensions on reservation wages ([Le Barbanchon et al., 2019; Lichter and Schiprowski, 2021](#)) using credible research designs (Diff-in-Diff, RD) while two other papers find no effect of benefit levels ([Kreuger and Mueller, 2016; Massenkoff, 2023](#)). While this is not an overwhelming amount of evidence it is fairly consistent: $\frac{\partial \pi}{\partial \text{UI}} = 0$ and $\frac{\partial \pi}{\partial \text{wage}} = 0$. To our knowledge, there is not much evidence whether target wages or other measures of job selectivity respond to UI.

In the standard search model the effect of UI on nonemployment durations goes either through effort or reservation wages. As discussed, there is significant evidence that UI affects search effort and some evidence that effort affects job finding rates. By contrast, it is not clear that UI affects reservation wages or other dimensions of job selectivity. This is also consistent with the fact that the literature has not found consistent positive effects of PBD extensions on reemployment wages, unconditionally or conditional on duration (see [Section 2.3](#)). Therefore this evidence seems to suggest that the effect of UI on nonemployment duration goes mostly through effort and not through job selectivity.

2.5.2 *Directed search vs. reservation wages*

The earliest search models were reservation wage models since they focused on the job acceptance decision as the key choice variable. These models were then augmented with a search effort decision, as in the model laid out in [Section 2.2](#). Similarly to the extent that surveys elicited information on the job search process

they typically asked about reservation wages as well as various measures of search intensity. However, in recent years there has been growing interest in directed job search models (see, for example the review by Wright et al. (2021)) in macroeconomics but also in the literature on the micro-foundations of job search. The interest in directed search models was likely in part driven by the fact that observed reservation wages are not straightforward to reconcile with the theoretical notion of reservation wages. For example, the fact that reservation wages are so high and not falling throughout the spell (Feldstein and Poterba, 1984; Krueger and Mueller, 2016) is confusing given that observed reemployment wages are much lower and falling throughout the spell.

From a theoretical perspective, it is also somewhat puzzling that reservation wages do not respond to UI extensions. In the model (Eq. 5) the reservation wage is directly linked to the value of unemployment $\phi_t = -\delta_t$, since it is, by definition, the wage at which an individual is indifferent between remaining unemployed and taking a job.¹⁵ However given that a UI extension typically increases nonemployment duration τ_t , i.e. $\tau_t > 0$ it has to be that either search effort or the reservation wage responds to τ_t . However both are directly linked to the value of unemployment and the model does not allow for τ_t to be positive without $\phi_t < 0$, therefore $\tau_t > 0$ implies $\phi_t < 0$, which in turn implies $\phi_t < 0$.

In a directed search model workers choose to apply for a specific job with a fixed wage at every point in time. The probability of actually receiving an offer is then a function of the wage target. See Nekoei and Weber (2017) for an example. In general the target wage should also be linked to the value of unemployment and thus respond whenever effort responds to UI. However if the distribution of offered wages has relatively few discrete mass points (e.g., offered wages for a given occupation are mostly constant), workers may not change target wages in response to small changes in UI.

An advantage of the directed search model is that the target wage is in principle easier to observe than the reservation wage either directly or in survey data since it is simply the wage of the last applied for job. For example, in papers that rely on platform data can either proxy the target wage as the posted wage of applied for jobs or impute the wage of the applied for job using occupation and other characteristics.

In general, the predictions of a directed search model with target wages and a reservation wage model are very similar, especially in the absence of reservation wage / target wage data. It seems plausible, that in practice both mechanisms play a role in the real world. Some jobs are posted with and some without wage information and workers have imperfect information about what a given job may pay. Thus job search likely involves a directed search

¹⁵ Using this relationship, Shimer and Werning (2007) argued that estimates of reservation wage elasticity can be used to infer the welfare implications of UI changes.

component where workers have to actively decide what types of jobs to apply for, but also face some uncertainty about what a job offer may look like and thus may reject offers that offers pay that is too low.

2.5.3 Duration dependence in reemployment wages

An important question is whether unemployment duration in itself has a negative, causal effect on job quality. Such a negative effect could arise for 2 reasons, which can be thought of as a supply side or a demand side effect. On the supply side, workers may become less selective with longer unemployment durations, i.e. lower their reservation wage or target wage. On the demand side, the long-term unemployed may be stigmatized by potential employers or they may lose skills that are valued by the labor market, thus facing lower demand for their labor.

To make this more precise let's frame this in the language of the basic search model from above.¹⁶ Denote the reemployment wage of a worker who takes a job at time t as ϕ_t , furthermore, let ϕ_t be the reservation wage in period t and μ_t be the mean of the wage offer distribution ($\phi_t \sim \mu_t$). Finally, assume that ϕ_t is linear in t and that only the mean of the wage offer distribution changes with t . Consider a small increase in nonemployment duration Δt . The effect of this small increase will be through affecting either ϕ_t or μ_t :

$$\frac{\partial \phi_t}{\partial t} = \underbrace{\frac{\partial \phi_t}{\partial \phi_t} \frac{\partial \phi_t}{\partial t}}_{=0} + \underbrace{\frac{\partial \mu_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial t}}$$

What do we know about the two channels? On the supply side, Table 2 provides an overview of the evidence that reservation wage or target wages are falling over the spell. The evidence suggests that reservation wages are falling by up to 5 % per year, though the estimates are not very precise. Note that reservation wages may be falling with duration, i.e. $\frac{\partial \phi_t}{\partial t} < 0$, without impacting reemployment wages. In particular, if the reservation wage is not binding, i.e. below the lowest possible wage offer or $\phi_t = 0$, then changing the reservation wage at t does not affect the reemployment wage, i.e. $\frac{\partial \phi_t}{\partial \phi_t} = 0$. In our basic search model the reservation will be not binding if the wage offer distribution is such that there are no wage offers below the reservation wage ($\phi_t = 0$). Such a reservation wage implies that all offers are accepted, which is an optimal strategy if there is little variance in wage offers, job offers are rare, the person is very impatient, or it is relatively easy to move to better jobs once employed. It may well be that this holds for at least some individuals. Also note that the effect of the reservation wage on

¹⁶The analysis here follows Schmieder et al. (2016). For a similar analysis based on a directed search model see Nekoei and Weber (2017).

the reemployment wage, $\frac{\partial \mu_t}{\partial \phi_t}$, is likely less than 1 unless the reservation wage is at a point in the wage offer distribution where the density is large. Therefore we would expect reemployment wages to fall less than reservation wages. This is in fact confirmed by [Massenkoff \(2023\)](#), who provides some helpful evidence by regressing log reemployment wages on log reservation wages and detailed controls. He finds an elasticity is around 0.54, confirming that reemployment wages should fall less than reservation wages.¹⁷

It therefore makes sense that in [Table 2](#), target wages seem to fall less than reservation wages, between 0 % and 1.5 % per year. Intuitively, target wages represent the typical job a worker might get, while the reservation wage is only the lower bound. Since the lower bound will fall faster than the average reemployment wage, reservation wages would fall faster.

Turning to the demand side, there is some direct evidence suggesting that wage offers may decline with unemployment duration, while other papers find evidence of no impact. As discussed above, some audit studies found lower callback rates for the longterm unemployed ([Eriksson and Rooth, 2014](#); [Kroft et al., 2013](#)), while others found no effect on unemployment durations on callbacks ([Farber et al., 2019](#); [Nunley et al., 2017](#)). There are also 2 papers that directly estimate how cognitive skills vary through the unemployment spell. [Edin and Gustavsson \(2008\)](#) uses panel data from Sweden that were part of the International Adult Literacy Survey. Using 2 waves, 1994 and 1998, that were conducted 4 years apart they can observe about 600 workers over time. They find that workers with unemployment spells had losses in literacy scores and that these losses were larger the longer the individual was out of work. The estimates suggest that 1 year of unemployment reduces literacy scores by about 5 percentiles of the skill distribution. By contrast, using higher frequency data from Germany and a broader measure of skills (cognitive and non-cognitive) ([Cohen et al., 2023](#)) find no evidence of skill depreciation either at the onset or during unemployment.

It is also possible to learn about the relationship between non-employment duration and reemployment wages from UI extensions. Using the basic search model as a framework, [Schmieder et al. \(2016\)](#) show that one can write the effect of a marginal change in t on reemployment wages as.¹⁸

$$\frac{\partial \mu_t}{\partial t} = \underbrace{\left[\frac{\partial}{\partial \phi_t} \frac{\partial \phi_t}{\partial t} \right]}_{\left[\frac{\partial}{\partial \phi_t} \frac{\partial \phi_t}{\partial t} \quad \frac{\partial}{\partial \mu_t} \frac{\partial \mu_t}{\partial t} \right]} -$$

¹⁷ We know from [Krueger and Mueller \(2016\)](#) and others, that reservation wages are high and workers accept reemployment wages below the stated reservation wage. This goes against the model, so then it is not clear what to expect for the magnitude of $\frac{\partial \mu_t}{\partial \phi_t}$. This makes the evidence in [Massenkoff \(2023\)](#) particularly valuable.

¹⁸ [Schmieder et al. \(2016\)](#) provide a more general decomposition that does not rely on linearity.

Thus the effect of a PBD extension on average reemployment wages is the combination of two effects: First, workers become more selective so that the reservation wages increase, which then leads to higher reemployment wages.¹⁹

Second, non-employment durations increase — 0 (either due to changes in the reservation wage or search effort), and thus reemployment wages decline due to the duration effect on wages. In the model always has a weakly positive effect on reservation wages so that: $\left[\frac{\partial}{\partial \phi_t} \frac{\partial \phi_t}{\partial t} \right] \geq 0$. This implies that

$\underline{\mu}_t$ provides an upper bound for the duration effect: $\left[\frac{\partial}{\partial \phi_t} \frac{\partial \phi_t}{\partial t} \quad \frac{\partial}{\partial \mu_t} \frac{\partial \mu_t}{\partial t} \right]$. As discussed before many papers find no effect of PBD on reemployment wages which would suggest that the duration effect is ≤ 0 .

Schmieder et al. (2016) finds a negative effect of PBD on reemployment wages (see Fig. 8A) and provide an estimate for $\underline{\mu}_t$ of -0.8% , which would imply that reemployment wages fall by at least 0.8% per month. The paper further shows that reemployment wages conditional on non-employment duration do not seem to be affected, which then implies $\left[\frac{\partial}{\partial \phi_t} \frac{\partial \phi_t}{\partial t} \right] = 0$, so that the upper bound of the duration effect is in fact the best point estimate. Furthermore, since the reemployment wage does not shift with respect to P (even though theory implies that $\frac{\partial \phi_t}{\partial t} = 0$), this implies that $\frac{\partial}{\partial \phi_t} = 0$ and therefore $\frac{\partial}{\partial \mu_t} \frac{\partial \mu_t}{\partial t} = \underline{\mu}_t$. Thus, this implies that the demand side of the duration effect on reemployment wages induces a -0.8% loss in reemployment wages per month of unemployment (or a 10% loss over a year). Compared to the estimates of $\underline{\mu}_t$, estimates of the effect on reemployment wages $\overline{\mu}_t$, tend to be less precise due to the large variance in reemployment wages. Thus, several papers found negative point estimates of the effect of UI extensions on wages (Card et al., 2007; Centeno and Novo, 2009; van Ours and Vodopivec, 2008), that are not statistically significant. Using a similar framework, Hernandez Martinez et al. (2023) sets out to estimate $\left[\frac{\partial}{\partial \phi_t} \frac{\partial \phi_t}{\partial t} \quad \frac{\partial}{\partial \mu_t} \frac{\partial \mu_t}{\partial t} \right]$ when reservation wages are binding by controlling for the reservation wage shift. After this correction, they find a very similar estimate, namely that reemployment wages are falling by about 0.75% per month of unemployment.

The clearest evidence for the effect of non-employment duration on wages would be an RCT that randomly varies time out of work, without affecting other channels that may affect wages. As the framework in Schmieder et al. (2016)

¹⁹ Note that the expectations operator is with respect to realizations of nonemployment duration t

makes clear, this is hard to do since every instrument that would affect non-employment duration by making unemployment more or less attractive will also affect reservation wages. The closest to such an experiment is perhaps Autor et al. (2015) for the related context of disability insurance (DI) applicants in the US. After applying for DI and before they receive a decision whether the application is approved, applicants cannot work without voiding their application. The paper uses exogenous variation in decision times stemming from randomly assigned examiners for applicants who eventually get denied and thus return to the labor force. They find that a one month increase in processing time reduces long-run annual earnings by about 2.4 %. This is a substantially larger negative effect than the estimates in the UI context imply, which might be due to the special nature of applying for disability or to the specific sample, but it does further underscore the potential negative effect of non-employment on earnings capacity. Overall, the evidence on the effect of non-employment duration on reemployment wages is somewhat mixed. On the supply side the implied estimates range from 0 to -2% per year, depending on whether one goes with the evidence from reservation wages, target wages or the evidence from UI extensions. On the demand side the estimates range from 0 to -10% per year. Many estimates are not very precise and 95 % confidence intervals cover a wide range of possible values. Given the importance of the parameter having more and better evidence would be clearly welcome.

2.5.4 Duration dependence in job finding rates

Policy makers have long been concerned with long-term unemployment, which has long been a challenge in Europe (e.g., Machin and Manning, 1999) and more recently in the US, after the Great Recession (Kroft et al., 2016). A key concern about long unemployment duration is, that being out of work itself has a detrimental effect on the labor market prospects of the unemployed, rendering them increasingly unable to return to work. This would explain the well-documented decline in job-finding rates as time out of work increases. For example the in Section 2.3, we saw that job finding rates often decline by around 40–50 % over the course of a year (see Fig. 7). The notion that non-employment duration has a causal effect on job finding is typically referred to as ‘true duration dependence’. However, whether there is actually true duration dependence, is not obvious. As we saw in Section 2.3.4, the decline in the hazard (and wages) can also be fully explained by heterogeneity between individuals (see Fig. 9) where workers with a high cost of effort and low wage offers remain unemployed longer and dominate the pool of the long-term unemployed. It is therefore hard to identify duration dependence from evidence on job finding rates and reemployment wages alone.

Mueller et al. (2021) and Mueller and Spinnewijn (2023) provide a helpful framework for analyzing duration dependence in job finding rates. Let τ_t be

the probability of finding a job at duration t for an individual. The observed duration dependence at duration t is the change in the average job finding probability from t to t : $\overbrace{t} - \overbrace{t}$. Observed duration dependence can then be decomposed into the change in the job finding rate within individuals and a dynamic selection component:

$$\overbrace{t} - \overbrace{t} = \overbrace{t} - \overbrace{t} + \overbrace{t} - \overbrace{t} \quad (18)$$

This framework suggests two approaches for identifying true duration dependence. Either to directly estimate $\overbrace{t} - \overbrace{t}$ or to estimate the degree of dynamic selection based on the observable worker characteristics.

Regarding the first approach, the main challenge is that one can never observe within-person changes in the job finding rate in the same unemployment spell, since once a job is found the job seeker leaves unemployment. However, recall, that in our basic model the job finding rate can be written as $t = t_t - t \phi_t$. Thus one can gain some insights from the new evidence on within-person changes in search effort and reservation wages. As we saw in [Table 1](#) and the discussion in [Section 2.4](#), search effort appears relatively flat in most studies and thus cannot explain the large observed decline in job finding rates. Similarly [Table 2](#) shows that reservation wages (and target wages) show a only a very small decline, which, as we discussed in [Section 2.5.3](#) likely only explains about a 0–2 % decline in reemployment wages over a year. Thus the behavior of job seekers alone, is unlikely to generate much duration dependence.

Of course, even if behavior does not change much through the spell, duration dependence in the job finding rate could still exist via changes in either the wage offer distribution t , i.e. due to skill depreciation or stigmatization, or the effectiveness of job search t . Perhaps the only evidence clearly suggesting a decline in the effectiveness of search comes from the audit studies by [Kroft et al. \(2013\)](#) and [Eriksson and Rooth \(2014\)](#), though as we discussed before the evidence is somewhat mixed. Similarly, the evidence in the previous section on duration dependence in wages is also somewhat mixed. Furthermore, although at least some studies point to substantial negative duration dependence ([Section 2.5.3](#)) in the order of up to 10 % wage losses over a year, it is not clear how this duration dependence in wages translates into duration dependence in the job finding rate.

The second approach to learning about duration dependence is to estimate the degree of dynamic selection instead in order to infer true duration

dependence from Eq. (18). Mueller et al. (2021) develop this approach in depth and show that the dynamic selection component is equal to the covariance between job finding rates in the current and next period. With this Eq. (18) can be written as:

$$\underbrace{t - t}_{\text{transitory}} = \underbrace{t - t}_{\text{dynamic selection}} - \underbrace{\overline{t} - \overline{t}}_{\text{total variance}} \quad (19)$$

This covariance between current and future job finding rates in turn consists of two components:

$$\underbrace{t - t}_{\text{transitory}} = \underbrace{t}_{\text{dynamic selection}} - \underbrace{t - t}_{\text{total variance}} \quad (20)$$

Thus dynamic selection is a function of the total variance in job finding rate minus the transitory part of this heterogeneity. If the job finding rate is constant within individuals, then the transitory component is zero and dynamic selection is simply a function of the total variance in job finding $t - t = t$. On the other hand if the job finding rate has no constant component and is instead iid, then the variance is equal to the transitory component and there is no dynamic selection.

The individual level job finding rate t is of course unobserved. However, we can learn about these components from the extent to which job finding rates are predictable from individual characteristics \cdot . Note that any predictable variance $t - t$ provides a lower bound for

t . Based on this, Mueller and Spinnewijn (2023) show how with a prediction model for the job finding rate one can estimate a lower bound for the covariance term $t - t - t$ and therefore an upper bound for true duration dependence. To understand the intuition: imagine there are two types of workers, say high and low skill, and we observe that low skill workers have longer unemployment duration than high skill workers. This can only be explained by between group differences in job finding rates, but if such between group differences exist, then the fact that low skill workers take up a larger share of the long-term unemployed in part explains why job finding rates are lower for long-term unemployed. Overall, the better one can predict job finding rates (or non-employment durations), the higher (the lower bound of) the share of observed duration dependence that is due to dynamic selection.

Mueller et al. (2021) show that with high quality data it is indeed possible to predict job finding rates relatively well. While some of the

variation is explained by typical labor market variables, a particularly valuable predictor stems from job seekers information about their job-finding probabilities. They obtain this information from the Survey of Consumer Expectations and the KM survey where workers are asked with what probability they expect to find a job in the future. They show that these elicited beliefs are in fact highly predictive of realized unemployment durations. Using the lower bound implied by Eq. 19, they show that the lower bound for dynamic selection is 52 % when using elicitations only and as high as 89 % when using beliefs and demographics. They also develop a more parametric statistical model, that allows to point-identify the two components of observed duration dependence. They find that heterogeneity explains 84.7 % of the observed decline, leaving 15.3 % of the decline to be due to true duration dependence.

Another way to identify heterogeneity in job-finding rates stems from using data where the same individual can be observed over multiple unemployment spells. Alvarez et al. (2023) use this approach to identify duration dependence and dynamic selection using data from Austria. They find little dynamic selection based on observables, but strong evidence for dynamic selection along unobservable dimensions and argue that dynamic selection accounts for a large share of observed duration dependence.

Mueller and Spinnewijn (2023) study duration dependence using detailed data from Sweden. An advantage of the Swedish data is that they can also observe multiple unemployment spells for the same individual. Extending the Mueller et al. (2021) by also looking at heterogeneity based on multiple spells, they find evidence that would suggest that 84 % of the observed decline in job finding rates is due to dynamic selection, leaving only about 16 % for true duration dependence.

Overall, this recent literature has made a strong case that dynamic selection is important, in particular along unobserved dimensions.²⁰ This leaves limited scope for true duration dependence, which in turn is consistent with the limited evidence of changing search effort and reservation wages / target wages over the unemployment spell.

2.5.5 Present bias vs. exponential discounting

One of the first and consistently documented behavioral biases in economics is the observation that individuals place extra weight on immediate pay-offs relative to the standard exponential discounting applied to a stream of future pay-offs. Building on earlier work by Strotz (1955), Laibson (1997) and O'Donoghue and Rabin (1999) established this notion of ‘present bias’ in the economics literature.

²⁰ Earlier papers that only explored a few observables characteristics concluded that the role of dynamic selection is fairly limited (Krueger et al., 2014; Schmieder et al., 2016; DellaVigna et al., 2017).

[Laibson \(1997\)](#) proposes to characterize the subjective present value of a future stream of utility flows as:

$$U_0 = \beta \sum_{t=1}^{\infty} \delta^t U_t$$

This implies a discount factor between today and the next period of $\beta\delta$, while the discount factor between any two periods in the future is simply δ . For long-run decisions individuals are relatively patient, while impatient over short run decisions. Furthermore, preferences are time-inconsistent: when making decisions about future trade-offs they would exhibit more patience than when making the same decision about an immediate trade-off. E.g., consider a person who is deciding how much to save in a retirement account. If the person faces the decision in the form of putting away a certain percentage in a future paycheck (say in period 1, so that it affects consumption utility U_0), the trade-off is guided by exponential discounting and the person may elect to save. On the other hand when making the decision at the point of receiving the payment, thus affecting consumption utility U_1 , the trade-off is guided by exponential discounting and the present bias factor β , which reduces the value of saving, potentially drastically.

[DellaVigna and Paserman \(2005\)](#) were first to systematically explore the implications of present bias for job search. They start by integrating present bias in the basic search model from [Section 2.2](#). To pin down behavior with present bias requires taking a stance whether individuals anticipate their own present bias for future decisions or not ([O'Donoghue and Rabin, 1999](#)). A sophisticated individual anticipates their own present bias for future decisions and as such has rational expectations. A naive agent incorrectly believes that they will behave as an exponential agent in the future.

Focusing on naive present bias, the value of unemployment of an unemployed individual with present bias becomes:

$$U_t = U_{t-\Delta t} - \beta \delta \left(U_{t-\Delta t} - U_{t-\Delta t + \Delta t} \right) \quad (21)$$

where the future value functions are the value functions of an exponential discounter ($\beta = 1$). A key feature of this problem is that short-run impatience (β) does not affect reservation wages, which are based on the value functions describing future pay-offs, but it does sharply reduce search effort since the pay-off from search is discounted by β , while the search cost is immediate. [DellaVigna and Paserman \(2005\)](#) explore the comparative statics of the search model with respect to patience. With present bias, a more impatient job seeker (smaller β) will search less, the reservation wage is unaffected, and the exit rate will go down. With only exponential discounting, a more impatient job seeker (smaller δ) will search less, but the reservation wage will also go down. The paper argues that under plausible assumptions the reservation wage channel

dominates and impatience actually leads to an increase in the job-finding rate. The paper tests the relationship between patience and the job-finding rate using the NLSY and PSID. They use proxies for impatience, such as having a bank account or smoking, and show that those are associated with a lower exit rate, even when controlling for detailed other characteristics. They take this as evidence for present bias.²¹

Further evidence on present bias stems from structural work. Paserman (2008) estimates a structural model in the spirit of DellaVigna and Paserman (2005) using data from the NLSY on unemployment duration and reemployment wages. By treating groups of workers as homogeneous, he can observe the reservation wage as the lowest accepted wage in that group. This indirect information on reservation wages together with job finding rates pin down the discounting parameters. The paper estimates β to be between 0.4 and 0.9, and δ between 0.996 and 1 (with lower wage workers being less patient). Since the paper nests the standard model ($\beta=1$), the low estimates of β , with relatively tight SEs serve as a rejection of the standard exponential discounting model. Several other recent papers incorporate savings decisions as well as present bias into the search model and provide structural estimates. These recent papers take reemployment wages as fixed so that the choice variables are search intensity π_t and savings, i.e. assets in period t : $\pi_t \pi_{t+1} \dots \pi_T$. The value of unemployment can then be written as:

$$\pi_t \pi_{t+1} \dots \pi_T = \sum_{\pi_t \in 0} \pi_t - \frac{\pi_t}{\pi_t} = \beta \delta \left[\frac{\pi_t}{\pi_t} + \frac{\pi_{t+1}}{\pi_t} + \dots + \frac{\pi_T}{\pi_t} \right]$$

$$\pi_t = \frac{\pi_t}{\pi_t} - \frac{1}{\pi_t}$$

where π_t is the return on savings. The first paper to estimate such a model was DellaVigna et al. (2017) using data from Hungary. The paper focuses on how job finding rates were affected by a unique policy reform that introduced a step function into the UI benefit path. The paper estimates a structural model to explain how the hazard rates and especially the spikes at the exhaustion point respond to the policy reform. A key takeaway is that when endogenous savings decisions are added to the model, then the model can only generate spikes in the hazard rate with a very high degree of impatience. The reason is that patient job seekers would anticipate UI exhaustion and save in advance to smooth the consumption drop, but then, since consumption declines smoothly at exhaustion, there is no sudden increase (spike) in the hazard rate. When estimating the search model with exponential discounting, they obtain $\delta=0.02$ on a biweekly

²¹ One caveat, is that this argument relies on the reservation wage effect dominating the search effort effect for exponential discounters. However, as discussed above, reservation wages in general are not very responsive and potentially not binding, which may cast some doubt on the strength of this evidence.

level, but this would imply an annual discount factor of around 0.06.²² Such a low annual discount factor is inconsistent with many other estimates from the literature as it would imply that individuals have an extremely short planning horizon and would, e.g., never save for retirement or other long-term goals. By contrast when the paper estimates the model with $\beta\delta$ -preferences, the resulting β of 0.58 and δ of 0.995. This implies a yearly discount factor of 0.46 for the first year and 0.88 for subsequent years, which is consistent with other estimates of β the resulting yearly discount factor is high enough to allow for long-term planning. DellaVigna et al. (2022) estimate a similar search model using the results from the SMS survey as moments. Across a number of model specifications they come to the same conclusion. The exponential model results in an implausible low δ and worse fit compared to a model with $\beta\delta$ -discounting, which typically results in estimates of β of around 0.4–0.5 with a plausible δ of around 0.995. As in DellaVigna et al. (2017), present biased is identified by the spike in the hazard rate, which the model can only fit if workers are impatient enough not to smooth consumption too much.

Another piece of evidence for present bias comes from Ganong and Noel (2019), who estimate essentially the same model but use the consumption path during unemployment as well as the job finding rate as moments to identify the model parameters. The key empirical pattern the model tries to rationalize is that consumption drops sharply at UI onset and at UI exhaustion but by a lower degree than income. (Ganong and Noel (2019) show that this is hard to rationalize even with a multi-type model as long as the types all have the same discount factor. The problem is that if workers are patient, they smooth consumption and consumption does not drop sharply at benefit exhaustion. But if workers are impatient, the drop in consumption at exhaustion is as large as the income drop and too large relative to the data. To solve the puzzle, Ganong and Noel (2019) introduce heterogeneity in the discount factor (either in δ or in β). Allowing for 2 types with different discount factors improves the fit of the exponential model. However, the exponential model would again require implausibly low δ . By contrast the model with $\beta\delta$ -preferences fits better with plausible parameter values β between 0.5 and 0.9.

Finally, Gerard and Naritomi (2021) estimate a similar model based on the consumption data from Brazil (see Section 2.4 above). A key innovation is that while the previous papers do not seek to distinguish different forms of present bias (usually focusing on naive, with the exception of Paserman (2008)), Gerard and Naritomi (2021) compare three versions: exponential, naive, and

²² As we discuss in the next subsection 2.5.6, the main innovation of DellaVigna et al. (2017) is to introduce reference-dependence into the search model. In the discussion here, we focus on the version of the model with reference dependence, but the results for the standard model with respect to impatience are qualitatively similar.

sophisticated present bias. Again they find that the exponential model implies implausible levels of impatience ($\delta=0$ on the monthly level). Naive and sophisticated $\beta\delta$ preferences both can generate the broad pattern in their data, but the sophisticated present bias model (with $\beta=0$) obtains a substantially improved fit. They argue that this is because a sophisticated agent anticipates self-control problems in the future, which leads her to save somewhat more in the present.

Overall, there is now very strong evidence for the importance of present bias. Over a wide range of contexts and with different types of empirical moment, models with present bias preferences perform substantially better and with much more plausible parameter estimates than models with only exponential discounting. It is also remarkable, that across all these contexts the estimates for β are quite consistently in a similar ballpark of around 0.5 to 0.7.

2.5.6 Reference dependence

Since [Kahneman and Tversky \(1979\)](#)'s seminal 1979 paper a key insight in behavioral economics has been that individuals evaluate payoffs relative to some benchmark or reference point and that they value losses relative to this reference point higher than gains of the same magnitude (a feature known as “loss aversion”), [DellaVigna et al. \(2017\)](#) introduce reference dependence into the standard search model. The paper models the reference point as the average income over the \geq previous periods:

$$\bar{r}_t = \frac{1}{n} \sum_{i=t-n+1}^{t-1} r_i$$

represents the length of adaption: the higher the longer an unemployed worker feels the loss utility from a drop in consumption.²³ Using this reference point, the paper models the utility from consumption as:

$$u_t(r_t) = \begin{cases} r_t - \eta & r_t \geq \bar{r}_t \\ \eta \lambda & r_t < \bar{r}_t \end{cases} \quad (22)$$

The utility function consists of consumption utility $u_t(r_t)$ and gain-loss utility $u_t(r_t) - u_t(\bar{r}_t)$. For consumption levels above the reference point ($r_t \geq \bar{r}_t$), individuals receive gain utility $u_t(r_t) - u_t(\bar{r}_t) \geq 0$, with a weight η . For consumption below the reference point ($r_t < \bar{r}_t$), the individual suffer loss utility $u_t(r_t) - u_t(\bar{r}_t) < 0$ with weight $\lambda\eta$. The parameter η captures the importance of gain-loss utility, while the parameter $\lambda \geq 0$ captures loss aversion. The standard search model is nested in this model for $\eta=0$.

²³ For a discussion of alternative assumptions for the reference point, such as expectations based reference points as in [Kőszegi and Rabin \(2006\)](#), see [DellaVigna et al. \(2017\)](#).

The unemployed choose search intensity τ_t and consumption c_t in each period. The value function of a job seeker can then be expressed as:

$$U(c_t, \tau_t) = U(c_{t-1}, \tau_{t-1}) - \delta \left[\frac{c_t}{\alpha} + \frac{\tau_t}{\beta} - \frac{c_{t-1}}{\alpha} - \frac{\tau_{t-1}}{\beta} \right]$$

$$\tau_t = \tau_{t-1} - \frac{c_t}{\beta}$$

The model makes predictions about the job finding rate that are consistent with the observed patterns in [Section 2.3](#): Newly unemployed workers experience strong loss utility and are thus eager to find a job, resulting in a high exit rate. As they stay unemployed, the reference point adapts to the lower income level and workers get accustomed to being unemployed and search effort and the job finding rate decline. As workers approach UI exhaustion point, however, they face a large drop in consumption relative to their reference point and search effort should increase up to the exhaustion point. However after benefits are exhausted, adaption sets in again among individuals who are still unemployed and search effort declines. As a result the time path in search effort should look very similar to the typical hazard rates documented before. However, we saw in [Section 2.3](#), that the basic search model with enough heterogeneity can also generate an initial decline in the exit rate, followed by a spike at the exhaustion point. In order to test for the empirical importance of reference dependence, [DellaVigna et al. \(2017\)](#) therefore turn to a unique reform in Hungary in 2005. Prior to the reform UI benefits were constant for 9 months. After the reform UI benefits were increased in the first 3 months and reduced for the following 6 months. From month 10 onwards there was no change. The crucial difference between the standard and reference dependent model is, that the standard model is purely forward looking. Thus conditional on not finding a job in the first 9 months, the standard model predicts the same job finding rate before and after the reform. By contrast in the reference dependent model, the reference point creates a backward looking mechanism. As such, after 9 months a worker in the pre-period would have a higher reference point than a worker in the post period with the same non-employment duration (but lower income in the previous 6 months). [DellaVigna et al. \(2017\)](#) show that this pattern indeed holds in the data. To provide a formal test between the reference dependent and the standard model, they estimate the model structurally with and without reference dependence. They show that the reference dependent model provides a much better fit than the standard model and in particular can explain, quantitatively and qualitatively, the higher hazard rates from 9 months onwards in the post-period, which the standard model fails to match. Another area where the standard and reference dependent model create different predictions is how search effort would evolve within individuals

over time. We saw in [Section 2.2](#) that if the only source of nonstationarity is UI exhaustion, then the standard model generates search effort that is increasing over time until the exhaustion point and then remains constant. The reference dependent model by contrast predicts a spike in search effort at UI exhaustion. The reference dependent model also predicts a reduction in effort at the beginning of the spell, but this may be dominated by the approaching exhaustion point (and thus the need to search more), so that this prediction is less distinct. As discussed in [Section 2.4](#), the two studies that provide clear evidence of the within person evolution of search effort around UI exhaustion [DellaVigna et al. \(2022\)](#), [Marinescu and Skandalis \(2021\)](#) both found a spike in effort at the exhaustion point consistent with the reference dependent model. It should be noted though that there could be other sources of nonstationarity, for example searching may become more costly over time (τ_t) or less productive (η_t). Furthermore skill depreciation (δ_t) may lead to lower job quality over time and thus may make search more attractive initially and less attractive later on. These other factors might also explain a decline in search effort over time and lead to a pattern resembling a spike in effort around the exhaustion point. [DellaVigna et al. \(2022\)](#) provide structural estimates of the standard search model without nonstationarity (apart from UI), the referenced dependent model and a model with time-varying search cost (calling it the 'discouragement' model). The estimates suggest that the standard model alone cannot fit the data (in particular the spike in effort), while both the reference dependent model and the discouragement model provide a good fit. The discouragement model can only do so however by also allowing for a very elastic (convex) search production function τ_t , with an elasticity of close to 2, which may seem implausible. On the other hand, the reference dependent model requires only an elasticity of 1, i.e. job finding is proportional to effort. The best fit is provided by a model with both reference dependence and search effort.

Reference points have also been proposed as a determinant of reservation wages. In principle, if pre-unemployment wages serve as a reference point, this could be an explanation for the relatively high reservation wages (on average close to the pre-unemployment wage) found in the literature. Reference dependence has also been suggested as a way to reconcile the cyclical properties of accepted and reservation wages ([Koenig et al., 2016](#)). There is also some experimental work consistent with this (see the discussion in [Cooper and Kuhn \(2020\)](#)). On the other hand, if previous wages would serve as a reference point, one would expect bunching in the post-unemployment wage distribution at the pre-unemployment wage (similar to bunching at round finishing times for marathon runners, see [Allen et al. \(2017\)](#)), which to our knowledge has not been found.²⁴

²⁴We are not aware of published work, but know informally that people have looked for this.

2.5.7 Biased beliefs

A key determinant of how much a job seeker searches for a job is how they perceive how their effort t translates into the probability of finding a job π_t . It seems plausible that workers have only a vague idea of what π_t actually looks like: job offers are rare and most workers have limited experience with unemployment over their life. It's also likely that the exact shape of π_t will be very individual specific so that it is limited how much one can learn from other. Let δ_t be the perceived job search productivity of a worker. The value of unemployment (assuming fixed wages) then becomes:

$$u_t = \frac{1}{\delta_t} - t - \delta_t \left[\begin{array}{cc} \pi_t & \pi_t \end{array} \right] - t \left[\begin{array}{c} \pi_t \\ \pi_t \end{array} \right]$$

This perceived job search production function may be biased relative to the true production function. [Spinnewijn \(2015\)](#) classified such misperceptions into two categories:

Baseline-optimistic (-pessimistic): A baseline optimistic job seeker overestimates the probability of finding a job at a given search effort (or at all levels of search effort: $\delta_t' > \delta_t$). Similarly, a baseline pessimistic person underestimates the probability of finding a job at a given search effort (or at all levels of search effort: $\delta_t' < \delta_t$).

Control-optimistic (-pessimistic): A control optimistic person overestimates the marginal return of effort: $\delta_t' > \delta_t$, while a control pessimistic worker underestimates the marginal return: $\delta_t' < \delta_t$.

Biased beliefs can lead to search effort being inefficiently too high or too low from the individuals perspective. The first order condition for search effort is:

$$\delta_t' = \delta_t \left[\begin{array}{cc} \pi_t & \pi_t \end{array} \right]$$

It is straightforward that a control-optimistic job seeker will search too much, since they overestimate the returns to their effort, while a control-pessimistic job seeker will search too little. Interestingly, a baseline optimistic worker with correct beliefs about the marginal returns to search effort will also search too little. The reason is that the baseline optimistic worker overestimates the probability of finding a job in future periods and thus overestimates δ_t' , but a higher value of unemployment leads to lower effort. Similarly, a baseline pessimistic worker will search too much. Whether or not workers have biased beliefs has potentially important policy consequences. For example, if job seekers underestimate the true productivity of search effort, interventions that unbias beliefs, such as providing information about job finding prospects could be welfare enhancing and potentially very cheap. Furthermore, [Spinnewijn \(2015\)](#) shows how the typical sufficient-statistics Baily-Chetty approach to calculating the welfare effects of UI reforms has to be modified to account for biases in beliefs.

There is relatively strong evidence that job seekers overestimate the probability of finding a job over a specified time period. For example, [Spinnewijn \(2015\)](#) reports a survey where the average job seeker expected to remain unemployed for an additional 6.8 weeks, while the true duration ended up about 23 weeks. This is clear evidence for baseline-optimism. Similar evidence for baseline-optimism can be seen in [Mueller et al. \(2021\)](#), where workers overestimate the probability of finding a job over the next 3 months and this overestimation actually increases for the long-term unemployed.

By contrast, it is difficult to obtain estimates of control-optimism/pessimism for two reasons: First, it is arguably harder to survey beliefs about marginal returns to effort than it is to ask about levels and second, it is hard to obtain unbiased estimated of the return to effort. Coming up with innovative designs either in the lab or the field could be a fruitful avenue for future work.

2.5.8 Locus of control

Biased beliefs are closely linked to the psychological concept of locus of control. Locus of control refers to the subjective belief of an individual that their actions determine important life outcomes. Psychologists refer to a person who believes that outcomes are determined by factors outside of their control as having ‘external locus of control’. On the other hand a person who believes that they have control over their life is referred to as having an internal locus of control. In the job search context, a person with internal locus of control would have relatively high β_t , while a person with external locus of control relatively low β_t . Note that this notion of locus of control is independent from whether the perceived differences in locus of control correspond to real differences (i.e. $\beta_t = \beta_t$) or due to biased beliefs (such that an internal locus may correspond to control-optimism, while an external locus to control-pessimism).

There is some evidence that suggest that locus of control is indeed predictive of job search behavior: [McGee \(2015\)](#) uses NLSY data to show that a more internal locus of control is associated with somewhat higher reservation wages and more time spent on search. A similar analysis is providec by [Caliendo et al. \(2015\)](#) using the IZA evaluation dataset of unemployed workers in Germany. They also report that higher internality is associated with higher search effort and higher reservation wages. They also find that job finding rates are indeed higher among workers with a more internal locus of control, suggesting that the search effort effect dominates the reservation wage effect.²⁵

²⁵ See [Cooper and Kuhn \(2020\)](#) for a longer discussion of Locus of Control and how it relates to other concepts in psychology.

2.5.9 Employer collusion / storable offers

In most models of job search, worker search and if they accept a job offer, they start the job in the following period. In practice, this process may not be as simple: negotiating the terms of a contract may take time leading to a delay between a job offer and job acceptance. Furthermore, there may be some flexibility about the job start date, which may be some time after the job is accepted. [Boone and van Ours \(2012\)](#) propose that negotiations over the job start date could explain some of the patterns of the observed exit hazards from unemployment and, in particular, the spike at UI exhaustion. Consider a worker who is relatively content being unemployed, while on UI receives and prefers this to working, while also wanting to avoid being unemployed beyond the UI exhaustion point. Suppose such a workers receives a job offer 2 months before UI exhaustion. In this case they might prefer to 'store the offer', i.e. stay on UI until they exhaust their benefits and start working then. Thus the worker might negotiate a job start date to coincide with benefit exhaustion. In such a world, even if job offers arrive at a constant rate, the unemployment exit rate would be lower prior to UI exhaustion, then spike right at the exhaustion point and fall again afterwards (when all offers lead to job start dates as early as possible).

[Boone and van Ours \(2012\)](#) develop such a 'storable offers' search model where job start dates can be delayed if job seekers and employers agree. Importantly employers are more likely to accept such delays, for job offers with permanent contracts which have higher value to the employer and where the cost of short term delays is relatively less important. The model predicts a spike in unemployment exit at the UI exhaustion point and a larger spike for exits into permanent jobs than for exits into temporary jobs. The paper tests this prediction in Slovenian data and finds a clear spike at 3 different exhaustion points (for 3 different PBD groups) for exits to permanent jobs but a much more muted spike for exits to temporary jobs. This evidence is suggestive of the role of strategic delays. However the probability of accepting a permanent contract might be correlated with other worker characteristics. We saw in [Section 2.3.4](#) that the basic search model can easily generate spikes in the hazard rate with enough heterogeneity. Therefore if worker types are sufficiently flexible and correlated with contract type, the basic model should also be able to generate different spikes for workers with permanent and temporary contracts.

Another way to test for strategic delays in job start dates is to directly measure the time gap between a job offer and a job start date. The storable offer model predicts that this gap should be larger for jobs that start at UI exhaustion, compared to jobs that start before or after. [DellaVigna et al. \(2022\)](#) calculate the job starting gap using the SMS survey data, which asked about job offer, acceptance and start dates. They show that the typical gap between job offer and start is around 30 days. For individuals starting a job in the month of UI exhaustion, the gap is only slightly larger with around 31.5 days, not

enough to explain the large spike at the exhaustion point. They also provide an alternative test for the storable offer mechanism. If individuals delay job start dates to start at the UI exhaustion point, then for people exiting at the exhaustion point search effort should be lower in the weeks prior to the job start compared with people exiting after the exhaustion point. While they show that search effort declines about 3–4 weeks prior to a job start, the pattern is virtually identical for job start dates in the exhaustion month as for other job start dates.

Another form of collusion between workers and firms that could explain the spike in the hazard at UI exhaustion could occur in the case of temporary layoffs that result in recall. For example, [Katz and Meyer \(1990\)](#) find a clear spike in recall rates at UI exhaustion. In principle, recalls can be quantitatively important: [Fujita and Moscarini \(2017\)](#) report a 50 % recall rate for the United States, while [Nekoei and Weber \(2015\)](#) report a 35 % recall rate in Austria (though other papers have found lower recall rates in other countries such as Hungary [DellaVigna et al. \(2017\)](#) or Germany [DellaVigna et al. \(2022\)](#)). However, recalls do not explain the spike in the hazard at UI exhaustion by themselves. [Katz and Meyer \(1990\)](#) shows that the spike for unemployment exits to new employers is larger than for recalls and [DellaVigna et al. \(2017\)](#) and [DellaVigna et al. \(2022\)](#) both show that the spike and general pattern of the job finding rate is virtually identical when including or excluding recalls.

Overall, while there is probably some collusion leading to the timing of job start dates, either for jobs with new employers or recalls, the magnitude of this seems very limited and this channel is unlikely to play a large role in either explaining the spike at exhaustion or the disincentive effect of UI. Another form of collusion between workers and employers in the UI context can occur with respect to the timing of job separations to exploit features of the UI system, e.g., to allow a bridge into retirement. We will return to this in [Section 3](#).

2.5.10 Learning / information

The standard search model assumes that workers are fully informed about the job offer arrival rate (or the productivity of job search) and the wage offer distribution. This may well be unrealistic given that these are hard to observe and likely exhibit substantial variation between individuals and across time. It seems possible then that workers have imperfect information about the job search process at the beginning of the unemployment spell and gradually learn throughout the spell as they observe the rate of job offers and their associated wages.

[Potter \(2021\)](#) develops a model where job seekers are imperfectly informed about the stochastic nature of the job search process. He assumes a search production function that takes on a Poisson process:

$$_t = - \lambda_t$$

which describes the probability of receiving a job for a given time spent on job search t , where λ captures the effectiveness of job search. Workers do not know the true value of λ , but instead form beliefs which take on the form of a gamma distribution with two parameters α and β , describing the mean and variance of the true location of λ . While search productivity is thus unknown, workers have perfect information about the wage offer distribution. Workers update their beliefs about search productivity based on how much they are searching and whether or not they receive offers. The longer a worker is searching without receiving an offer the more she updates her belief that the productivity of job search λ is low. If she receives an offer her beliefs about λ increase. Thus as workers remain unemployed and search without offers, they become more pessimistic about the productivity of search effort. This pessimism has two impacts on search effort, it reduces the perceived productivity of search, which leads to lower search effort, but it also reduces the value of unemployment since search in future periods is also less productive. The paper shows that the first effect typically dominates and that this leads to a reduction in search effort and reservation wages over the course of the unemployment spell. The main piece of empirical evidence provided by the paper is that the model predicts that learning is a function of time spent on job search, not on unemployment duration by itself. Thus of two individuals with the same unemployment duration, the one who has spent more time searching in prior periods without obtaining an offer should exert less effort today. On the other hand individuals who have received (and rejected) job offers in the past positively update their beliefs about the productivity of search and search more today. [Potter \(2021\)](#) tests this prediction in the KM survey and indeed finds that job search in the current period is negatively associated with job search in prior periods and positively associated with previous job offers. The paper also goes on to estimate the model structurally and finds that job seekers at the beginning of the unemployment spell overestimate their job finding prospects by about 60 %.

An alternative explanation for a negative effect of past search on present search could be stock-flow matching ([Coles and Petrongolo, 2008](#)). Suppose that for a given job seeker there is a finite set of vacancies that match her qualifications or fit her interests. When becoming unemployed a worker can apply to all jobs among this stock of vacancies. However, once the job seeker has applied to the whole stock, going forward she is constraint by the available flow of new vacancies that open up in her field. A worker who searches more in one period may be more likely to exhaust the stock and thus be forced to search less in the next period. The stock-flow model can explain the negative effect of past search on current search, but does not explain why past offers have a positive effect on current search. However having more direct evidence on learning and stock-flow models would be very helpful.

2.6 Discussion

The advent of high-quality administrative data with credible, causal research designs has revealed several clear stylized facts about the job search process. However, such data alone is not sufficient to learn about important underlying mechanisms that drive job search. The arrival of a broad array of data that sheds light on the underlying mechanisms of job search has dramatically deepened our understanding.

Some of the core lessons are that search effort is probably more important in shaping search outcomes than reservation wages. Job seekers likely exhibit present bias that leads to too little search. Dynamic selection is an important driver of changes in aggregate hazards and reemployment wages, limiting the extent to which there is true duration dependence. Several other refinements to the search model have been proposed and found some supportive evidence, such as reference dependence, learning, or biased beliefs about search effort productivity.

Why do job seekers spend so little time on job search? Papers have shown time and again that unemployed workers spend only about 60–90 min per day on job search, a tiny fraction of the time they would work at a job. Either the (perceived) returns to searching more must quickly diminish within a day or the marginal cost of search must increase rapidly as workers search even just above an hour. The problem is that low returns to search explanations appear at odds with the fact that stronger incentives via lower UI, clearly affect search outcomes. On the other hand it is not clear why the cost of job search should go up so fast to keep search effort at such a low level. Such a high cost may come from psychological factors. For example, [Ahammer and Packham \(2023\)](#) find big negative effects of unemployment on mental health. Being unemployed and receiving rejections may also impact workers self-esteem, a theory proposed by [Kőszegi et al. \(2022\)](#). More rigorous research on these outcomes and how they shape job search would be very interesting.

3 Design of UI policy

Unemployment Insurance (UI) provides income replacement to workers who lose their jobs involuntarily. This section discusses the design of Unemployment Insurance. We first present standard frameworks of optimal UI initially introduced by [Baily \(1978\)](#) and extended by [Chetty \(2006, 2008\)](#). Second, the section shows how to quantify the welfare effects of UI, with a specific emphasis on the most recent estimates of the social value of UI. It compares various UI policies using new estimates of the Marginal Value of Public Funds of UI policies in the US and in Europe (MVPFs, see [Hendren and Sprung-Keyser \(2020\)](#)). Third, it discusses how to take into account effects on wages, and on pre-unemployment separation rates when designing UI programs. Fourth, the section asks whether benefits should vary within the

unemployment spell or over the business cycle, and discusses how UI interacts with other social policies. Fifth, it discusses macro effects of UI. Overall, we draw seven lessons from the section. Before presenting the Baily-Chetty framework, we recall briefly the main institutional features of UI policies.

3.1 The structure of unemployment insurance policies

Unemployment Insurance provides income replacement to workers who lose their jobs involuntarily.²⁶ In general, not all job losers are eligible for UI benefits. To be eligible, workers must have worked for a minimum period or earned a minimum amount of wages (and thus have significantly contributed to the UI fund). For example, in France, the previous work requirement amounts to six months over the two years before separation (2024 rules). In the US, the previous work requirement is known as the monetary requirement. In California, in 2024, workers must have earned at least \$1300 in one of the quarter of the year before losing their job. To be eligible, workers must also satisfy a non-monetary requirement. They must be deprived of work involuntarily, because they have been laid off. Job quitters and workers fired for misconduct are not eligible for UI benefits (in some countries, they may be after a waiting period). They must be searching for jobs actively.

When eligible, and conditional on registering their claim, UI claimants receive weekly or monthly benefits for a fixed period of time. The level of benefits is usually set as a fraction of previous wages, and subject to a maximum amount. The corresponding replacement rate varies across countries (e.g., around 80 % in Sweden vs. around 60 % in France) and across workers in the same country. The replacement rate generally decreases with pre-unemployment wages (either by design or because there are caps at maximum benefit level). The fixed period of time during which benefits are paid is known as the Potential Benefit Duration (PBD). The PBD may also vary as a function of pre-unemployment work experience. In some exceptional cases, there is no exhaustion of UI benefits after a fixed period (e.g., in Sweden in the 2000s, or in Belgium). While in simple UI systems, benefits are constant within the claiming period (and until the end of the PBD), some countries implement more complex schedules where benefit levels decrease with unemployment duration (for example in Hungary, Sweden, or Spain).

In this section, we discuss how to choose both benefit levels and PBD. We present next how welfare considerations can guide the policy choice.

3.2 The welfare effects of unemployment insurance

The design of Unemployment Insurance is guided by its effects on workers' welfare. Unemployment Insurance increases workers' welfare as it provides

²⁶ See Schmieder and von Wachter (2016) for a more detailed description of the structure of UI across various countries.

replacement income and allows workers' to smooth consumption over time. Unemployment Insurance also bears welfare costs. As highlighted in the previous section, generous unemployment insurance slows job finding theoretically and empirically. The workers' behavioral response increases government spending on benefits, which is costly. It requires to increase taxes to satisfy the government budget constraint and in turn, reduces workers' welfare. Unemployment Insurance generates a fiscal externality. [Baily \(1978\)](#) first highlighted those tradeoffs in his seminal work. We describe a simple version of the [Baily \(1978\)](#) approach. We discuss the general application of the Baily approach as advocated by [Chetty \(2006\)](#). We introduce the Baily-Chetty approach in a dynamic environment, which allows to compare the welfare effects of various UI policy parameters (as in [Schmieder and von Wachter \(2016\)](#)).

3.2.1 Baseline framework

The simple model has one single period. At the beginning of the period, workers are unemployed and receive benefits β . They choose search effort which pins down their job finding rate, and sets the expected length of their unemployment spell $\bar{\tau}$. Workers face convex increasing search costs: $\psi(\bar{\tau}) \geq 0, \psi''(\bar{\tau}) > 0$. When employed, workers receive a gross wage \bar{w} and pay taxes τ . In the simple model, we assume that wages are fixed. This is an important departure of the simple Baily framework from the classical job search model of the previous section. We discuss this point later.

Workers derive utility $U(\bar{\tau}, \bar{w})$ when unemployed and $U(\bar{\tau}, \bar{w} - \tau)$ when employed. The utility functions can be different between both states, but they are both increasing and concave in consumption. In the simple model, workers consume their state-specific income and cannot transfer income across states and do not have asset. Workers then solve the following problem:

$$U(\bar{\tau}, \bar{w}) - U(\bar{\tau}, \bar{w} - \tau) = \psi(\bar{\tau}) \quad (23)$$

Workers' optimal behavior is characterized by the first order condition:

$$\frac{\partial U}{\partial \bar{\tau}}(\bar{\tau}, \bar{w}) - \frac{\partial U}{\partial \bar{\tau}}(\bar{\tau}, \bar{w} - \tau) = \psi'(\bar{\tau}) \quad (24)$$

It states that at the optimum, the marginal cost of search equals the marginal return of switching to employment. It captures the same tradeoff as the more involved dynamic first order condition of the job search model ([Eq. 6 in Section 2.2](#)). [Eq. \(24\)](#) implicitly defines the optimal search effort, and consequently unemployment duration as a function of unemployment benefits: $\bar{\tau}^*$. Differentiating [Eq. \(24\)](#) yields that search effort decreases with benefits

$$\frac{\partial \bar{\tau}^*}{\partial \beta} = -\frac{\psi''(\bar{\tau}^*)}{\psi'(\bar{\tau}^*)} < 0.$$

The social planner chooses the level of benefits β and of taxes τ to maximize workers' welfare. The social planner takes as given workers' behavioral reactions to unemployment benefits (she cannot enforce search effort directly). It is subject to a budget constraint, where benefits are financed through taxes:

$\tau = \tau$. Formally, the social planner solves the following problem:

$$\tau - \tau = -\tau - \psi \quad (25)$$

such that

$$\begin{cases} -\tau - = \psi' \\ - = \tau \end{cases}$$

We note that the budget constraint defines taxes as a function of benefits: $\tau = -$. This allows to write workers' welfare as a function of benefits only and simplifies solving the social planner's problem. Workers' welfare derivative wrt τ then writes:

$$\begin{aligned} \frac{\partial}{\partial \tau} &= - -' - -' - \tau \frac{\partial}{\partial \tau} \\ &= \underbrace{-\psi'}_{=0} - \tau - \end{aligned} \quad (26)$$

where the last term is zero because of the first order condition of the workers' program (Eq. 24). To write τ , we differentiate the government budget constraints. To keep the budget balanced after a benefit increase, the social planner needs to increase taxes through two channels. First, holding workers search effort constant, taxes have to increase by $-$. Second, as workers spend more time unemployed, taxes must be further increased by $-$. Replacing this expression of τ in Eq. (26), we obtain:

$$\frac{\partial}{\partial \tau} = - -' - -' - \tau \frac{\partial}{\partial \tau} - \quad (27)$$

$$\frac{\partial}{\partial \tau} = \underbrace{- -'}_{-\eta} - \underbrace{- -' - \tau}_{-\eta} - \underbrace{\frac{\partial}{\partial \tau} - \tau -}_{-\eta} \quad (28)$$

where η is the elasticity of unemployment duration ($-$) with respect to the benefit level $-$. The first term corresponds to welfare gains. As the marginal utility when unemployed is higher than the marginal utility when employed ($' - \tau$), transferring income and consumption from the employment state to the unemployment state increases workers' welfare. On the other hand, the second term is negative (recall that $\eta < 0$, as > 0). This corresponds to the welfare cost of providing insurance due to the fiscal externality. As unemployment duration increases, extra taxes are levied on wages and workers' welfare when employed decreases. At the optimum, the social planner chooses the benefit level so that marginal welfare gain and cost are equal:

$$\frac{\partial}{\partial \tau} = -\eta \quad (29)$$

where (resp.) is the consumption when unemployed (resp. when employed). Eq. (29) is known as the Baily-Chetty formula. This formula provides a direct mapping between theoretical welfare effects and empirical counterparts. Many studies estimate the elasticity of unemployment duration wrt benefit generosity. The previous section describes some of them in details, and we discuss the order of magnitude of elasticity estimates from a wider review of the literature in the next section. Estimating the welfare gains involves quantifying the marginal utility change from employment to unemployment. We discuss empirical strategies to identify this change in the next section.

The baseline Baily framework makes strong assumptions on workers' behavior, for example about their access to other consumption smoothing instruments. Following Baily (1978), several studies (Brown and Kaufold, 1988; Flemming, 1978; Lentz, 2009) successfully enriched the underlying job search model showing that the optimal level of benefits depends on various primitives. Those papers eventually quantify the optimal level of UI generosity performing structural estimation of the underlying model. While they provide important and relevant quantifications, including that of deep primitive parameters, Chetty (2006) shows that the reduced-form quantification of the Baily formula is actually sufficient to assess UI optimality. Under a general class of models, the Baily-Chetty formula holds.

To illustrate Chetty (2006) point, we take the example of allowing workers to borrow against their future wages. Suppose that in the augmented model, consumption when unemployed is $=$ with 0 and consumption when employed is $= -\tau -$. Workers now choose both search effort and borrowings $.$ This yields an extra first order condition related to borrowing choice $- \tau' = \tau'$, while the first order condition related to search effort remains the same. In the augmented model, the social planner also takes into account that workers choose to smooth consumption over states. However, as workers already optimize over their borrowings, the envelope condition holds and the first order condition of the social planner remains the same as in the baseline model. This implies that the Baily-Chetty formula holds. Intuitively, for any given level of benefits $,$ the change in marginal utility across states in the augmented model is lower than in the baseline model. Consequently, the optimal level of benefit in the augmented model may be lower than in the simple model. That being said, the optimal level of benefits is such that the Baily-Chetty formula holds. It remains sufficient to identify two statistics in the data—the change in marginal utility across states and the elasticity of unemployment duration wrt benefit level—to test whether unemployment insurance is optimally set.

3.2.2 Dynamic framework

While the static version of the Baily-Chetty formula captures the key trade-off inherent in formulating optimal UI policy, it does not directly speak to the

design of Potential Benefit Duration, a key policy variable in practice. For this reason, it is useful to consider a dynamic version. Here, we closely follow the model in Schmieder and von Wachter (2016), which delivers the dynamic Baily-Chetty style formula for both changes in benefit levels (Chetty, 2008) and for PBD extensions (Schmieder et al., 2012).²⁷ The Unemployment Insurance policy is implemented through two main parameters: benefit level and Potential Benefit Duration. Under standard UI rules, the flow of UI benefits β_t is equal to constant β until unemployment duration reaches the maximum PBD τ when it drops to 0. The dynamic framework allows to characterize both the optimal benefit level and optimal PBD.

The model is set in continuous time. We consider workers becoming unemployed at date $t=0$, and we denote π_t the value function at the beginning of the spell. At each date t , unemployed workers search with intensity ψ_t , normalized so that it represents the instantaneous job finding rate. Searching with intensity ψ_t incurs cost of $\psi_t \pi_t$, which may vary over time. As in the baseline model, wages are fixed and equal to $\beta - \tau$. Jobs are permanent over the problem time horizon τ . As previously, workers are hand-to-mouth. The consumption when unemployed c_t is equal to their benefits at date t : $c_t = \beta_t$. The consumption when employed is $c_t = \beta - \tau$. The value function of the unemployed writes:

$$\pi_t = \int_0^t \left[\beta_{t-s} - c_{t-s} - \psi_{t-s} \right] ds \quad (30)$$

where $\pi_t = \pi_0 - \int_0^t \psi_s ds$ is the survival rate of unemployed workers. Given a sequence of job finding (or exit) rates from date 0 to date t , ψ_t is the share of the initial unemployed pool still searching for jobs at date t (i.e. the survival rate until time t). Workers choose the search effort sequence ψ_t to maximize their welfare π_t . Let us denote π^* workers' welfare under the optimal search effort. We rewrite the welfare function to make the dependence in the policy parameters explicit:

$$\pi^* = \int_0^\tau \beta_t - c_t \int_t^\tau \beta_{t-s} ds - \int_0^\tau \beta_t - \tau \int_0^\tau \psi_{t-s} ds \quad (31)$$

where β_t and c_t are to be understood as the optimal workers' choice from now on.

As previously, the social planner maximizes workers' welfare under the budget constraint. Let us denote the expected duration of receiving benefits $\tau = \int_0^\tau t \psi_t dt$ and the expected non-employment duration $\tau' = \int_0^\tau t \pi_t dt$. The budget constraint is $\tau = \tau' - \tau$, which defines τ as a function of β and

²⁷ The key simplification in Schmieder and von Wachter (2016) is to assume hand-to-mouth consumers and to set the model in continuous time. The resulting Baily-Chetty formulas are however virtually identical to the general case with endogenous consumption and capture the same intuition.

implicitly of τ through both expected duration τ' and τ . The social planner problem writes:

$$\frac{\tau}{\tau} = \frac{\tau'}{\tau} - \frac{\tau}{\tau} \quad (32)$$

We first differentiate the social planner objective wrt the benefit level. As previously, any endogenous change in search effort has no effect on marginal welfare (envelope theorem). We obtain the following expression for the change in welfare:

$$\frac{\partial}{\partial \tau} = \frac{\tau'}{\tau} - \frac{\tau'}{\tau} - \frac{\tau}{\tau} - \tau \frac{\partial}{\partial \tau} \quad (33)$$

From the budget constraint, we have $\frac{\tau}{\tau} = \frac{\tau'}{\tau} - \tau$. After rearranging and rescaling, the marginal welfare effect of an increase of \$1 in instantaneous benefit (expressed in marginal utility of employed workers) writes:

$$\frac{\partial}{\partial \tau} \left(\frac{\tau}{\tau} - \tau \right) = \underbrace{\frac{\tau'}{\tau} - \frac{\tau'}{\tau} - \tau}_{\frac{\partial}{\partial \tau}} - \underbrace{\left(\frac{\tau}{\tau} - \tau \right)}_{\frac{\partial}{\partial \tau}} \quad (34)$$

Following a \$1 increase in instantaneous benefits, the unemployed receive a total mechanical transfer of τ dollars (over the covered unemployment spell without behavioral changes). As usual, transfers related to behavioral changes do not contribute to welfare. The value of the mechanical transfer depends on the difference between the marginal utility of unemployed and employed. In sum, compared to the welfare analysis in the static model, welfare gains are unchanged. On the contrary, the expression of welfare costs now involves two channels. When they slow down job finding, unemployed receive extra benefits that lead the social planner to raise taxes (by τ). This channel is similar as in the static model. However, in the dynamic model, there is an extra behavioral cost related to the higher share of unemployed whose benefits exhaust. Even if they do not receive extra benefits after exhaustion, they are not employed and do not pay taxes. Consequently, the social planner further increases taxes on the employed to balance the budget (by $-\tau$).

We now turn to the first order condition related to an increase in Potential Benefit Duration (τ'). The detailed derivation is reported in the online Appendix. The marginal welfare effect of an increase in PBD (also expressed in marginal utility of employed) writes:

$$\frac{\partial}{\partial \tau'} \left(\frac{\tau}{\tau} - \tau \right) = \underbrace{\frac{\tau'}{\tau} - \frac{\tau'}{\tau} - \tau}_{\frac{\partial}{\partial \tau'}} - \underbrace{\left(\int_0^{\tau} \frac{t}{\tau} t - \tau \right)}_{\frac{\partial}{\partial \tau'}} \quad (35)$$

where we define $\tau' = \tau - 0$. When increasing PBD, the social planner transfers income to all workers who would have exhausted their benefits otherwise. As those workers represent a surviving share η of the initial pool, the total mechanical transfer has a dollar value equal to τ' . The effect on the exhaustee utility of a \$1 transfer is $\tau' - 0$, which corresponds to the average marginal utility between 0 and τ . Consequently, τ' is comprised between τ and 0 . The behavioral cost has a new first component compared to the behavioral cost of a τ increase. In that case, the first term $\int_0^{\tau} \frac{1}{t} dt$ represents benefits paid to workers who reach the exhaustion date because they slow down their job finding following the PBD extension.

To compare the welfare effects of both policy changes, we rescale Eqs. (34) and (35) so that they each represent the effect of a one dollar transfer to the unemployed. This creates the classic dynamic versions of the Baily-Chetty formula for benefit levels and durations:

$$\frac{\partial u_{\text{un}}}{\partial \tau}, = \underbrace{\frac{\tau' - \tau}{\tau}}_{\text{benefit level}} - \underbrace{\left(\eta \frac{\tau'}{\tau} - \eta \frac{\tau}{\tau} \right)}_{\text{duration}} \quad (36)$$

$$\frac{\partial u_{\text{un}}}{\partial \tau}, = \underbrace{\frac{\tau' - \tau}{\tau}}_{\text{benefit level}} - \underbrace{\left(\int_0^{\tau} \frac{1}{t} dt - \frac{\tau}{\tau} \right)}_{\text{duration}} \quad (37)$$

where η is the elasticity of expected duration of covered unemployment wrt benefit level and η is the elasticity of non-employment duration.

At the social planner optimum, the marginal effect of either policy parameter on workers' welfare is equal to zero. Consequently, to test if the current UI system is optimal, it is sufficient to compute estimates of the consumption-smoothing value and estimates of the behavioral costs, to take the difference between the two estimates and to compare it to zero. We now review estimates of each term.

3.3 Quantification of behavioral costs

We first review estimates of the behavioral costs and then estimates of the consumption smoothing value.

The behavioral costs of Unemployment Insurance programs depend on their disemployment effects: the elasticity of covered unemployment and non-employment duration wrt benefit level, and the marginal effect of potential benefit duration on the survival curve and on non-employment duration. There is a large literature estimating those UI effects on labor supply. We discuss some recent papers estimating UI effects on labor supply in Sections 2.3 and 2.4. The overall evidence is summarized in excellent reviews (Cohen and Ganong, 2024; Krueger and Meyer, 2002; Lopes, 2022; Meyer, 2002; Schmieder and von Wachter, 2016). The recent meta-analysis of Cohen and Ganong (2024) gathers almost 60 UI elasticity estimates from 52 studies

published before 2022. Their meta-analysis focuses on the effects of replacement rates and of PBD on either non-employment duration, or covered unemployment duration. Fig. 10 plots the estimates collected by Cohen and Ganong (2024) by publication dates. Elasticity estimates are almost all strictly positive implying disemployment effects of both Potential Benefit Duration (circles in orange) and benefit level (triangles in blue). Before interpreting their magnitude, we discuss recent trends in estimation methodology.

Since the mid-1990s, empirical research on UI effects is an important contributor to the credibility revolution. It has developed a series of specific designs to identify causal elasticity estimates. The main identification threat in empirical UI studies is a classical selection issue. Unemployed workers select into unemployment insurance categories based on unobservables. For example, in many countries, the Potential Benefit Duration depends on past work history. High-experience workers are eligible to longer PBD when they become unemployed. Then, comparing workers with long vs. short PBD does not allow to identify the causal effect of PBD as it is confounded by workers' unobserved productivity. In this example, observing past work experience helps to solve the selection issue. However, how much it helps depends critically on the PBD rule itself and on the data quality. When PBD is a deterministic function of past work experience with some discontinuous jumps, causal effects can be obtained through Regression Discontinuity Designs (Card et al., 2007). One can then compare the unemployment duration of workers in high vs low PBD categories in a neighborhood of the

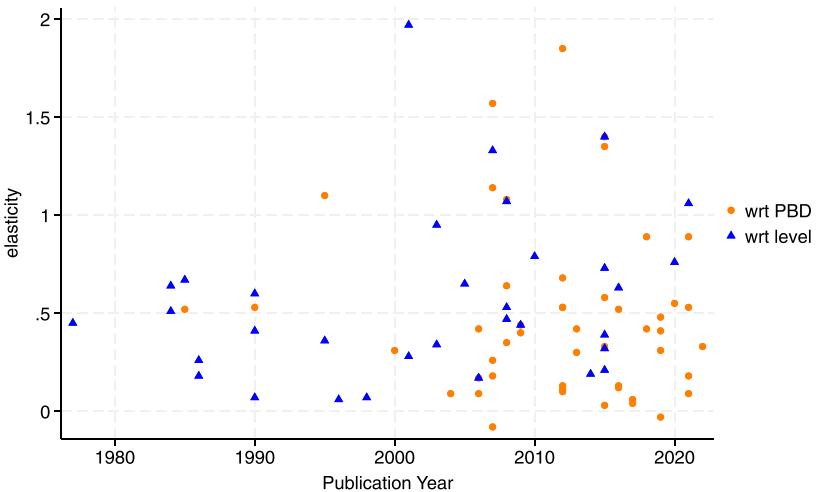


FIG. 10 Elasticity estimates of unemployment duration wrt potential benefit duration or benefit level. Notes: This figure presents elasticity estimates gathered in the review by Cohen and Ganong (2024) (Appendix Table B-1 and B-2 from which we exclude six outliers below -1 and above 2). The estimates are for the elasticity of unemployment duration wrt Potential Benefit Duration (PBD) in orange circles and wrt benefit level in blue triangles. The x-axis corresponds to the year of publication of the study.

PBD-rule discontinuity cutoff. Such a quasi-experimental design requires high-quality data on previous work experience to correctly identify workers in a neighborhood around the cutoff, and large initial samples so that the final selected sample of local comparison is large enough to detect effects with reasonable statistical power. Because of those data constraints, a significant share of available UI elasticity estimates since the 2000s are from Europe where large administrative datasets are accessible to researchers.

The UI empirical research also features some of the first applications of Regression Kink Designs (Card et al., 2015a, 2015b; Landais, 2015), which leverage discontinuous changes in the slope between the policy variable of interest and a running/selection variable. This design is used to identify the effects of benefit level specifically, as caps on benefit levels generate such kinks in the relationship between previous wages and benefits.

Another widely-used design in empirical UI studies is the difference-in-difference methodology. In the ideal case, the DiD design leverages an exogenous reform in policy parameters that affects only a subpopulation of workers and thus yields a natural control group. This method is less demanding in terms of data quality. It can be implemented either with administrative data or with survey data to the extent that they correctly identify workers impacted by the reform vs the untreated control group. One important identification assumption is that of exogenous reform, which in practice may be violated (Card and Levine, 2000). When labor market conditions worsen, policymakers may increase the generosity of unemployment insurance, as a countercyclical stabilization policy. Then, labor market conditions confound the effect of UI generosity. In some countries, like the US, the UI generosity rules are even countercyclical by design. When state-level unemployment rates reach certain pre-determined thresholds, Potential Benefit Duration is increased through the Extended Benefit and Emergency Unemployment Compensation (EUC) programs (Rothstein, 2011). Such countercyclical rules push US UI research to develop trigger design that control for a flexible (but parametric) function of unemployment when regressing unemployment duration on state-level PBD. Causal identification is then obtained from the discontinuous jump of PBD at unemployment triggers. In recent work, Chodorow-Reich et al. (2019) further leverage ex-post revision in state-level unemployment rate to focus identification on trigger events generated by measurement errors. An alternative strategy is to analyze non-automatic and politically-motivated changes in UI rules (see Card and Levine (2000), Johnston and Mas (2018) for such studies in the US). In Europe, few countries have automatic countercyclical rules (except France recently) and many studies implement DiD designs credibly (for example (Le Barbanchon et al., 2019); van Ours and Vodopivec, 2008).

In their review of 22 studies, Schmieder and von Wachter (2016) report that the average PBD elasticity is 0.41, and the average elasticity wrt replacement rates is 0.6 (Cohen and Ganong, 2024). who review twice as many papers find average of published elasticities in the same ballpark: 0.49 wrt PBD and 0.40 wrt replacement rate.

Beyond the average disemployment effects, those studies find interesting heterogeneity between the US and Europe, and over the business cycle. That heterogeneity has implications for targeting UI policies we discuss later.

The elasticity estimates also vary depending on the type of duration outcome considered. Some studies rely on administrative unemployment registers that record covered unemployment only. Such registers measure the duration between the first and last benefit payment in a given claiming spell. The duration outcome then corresponds to the variable τ defined in the previous Baily-Chetty formulas. When studies rely on matched unemployment-employment registers (such as social security data), they can record both covered unemployment (τ) and nonemployment duration (variable τ^* of the previous Baily-Chetty formula). In survey-based studies, the focus is rather on nonemployment duration, as unemployment receipts variables (when available) typically suffer from measurement error. [Schmieder and von Wachter \(2016\)](#) find that nonemployment duration elasticities are smaller than unemployment duration elasticities.

The Baily-Chetty formula (36) shows that the behavioral cost of UI benefit levels depends on both the elasticity of covered unemployment duration and of non-employment duration. However, as few studies report both elasticity types, it is useful to make an extra approximation to relate both elasticities and simplify the Baily-Chetty formula before taking it to the data. Under a constant hazard rate assumption $\lambda = \eta \tau$, we have $\lambda = -\xi$ where $\xi = \frac{\eta}{\lambda} = \frac{\eta}{\eta \tau} = \frac{1}{\tau}$. Consequently, the behavioral cost associated with a \$1 increase in unemployment benefits (and expressed in marginal utility when employed) writes:

$$\eta \lambda \xi = \frac{\eta}{\tau}, \text{ where } \lambda \text{ is the survival rate at benefit exhaustion.}$$

Behavioral Cost of UI Benefit Levels ([Schmieder and von Wachter, 2016](#)) report behavioral cost estimates from 18 studies from 5 countries (out of which 11 estimates are from the US). We explain how they measure the different components of the behavioral cost expression. First, they gather the elasticity estimates (η) from previous studies. Then, for each study, they also gather the other quantities of the behavioral costs: the survival rate at benefit exhaustion λ , the average hazard rate λ (to obtain the factor ξ), and the ratio of taxes over benefit level τ . [Schmieder and von Wachter \(2016\)](#) present cost estimates under two scenarios for tax rates. From the perspective of the UI agency, the relevant tax (τ) is the workers' contribution rate for unemployment insurance. On average, in OECD countries, the UI tax rate is of the order of 3 %. If we rather assume that the budget of the UI agency is integrated in the general government budget, then the relevant tax rate is higher and amounts to the total tax rate on labor income (around 30 %). Using the UI tax rate of 3 %, the behavioral cost for each additional \$1 transfer of UI benefits varies between \$0.06 and \$0.95, with a median of \$0.35. Taking the median estimate, for every dollar of mechanical transfer to UI claimants, \$1.35 has to be raised in taxes: \$1 of mechanical transfer and an additional \$0.35 because of the loss of tax revenues due to workers changing their behavior. Using the full labor tax wedge, the median behavioral cost is significantly higher at \$0.81.

Behavioral Cost of UI PBD As before, under the assumption of constant hazard rate, the expression of the behavioral cost of increasing Potential Benefit Duration simplifies. It writes: $-\xi \frac{\tau}{\gamma}$. Schmieder and von Wachter (2016) report the marginal effects of PBD on nonemployment duration — for eight European studies and five US studies. In Europe, the median estimate is 0.13: one month increase in PBD translates into a nonemployment duration by around 4 days. In the US, the mean estimated marginal effect is twice as large (0.28). Under the assumption of UI tax rate, the behavioral cost of \$1 transfer to UI exhaustees through PBD extension is between \$0.11 and \$2.13 with a median at \$0.60. As expected, under the assumption of general labor tax wedge, the behavioral cost rises to \$1.78. Overall, the behavioral cost of PBD increase is larger than the behavioral cost of benefit increase.

LESSON 1: The behavioral costs of providing UI are substantial.

3.4 Quantification of the social value of UI changes

To assess whether a benefit rise increases welfare, the behavioral costs estimated in the previous section are to be compared to estimates of the social value of more generous UI. Recall from the previous Baily-Chetty formula, that the social value of a \$1 transfer from the employed state to the unemployed state is $\frac{1}{\gamma} - \frac{1}{\gamma'}$. The social value depends on the marginal rate of substitution across covered state and contributing state. The empirical literature quantifying the marginal social value is less numerous than the empirical literature estimating disemployment effects, but it is growing fast. After the key seminal contributions of Gruber (1997) and of Chetty (2008), the availability of new types of consumption data (for example Ganong and Noel (2019), Kolsrud et al. (2018)) and the development of new identification methods (for example Landais and Spinnewijn (2021)) provides new insights on the social value of unemployment insurance. We first describe the various quantification methods of the social value of UI and then discuss their applications and corresponding estimates from recent papers.

3.4.1 How to quantify the social value of UI?

We review the four main approaches to quantify the social value of UI. We follow their publication order.

The Classical Consumption-Based Approach (Gruber, 1997) The first classical approach to estimate the marginal social value of UI is due to Gruber (1997). The approach rests on important assumptions about the utility function Gruber (1997). assumes that the utility functions of unemployed and of employed workers are the same ($U_U = U_E$). In addition, the approach assumes that the utility function has constant relative risk aversion (CRRA) and can be written as: $U = \frac{1}{1-\gamma} e^{-\gamma Y}$. Then the marginal utility writes:

' = $-\gamma$. Taking a first-order approximation of the marginal utility function, the marginal social value of UI becomes:

$$\frac{\psi' - \tau'}{\psi} \approx \gamma \frac{-\tau}{\psi} \quad (38)$$

The expression shows that, with data on consumption across states and a CRRA estimate, one can estimate the social value of UI. Gruber (1997) uses consumption data from the Panel Study of Income Dynamics (PSID) in the US. In the PSID survey, food consumption drops by 6.8 % when UI eligible workers become unemployed. There is a wide range of estimates for the CRRA parameter. Taking $\gamma = 2$ as a focal CRRA estimate, Gruber (1997) obtains that the marginal social value of \$1 transfer of unemployment benefits is \$0.13. In other words, workers would be willing to pay \$1.13 when they are employed to receive \$1 when they become unemployed. Such a marginal social value is lower than behavioral costs implied by the disemployment effect estimates available in the late 1990s (and still lower than the median of estimates available today). The comparison suggests that the US level of UI benefits is higher than the optimal level from the Baily-Chetty approach, unless the CRRA γ coefficient is significantly larger than 2.

The Liquidity to Moral Hazard Ratio Approach (Chetty, 2008) To address the limitations of Gruber (1997) approach, Chetty (2008) develops a sufficient-statistics approach that does not require consumption data, nor assumptions on the utility function parameters. The key idea is to leverage another type of policy variation in the same context, for example changes in severance payments. Chetty (2008) shows that the various behavioral search responses allow to identify the social value of UI. The underlying job search model is extended to allow unemployed workers to have savings at the beginning of their spell. Workers' optimal behavior in Eq. (24) is marginally modified as:

$$-\tau - \psi' = \psi'' \quad (39)$$

Let us consider a marginal increase in savings. Savings decrease the gap in consumption across the unemployed and employed states and workers decrease search effort:

$$\frac{\partial}{\partial} = \frac{\psi' - \tau - \psi'}{\psi''} = 0 \quad (40)$$

Recall that the marginal effect of a benefit increase on search effort writes: $\frac{\partial \tau}{\partial \psi} = -\frac{\psi'}{\psi''}$. Combined with Eq. (40), we obtain a new expression for the marginal social value of UI:

$$\frac{\psi' - \tau'}{\psi} = \frac{-\partial \tau / \partial \psi}{-\partial \psi / \partial \psi} = 1 \quad (41)$$

The above expression shows that the marginal social value of UI is identified as a simple combination of the job search effects of both benefit level and savings. Those two effects are sufficient statistics to assess the Baily-Chetty optimality of current UI levels.²⁸ The sufficient statistics do not require consumption data.

The approach allows to use quasi-experiments to identify both job search effects, ensuring the credibility of the quantification exercise. Chetty (2008) estimates the effect of severance payments (equivalent to savings in the simple one-period static model). Severance payments are one-time monetary transfers that workers receive from their employers at lay-offs. The minimum amounts of severance payments are mandated by law in many countries and increase with tenure. In some countries, the rule features discontinuities: only workers with a certain job tenure are eligible (Card et al., 2007). Such discontinuities can be leveraged to obtain exogenous variations in severance payments.

Another important insight from Chetty (2008)'s approach is an alternative interpretation of the behavioral search response. Namely, Chetty (2008) highlights that the search response can be decomposed into two channels: substitution and income. On the one hand, higher benefits reduce the net wage ($- \tau -$), and lower search effort through a substitution effect. On the other hand, higher benefits also increase liquidity and reduce search through an income effect. The income / liquidity effect corresponds formally to $\partial \partial$. The pure moral hazard cost is then the difference between the total effect on search and the income effect: $\partial \partial - \partial \partial$. Consequently, Eq. (41) shows that the value of insurance is the ratio between the liquidity effect and the pure moral hazard effect. Hence, Chetty (2008)'s identification strategy is referred as the *Liquidity to Moral Hazard Ratio* approach. For liquidity constrained workers, we expect large liquidity effect.

Landais (2015) and Huang and Yang (2021) adapt Chetty's approach, when there are no quasi-random variations in severance payment available in the context at hand. Landais (2015) shows that the time profile of benefits can identify the Liquidity-to-Moral-Hazard ratio. Huang and Yang (2021) leverage exogenous variations in reemployment bonus to directly estimate $\partial \partial$. This corresponds to the pure moral hazard effect (note that $\partial \partial = \partial \partial - \partial \partial$). Then, the liquidity effect $\partial \partial$ can be recovered indirectly in difference with the total effect of benefits.

The Marginal-Propensity-to-Consume Approach (Landais and Spinnewijn, 2021) Recently, Landais and Spinnewijn (2021) propose a third approach to quantify the social value of UI. It combines ingredients from the two previous approaches: consumption data and optimality results from workers behaviors in response to non-UI shocks Landais and Spinnewijn (2021)'s approach leverages estimates of the marginal propensity to consume

²⁸Indeed, the behavioral cost is already identified thanks to the job search effects of benefits.

(MPC) of both employed and unemployed out of extra income. To present the MPC approach, we consider again the static one-period Baily-Chetty model.²⁹ We extend the model so that workers can adjust their consumption from their own labor income / benefits (denoted τ). They are no longer hand-to-mouth strictly speaking. For example, workers may borrow / save to smooth consumption or their spouses may supply extra labor to generate additional income. Formally, workers undertake actions τ where τ denotes the workers' state (unemployed or employed). Action τ raises consumption at price ψ . Formally, workers solve the following problem:

$$-\psi \quad (42)$$

$$\begin{cases} = - \\ = - \\ = - \end{cases}$$

When τ represents spouses' working hours, ψ is the inverse of the spouses' wage rate. When workers adjust through the financial markets, the shadow price is related to interest rates. The first order conditions wrt the adjustment variables and τ write:

$$\frac{\partial}{\partial} = - \frac{\partial}{\partial} \quad (43)$$

$$\frac{\partial}{\partial} = - \frac{\partial}{\partial} \quad (44)$$

As already noted in [Section 3.2](#), the extra actions available to workers are second-order when assessing the social value of UI (because of the envelope theorem). In the extended model, the social value still amounts to the relative marginal utility of consumption across states, equal to the marginal rate of substitution ($\frac{\partial}{\partial} = \frac{\partial}{\partial} \frac{\partial}{\partial}$) minus one. Using the optimality of workers behaviors (above FOCs), the MRS writes:

$$= \frac{\frac{\partial}{\partial}}{\frac{\partial}{\partial}} = - \frac{\frac{\partial}{\partial}}{\frac{\partial}{\partial}} \quad (45)$$

The MPC approach quantifies the relative prices of smoothing consumption and bounds the relative marginal disutility costs $\frac{\partial}{\partial} \frac{\partial}{\partial}$. To identify unobserved prices τ , let us consider an income shock δ . After implicit

²⁹ [Landais and Spinnewijn \(2021\)](#) consider both a more complex one-period model and a dynamic model. Our simple model illustrates the method intuition which carries over to general models.

differentiation of the first order condition (43), the consumption response when unemployed writes:

$$\frac{\partial}{\partial} = \frac{\frac{\partial}{\partial} \frac{\partial}{\partial}}{\frac{\partial}{\partial} \frac{\partial}{\partial} \frac{\partial}{\partial} \frac{\partial}{\partial}} \quad (46)$$

We obtain the consumption response when employed along the same lines. After taking the odds ratio of the MPCs and forming their ratio across states, we obtain:

$$\frac{\frac{\partial}{\partial}}{\frac{\partial}{\partial}} = \frac{\sigma}{\sigma} \quad (47)$$

where we denote $\sigma \equiv -\frac{\partial}{\partial} \frac{\partial}{\partial}$ and $\sigma \equiv -\frac{\partial}{\partial} \frac{\partial}{\partial}$ parameters that capture the curvature of the utility function wrt consumption and wrt action . Combining Eqs. (45) and (47), the MRS writes:

$$= \frac{\frac{\partial}{\partial}}{\frac{\partial}{\partial}} \underbrace{\frac{\sigma}{\sigma}}_{\geq} \underbrace{\frac{\frac{\partial}{\partial}}{\frac{\partial}{\partial}}}_{\geq} \quad (48)$$

Eq. (48) shows that MPCs indeed identify the MRS, but still require assumption on utility functions. Those assumptions are weaker than in the consumption-based approach, as what matters is *relative* marginal utilities or *relative* utility curvature *across states*. If preferences over consumption and resources are exponential functions stable across states, then the ratio of utility curvature is equal to one whatever the exact value of risk preferences. For utility functions with decreasing absolute risk-aversion (DARA), the curvature ratio can be bounded above one. This is because we can reasonably assume that unemployed consume less than employed workers () and devote more resources to support their income (). It is also reasonable that when unemployed devote more resources, their marginal cost is higher than for employed workers, so that the third factor in Eq. (48) is also greater than one. The weaker assumptions of the MPC approach come at the cost of point identification. In the end, the MPC approach provides a lower bound on the MRS and on the social value of UI. How informative the lower bound is depends on how its estimate compares to previous estimates obtained from the literature. We discuss estimates in the next section.

In practice, MPC estimates can be identified in quasi-experiments, ensuring the identification credibility of the whole method (as in the Sufficient-Statistics approach). An important requirement though is that the quasi-experiments are homogeneous between employed and unemployed workers. For example, Landais and Spinnewijn (2021) exploit large variations in welfare benefits provided by municipalities across types of households.

The Revealed-Preference Approach The most direct method to identify the social value of UI is to study workers' choices to buy insurance. Building on the MPC-approach model, suppose that workers can get extra UI coverage at rate $\frac{\partial}{\partial}$. They would buy insurance if and only if $\frac{\partial}{\partial} - \frac{\partial}{\partial} \geq 0$. Rearranging the terms, we obtain that workers will buy extra coverage if their MRS is above the expected price:

$$\frac{\partial}{\partial} \equiv \frac{\partial}{\partial} - \frac{\partial}{\partial} \quad (49)$$

Note that the relevant price $\frac{\partial}{\partial}$ depends on the individual-specific job finding rates $\frac{\partial}{\partial}$ (or individual unemployment risks). The Revealed-Preference (RP) approach has two important requirements: observing UI coverage choices and precise data on perceived unemployment risks. As UI is a mandatory insurance in many countries, the RP method has not been used in UI studies with the notable exception of [Landais and Spinnewijn \(2021\)](#). Swedish workers have income-related UI benefits (instead of a flat benefit level) if they pay a uniform premium. Iceland, Denmark and Finland are three other countries with voluntary UI schemes.

Other Approaches For the sake of completeness, we briefly discuss two other methods to assess the social value of UI. First, [Shimer and Werning \(2007\)](#) observe that there is a direct relation between the value of unemployment and reservation wages. Recall from [Section 2.2](#) that, in the standard job search model, we have $\phi_0 = \frac{\partial}{\partial} - \delta \frac{\partial}{\partial}$ where ϕ_0 is the reservation wage and $\frac{\partial}{\partial}$ the expected value of unemployment, both at the beginning of the unemployment spell. δ is the discount factor. Consequently, with data on reservation wages, we can identify the effect of a marginal increase in benefit on workers welfare (see [Le Barbanchon et al. \(2019\)](#)).

Second, [Hendren \(2017\)](#) shows how consumption responses *before* job loss to changes in perceived unemployment risk identify the social value of UI. [Hendren, \(2017\)](#)'s method builds on the Consumption-Based approach as it requires to observe consumption path before unemployment. In addition, identification requires measuring workers' beliefs about future job loss (available in the Health and Retirement Study for example). Identification rests on the following mechanism. When forward-looking workers learn about future job loss, they decrease consumption all the more that they are willing to increase precautionary savings and transfer income towards the unemployed state. This assumes workers' optimization (as in the MPC approach). Formally, the social value of UI writes:

$$\frac{\partial'}{\partial} - \frac{\partial'}{\partial} \approx -\gamma \left[\frac{\partial}{\partial} \right] \quad (50)$$

where $\left[\quad \right]$ is the average relationship between consumption when employed (before any job loss) and beliefs about future employment . The main advantage of [Hendren \(2017\)](#) approach compared to the classical consumption-based approach is that it allows for state-dependent utility function ($' \neq ' \neq ' \neq '$).

3.4.2 Selected review of social value estimates

In [Table 4](#), we report estimates of the social value of UI benefit increase.³⁰ We select a subset of studies representing the four main approaches: Consumption-Based, Liquidity-to-Moral-Hazard ratio, MPC and Revealed Preference approach. Studies using the same approach are grouped into panels. Our objective is not to be exhaustive, but to represent estimates obtained with different approaches and data sources. [Table 4](#) comprises sixteen social value estimates from ten studies.

The underlying data are drawn from the United States, Sweden, Austria and Brazil. This is already a significant coverage of countries, but there is room for further research to widen the scope of estimates across even more countries. Studies cover estimates from the 1970s to recent years (up to 2015). A noticeable feature of [Table 4](#) is the richness of the data sources mobilized by the empirical UI literature. As observing the consumption path of unemployed workers is key to estimate the social value of UI (in the Consumption-based and MPC approaches), the empirical literature has made recent breakthroughs in data sources. After the seminal use of survey data on consumption by [Gruber \(1997\)](#), the literature flourishes in administrative data: Swedish tax registers in [Kolsrud et al. \(2018\)](#), high-frequency banking transaction data in the US study of [Ganong and Noel \(2019\)](#), and Brazilian VAT receipts matched with employment registers in [Gerard and Naritomi \(2021\)](#). As discussed previously, such granular data allow to observe the precise timing of consumption in relation to job loss and to UI benefit receipts and exhaustion.

Compared to the initial estimate of 6.8 % from [Gruber \(1997\)](#), the recent literature find larger consumption drop at job loss reported in Panel A. [Rothstein and Valletta \(2017\)](#) document a 10 % drop in consumption in more recent US survey data (SIPP) down to 20 % drop during recession times. Using Swedish tax registers, [Landais and Spinnewijn \(2021\)](#) obtain that consumption drops by 12.6 % at job loss. [Ganong et al. \(2022\)](#) observe a 6.1 % consumption drop at job loss in bank data in the US, which is very close to [Gruber \(1997\)](#)

³⁰ We select a subset of studies from [Table 3](#) of ([Schmieder and von Wachter, 2016](#)) and update the table with more recent studies. Namely, we do not report every study using the PSID dataset.

TABLE 4 Estimates of the social value of UI benefit increase.

Study	Range of years	Country	Data source	Key moment	Social value
Panel A: Consumption-Based Approach					
Gruber (1997)	1968–1987	United States	PSID, food only	At job loss: 6.8 %	0.136
Rothstein and Valletta (2017)	2001 panel	United States	SIPP	At job loss: 10.0 %	0.2
Rothstein and Valletta (2017)	2008 panel	United States	SIPP	At job loss: 20.0 %	0.4
Ganong and Noel (2019)	2012–2015	United States	JPMC1 checking account	At job loss: 6.1 %	0.122
Landais and Spinnewijn (2021)	2000–2007	Sweden	Tax records	At job loss: 12.9 %	0.258
Ganong and Noel (2019)	2012–2015	United States	JPMC1 checking account	UI exhaustees: 25 %	0.5
Gerard and Naritomi (2021)	2010–2015	Brazil	VAT receipts, RAIS registry	UI exhaustees: 17 %	0.34
Hendren (2017)	1992–2013	United States	HRS-PSID	29 % after future job loss news	0.58

Panel B: Liquidity to Moral Hazard Approach

	Card et al. (2007)	1981–2001	Austria	Social Security Registry	Job Finding Response to
Chetty (2008)	1985–2000	United States	SIPP	Severance pay, RD	1.4
Landaas (2015)	1970s–1984	United States	CWBH	Severance pay, OLS	1.5
Huang and Yang (2021)	2001–2011	Taiwan	Admin. registers	Time profile of benefits, RKD	0.88
				Reemployment bonus, RKD	0.5–1.5

Panel C: Marginal Propensity to Consume Approach

	Landaas and Spinnewijn (2021)	2000–2007	Sweden	Tax records	Consumption response to welfare benefits	>0.59
				Tax records, survey on Unemp	Choice of UI scheme	1.13, 2.13

Panel D: Revealed Preference Approach

	Landaas and Spinnewijn (2021)	2000–2007	Sweden	Tax records, survey on Unemp	Choice of UI scheme	1.13, 2.13

Notes: This table presents estimates of the social value of UI benefit increase across a selected set of studies. The first panel A reports estimates following the Consumption-Based approach introduced by Gruber (1997), and extended to study the social value of PBD extension. To compute the social value from the consumption loss, we set the CRRA parameter to a conservative value of $\gamma = -1$. When studying the social value of PBD extension, we use consumption drop for UI exhaustees. The second panel B reports estimates following the Liquidity-to-Moral-Hazard approach introduced by Chetty (2008). To disentangle moral hazard from liquidity, they use job finding responses to UI and to another policy listed in the column entitled “key moments” Card et al. (2007), Huang and Yang (2021). Landaas (2015) use response to PBD extension and Chetty (2008) to benefit level increase. The third and fourth panels C and D report estimates from the MPC and RP approaches respectively (for any unemployed worker, not only UI exhaustees). The MPC approach identifies a lower bound for the social value of UI. Landaas and Spinnewijn (2021) do not highlight one specific RP-based estimate, we report here the average social value under both extreme beliefs models of unemployment risk. Other studies using the same PSID data are Chetty and Looney (2006), Cochrane (19910), Kroft and Notowidigdo (2016), Stephens (2001) and Chetty and Szeidl (2006). Browning and Crossley (2001) also provides Consumption-based evidence for Canada.

though. We translate the consumption drops at job loss into the social value of UI benefit increase in the last column of [Table 4](#). We assume that the CRRA parameter of the utility function is equal to two. We obtain social value estimates ranging from \$0.12 to \$0.4 (corresponding to the first five estimates in the top panel). They imply that workers would be willing to decrease their consumption when employed by \$1.12–\$1.4 for an extra dollar of consumption when unemployed. The social value estimates are very sensitive to the CRRA parameter. Assuming that the CRRA is equal to five, the social value estimates would range from \$0.30 to \$1. Such changes are pivotal for optimal UI design. In the Baily-Chetty framework, social value is compared to behavioral cost. When we use consumption drop *at job loss* to compute the social value of UI, the relevant behavioral cost is the one induced by a transfer of benefits through a monthly benefit increase (as opposed to an extension of PBD). In the Baily-Chetty Equation (), the relevant consumption when unemployed is before UI exhaustion. From the previous section, the median behavioral cost of monthly benefit increase is \$0.35. Consequently, most social value estimates are lower than the median behavioral cost when we assume $\gamma = 2$, but higher when we assume $\gamma = 5$.

In the last rows of Panel A, [Table 4](#) reports estimates of the drop of consumption for UI exhaustees (compared to their pre job loss levels). These estimates are difficult to compute in survey data as they require very precise measure of benefit receipts and high frequency observations. The new administrative data sources overcome the challenge. We expect the consumption of UI exhaustees to be significantly lower than the consumption of short-term unemployed, as they no longer rely on UI benefits and they may have lower savings after spending some time unemployed. Indeed, [Ganong and Noel \(2019\)](#) find that UI exhaustees have consumption expenditures 25 % lower than before job loss (compared to 6.1 % just after they become unemployed). [Gerard and Naritomi \(2021\)](#) report a consumption drop for UI exhaustees of 17 %, which translates into a social value estimate of \$0.5. The social value estimates are lower than the median estimate of the behavioral costs of benefit increase through PBD extension (the relevant policy instrument for UI exhaustees). In the previous section, the corresponding median behavioral cost is \$0.6 and even higher when computed with the full labor tax wedge. As above, the conclusions drawn from the Baily-Chetty exercise depend heavily on the CRRA assumption. Another consideration is that with such large changes in consumption at exhaustion, the first order approximation underlying [Eq. \(38\)](#) may not be valid any more. Taking the estimate from [Ganong and Noel \(2019\)](#), the social value estimate increases from \$0.5 to \$0.77 when we do not linearize the utility function (with $\gamma = 5$).

In the next three panels of [Table 4](#), we report social value estimates from approaches that do not require explicit assumption on the CRRA parameter. Overall, the corresponding studies find higher social value estimates (above \$0.5). In Panel B, we report findings from four studies applying the liquidity to moral

hazard ratio approach. [Card et al. \(2007\)](#) and [Chetty \(2008\)](#) both find similar estimates around \$1.5, despite their differences in countries and in empirical designs to estimate the job search response to severance pay [Landais \(2015\)](#) finds a lower estimate at \$0.89 when he uses variations along the time profile of unemployment benefits to disentangle liquidity and pure moral hazard effects. [Huang and Yang \(2021\)](#) find estimates in the same ballpark \$0.5–1.5, as they use the job finding response to a reemployment bonus to identify the marginal utility when employed (moral hazard denominator).

In the last two panels, we report estimates of the MPC and Revealed-Preference approaches on the same sample of workers from [Landais and Spinnewijn \(2021\)](#). They find that the MPC approach delivers a lower bound for the social value, as high as \$0.59. This is higher than the consumption-based estimate that they can compute on the same sample of workers. The Revealed-Preference estimates are even higher and the average estimate ranges from \$1.13 to \$2.13 depending on how unemployment risks are estimated. When workers are assumed to have correct beliefs about unemployment risks (based on the information available in the register dataset), the mean social value is \$2.13. However survey data eliciting subjective beliefs point to important risk misperception. This leads the RP method to overestimate the MRS. After correction, the social value is lower, but still significantly higher than consumption-based estimates. Beyond the average, [Landais and Spinnewijn \(2021\)](#) document significant heterogeneity in social value across groups of workers with a first quartile at \$0.8 and a third quartile at \$1.73 (estimated under misperceived risks).

While the approaches in Panel B to D do not rely on explicit CRRA parameters, their estimates imply some high values of the CRRA parameters. [Chetty \(2008\)](#) states that his social value estimate could be rationalized in a model with CRRA equal to 5. To match the lower bound of their MPC-based estimate, [Landais and Spinnewijn \(2021\)](#) would have to multiply the observed consumption drop by a CRRA parameter equal to 4. Further research is needed to deliver more estimates through the Liquidity to Moral Hazard ratio approach, the MPC approach and the RP approach. This would help to draw stronger conclusions on whether the consumption-smoothing value of UI is above or below the median behavioral costs (around \$0.5).

Further research is also needed on the reasons explaining the differences of estimates across methods. Whether the type of consumption observable in data could drive the difference between Consumption-Based and Liquidity-to-Moral-Hazard ratio approaches is debated. Another important candidate explanation is whether marginal utility depends on state. State-dependent utility leads to severe bias in the Consumption-Based approach. Last, failing to account for behavioral frictions also generate different bias across methods.

LESSON 2: Estimates of the social value of UI differ widely across identification methods. The most recent methods which are robust to risk-aversion assumptions yield significantly higher estimates.

In the previous section, we systematically compare the social value estimates with the behavioral cost estimates. We then follow the traditional Baily-Chetty approach to test whether one single policy instrument, either PBD, or benefit level, is set at its optimum. The Baily-Chetty framework derives the marginal welfare effect of a transfer of \$1 to unemployed workers. If the welfare effect of a marginal change in benefits is zero (social value of UI equal behavioral cost), then the social planner already maximizes welfare and no further policy changes are needed. If the welfare effect is positive (social value of UI greater than behavioral cost), then increasing the generosity of UI is the recommended policy. When the welfare effect is negative, the opposite policy recommendation holds. While the traditional Baily-Chetty approach is an established and useful policy assessment tool, it is often implemented for one policy instrument only and independent of changes in other policy instruments. This is an important limitation, as in practice social planners may leverage various policy instruments and policy-makers need empirical guidance on which policy is to be expanded vs another one. The next section presents the unified framework of [Hendren and Sprung-Keyser \(2020\)](#) which addresses this limitation, and allows for policy comparisons, taking into account redistribution effects.

3.5 The marginal value of public funds

This section discusses how to compare various UI policy changes against one another and against other government policies, within the unified framework of [Hendren and Sprung-Keyser \(2020\)](#). We briefly review how their unified framework yields the Marginal Value of Public Funds (MVPFs) as a key criteria for policy assessment. We then compute UI policy MVPFs using the various previous estimates in [Sections 3.3 and 3.4](#). We draw tentative comparisons between policies increasing benefit level vs. PBD, and we discuss how redistribution effects are accounted for in the MVPFs analysis (while absent from the traditional Baily-Chetty approach).

The MVPFs framework The government considers the following social welfare defined as the weighted sum of individual utilities : $\psi = \sum \psi_i$ where ψ_i is the social welfare weight of individual i . The government evaluates the welfare effects of a policy π with upfront initial spending π_0 . The social welfare change writes:

$$\Delta \psi = \eta \sum \psi_i \Delta u_i \quad (51)$$

where Δu_i is individual i 's willingness-to-pay for policy π out of her own income. Formally, $\Delta u_i = \frac{\partial u_i}{\partial \pi}$ with λ_i the marginal utility of individual i . The first factor η is the average social marginal utility of the beneficiaries of the policy.³¹

³¹ The average social marginal utility writes: $\eta = \frac{1}{\sum} \eta_i \Delta u_i = \sum \psi_i \lambda_i \Delta u_i / \sum \psi_i$

On the cost side, we denote the (present discounted) value of government budget and η — the net impact of policy η on government budget. The net impact includes both the initial cost of the program and all other effects of behavioral responses on the government budget. The Marginal Value of Public Funds of policy η is defined as:

$$= \frac{\Sigma}{\text{---}} = \text{---} \quad (52)$$

The effect of policy η on social welfare per dollar of government expenditure writes: η . Introducing the MVPFs conveniently separates the policy effect into two factors. The first factor η captures redistribution effects through social welfare weights. The second factor η gives unit weights to all beneficiaries but captures all relevant fiscal externalities.

The MVPFs framework allows to construct hypothetical budget-neutral policy changes. Suppose that the government increases spending on policy A by an amount η and reduces spending on policy B by the same amount (to keep the same budget). The policy shift from B to A increases social welfare if and only if:

$$\eta \quad \eta \quad (53)$$

This clarifies that key inputs for policy choices are estimates of the MVPFs for a large set of policies. We compute the MVPFs estimates for various UI policies. This is useful to compare different types of UI policies one against the other, but also UI policies to policies in other domains.

Before moving to MVPFs estimates, we highlight the conceptual similarities and differences between the MVPFs and Baily-Chetty approach. First, the unified approach of ([Hendren and Sprung-Keyser, 2020](#)) allows for heterogeneity in social marginal utility across policy beneficiaries, while the standard Baily-Chetty approach does not. Second, the MVPFs approach ultimately compares various policies, while the traditional Baily-Chetty approach focuses on the optimality of one policy only. Third, both approaches turn out to be very similar on accounting for government costs. The effects of behavioral responses on the government budget included in the MVPFs denominator are the same as the behavioral costs due to general fiscal externalities in the standard Baily-Chetty formula. The only difference is that the MVPFs computation does not require the government to close the budget constraint.³²

The MVPFs of UI policies We compute the MVPFs of the two UI policies we considered so far. They increase UI generosity either through an increase in benefit levels (policy η) or through an increase in PBD (policy η). [Hendren and Sprung-Keyser \(2020\)](#) compute the MVPFs for those policies using

³² Closing the budget constraint may induce extra behavioral costs on government budget as the government raises revenue to fund UI (eg through income effects). They do not appear in the standard Baily-Chetty approach though.

estimates from the US. We follow their UI MVPFs interpretation updating the WTP estimates and expanding the analysis beyond the US. The WTP for \$1 dollar of UI benefits is equal to the MRS between the unemployed and employed states. For policy π , the extra dollar is transferred to unemployed before their benefit exhaustion (with unemployment duration t before PBD τ). For policy π' , it is transferred to UI exhaustees (such that $t = \tau$). In line with the previous Baily-Chetty formula, we then have:

$$= \frac{\pi' - \pi}{\pi} = \frac{\pi' - \pi}{\pi} \quad (54)$$

$$= \frac{\pi' - \pi}{\pi} = \frac{\pi' - \pi}{\pi} \quad (55)$$

Consequently, we obtain π' and π estimates using the social value estimates in the previous section. For the main analysis, we adopt the Consumption-Based estimates with a CRRA parameter equal to four (instead of two). This makes corresponding Consumption-Based estimates closer to estimates obtained with the three other approaches (Liquidity-to-Moral-Hazard ratio, MPC and RP). For US WTP estimates, we use as consumption drop estimates: 9 % at job loss ([Hendren and Sprung-Keyser, 2020](#)) and 25 % for UI exhaustees ([Ganong et al., 2022](#)).³³ For Europe, the only available consumption drop estimates in Table 4 are from Sweden: 4.4 % at job loss and 9.1 % for long-term unemployed ([Kolsrud et al., 2018](#)). This is not ideal, as Sweden has a relatively high replacement rate and no UI exhaustion over the period of estimation. Consequently, one should take our European MVPFs with a grain of salt.

The denominators of the MVPFs are simply the \$1 spent on UI benefit increase augmented with the effect on government budget due to behavioral reactions. Assuming that behavioral reactions are related to search effort only, the second term corresponds to the behavioral costs of the Baily-Chetty formulas (36) and (37). The behavioral reactions generate negative fiscal externality increasing the net cost of the UI policy. Formally, we have:

$$= \quad (56)$$

$$= \quad (57)$$

We use the behavioral cost estimates analyzed in [Schmieder and von Wachter \(2016\)](#). As the MVPFs framework adopts the point of view of the general government (not of the UI agency), we choose the full labor wedge as relevant tax for fiscal externality. Appendix Tables B1 and B2 report 34 MVPFs estimates for European and US policies resp.

³³ For the consumption drop at job loss in the US, the 9 % estimate corresponds to the 6 % estimates from [Ganong et al. \(2022\)](#) and [Gruber \(1997\)](#) corrected for pre-job loss drop in consumption documented in [Hendren \(2017\)](#). It is close to the 10 % estimate of [Rothstein and Valletta \(2017\)](#) from the 2001 PSID panel.

[Fig. 11](#) plots the distributions of the MVPFs by continent in Panel 44 and 45. European MVPFs vary between .24 and .99. The median of the 21 EU estimates is .51. All estimates are below the reference value of one, which corresponds to simple nondistortionary transfers from the government to an individual. In Panel 11b, the 13 US estimates vary between .51 and 1.18 with a median at .78. Four US estimates are above the reference value of one. Of course, such a conclusion depends highly on the CRRA choice and on the tax definition.

In Panel 11c, we take the perspective of either the US federal government or a European government whose budget is consolidated across countries. We ask whether such a government should increase benefit levels while reducing PBD, or the opposite. We compute the median MVPF_{PBD} and $\text{MVPF}_{\text{Benefit level}}$ within Europe and the corresponding median estimates in the US separately. In Panel

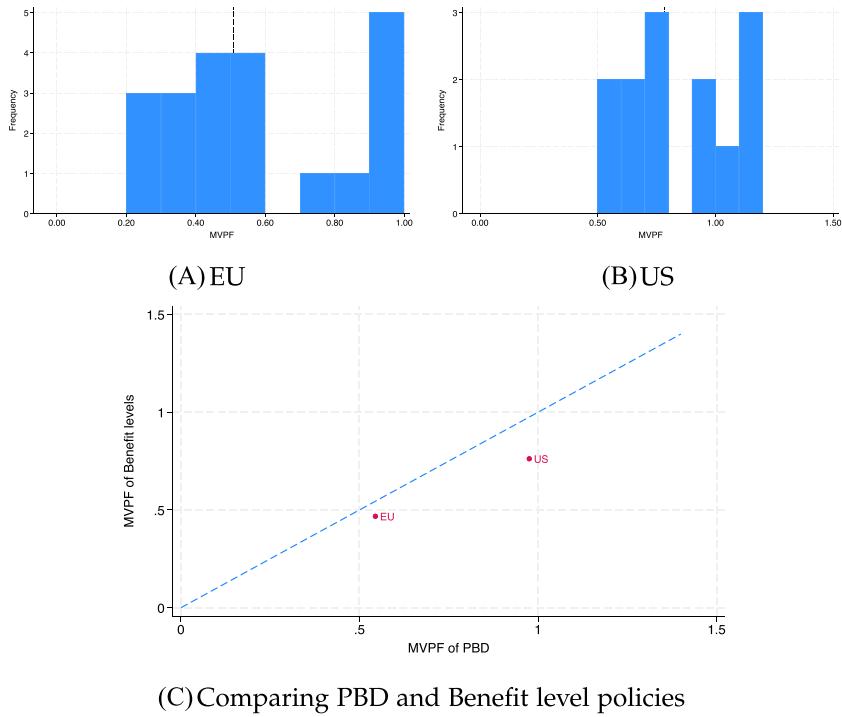


FIG. 11 The marginal value of public funds (A): EU (B): US (C): Comparing PBD and Benefit level policies **Notes:** The figure reports estimates of the Marginal Value of Public Funds for UI policies. To compute the MVPFs, we use WTP estimates from [Table 4](#) (Consumption-based approach) and the behavioral costs estimates from [Schmieder and von Wachter \(2016\)](#) (with some updates). See main text for details. Panel 44 shows the MVPFs distribution in Europe. The dashed vertical line corresponds to the median value. Panel 45 shows the MVPFs distribution in the US. For each continent, we compute the median MVPFs for each UI policy (PBD vs benefit level) and we plot MVPF_{PBD} against $\text{MVPF}_{\text{Benefit level}}$ in Panel 46. The dashed blue line corresponds to the 45 degree line.

46, we plot the median estimates vs the median estimates, together with the 45 degree line. We find that the EU is located almost on the 45 degree line. Consequently if the beneficiaries of the two policies have equal social marginal utility, the government will be indifferent between spending on either policies. There is no welfare gain for changing the policy-mix. On the contrary, the US is located below the 45 degree line. Under the same assumption of equal social marginal utility across each policy beneficiaries, the US government can increase welfare by increasing spending on PBD and decreasing spending on benefit level by the same amount ($\eta_{PBD} - \eta_{Benefit}$). This policy recommendation holds as long as the relative social marginal utility of PBD- vs benefits-increase beneficiaries $\eta_{PBD} / \eta_{Benefit} > 1$. Of course, such US policy recommendation depends heavily on PBD elasticity estimates, which are only three in our sample. In addition, we do not account for uncertainty in those estimates. Further research is needed to alleviate those two weaknesses. The above analysis rather shows the flexibility of the MVPFs framework.

One important question remains though. What would be a reasonable quantification for the relative social marginal utility between the PBD-extension and the benefit-increase beneficiaries ($\eta_{PBD} / \eta_{Benefit}$)? In principle, the average social marginal utility depends first on government social welfare weights of beneficiaries, and second on their marginal utility when employed.³⁴ Let us consider that the government puts the same social welfare weights on all job losers. Then, differences in social marginal utility are driven by differences in average individual marginal utility when employed. UI exhaustees who benefit from PBD extension are generally negatively selected on potential wages compared to the average pool of unemployed (see negative duration dependence of wages in [Section 2](#)). Their marginal utility when employed is thus greater than that of average UI claimants. It is then reasonable to consider $\eta_{PBD} / \eta_{Benefit} > 1$. Consequently, a policy-mix with some redistribution objective would tilt towards PBD-extension rather than benefit-increase even more than the relative $\eta_{PBD} / \eta_{Benefit}$ suggests.

The previous MVPFs estimates are useful to compare UI policies to any other policy (e.g., training policies or other educational policies). For example, [Hendren and Sprung-Keyser \(2020\)](#) shows how UI policies have lower MVPFs than educational policies in the US. Further research is needed to confirm this finding in Europe. Comparing UI policies and educational policies also require to compute their respective average social marginal utility. For UI policies, this amounts to the average social marginal utility of employed workers at risk of becoming unemployed, which further research would need to quantify.

³⁴ For the sake of simplicity, we assume that there is no heterogeneity in the individual WTP within each beneficiary group. Otherwise, a third component depending on the product between individual social marginal utility and WTP matters.

3.6 UI effects beyond unemployment duration

The design of UI policy requires a priori to also consider effects beyond job finding rates, such as effects on reemployment wages, and on job separation. In this section, we provide some informal insights on the channels through which such other effects impact UI design, and some empirical evidence on their magnitude.

3.6.1 UI effects on wages

To analyze wage effects, the UI literature modifies the underlying job search model underlying the Baily-Chetty framework in [Section 3.2](#). Whether the model assumes random search or directed search, the introduction of wages does not change the social value part of the Baily-Chetty formula ([Chetty, 2006, 2008; Nekoei and Weber, 2017](#)).³⁵ As explained in [Section 3.2](#), as long as wages are a choice variable for workers, direct UI effects on wages do not contribute to the social value (because of the envelope theorem). Outside of the Baily-Chetty framework, direct wage effects may matter. For example, when the social planner maximizes welfare beyond the private value of UI for unemployed, UI externalities through wages may enter in the social value of UI (see [Section 3.8](#) for example).

On the contrary, UI effects on wages may contribute to the social cost of the Baily-Chetty formula. When taxes are proportional, wage increases due to UI loosen the government budget constraint. This triggers a positive fiscal externality that counteracts the standard negative fiscal externality related to disemployment effects. Assuming that search is directed as in [Nekoei and Weber \(2017\)](#), we derive the UI welfare effect in the one-period static model (see online appendix for computation details). The marginal welfare effect of UI benefit increase writes:

$$\frac{\partial \text{social value}}{\partial \text{UI benefit}} = \frac{\eta' - \eta' - t}{\eta' - t} - \eta_{-} - \eta_{+} \quad (58)$$

where all notations are previously defined, except the proportional tax rate t and the elasticity of wages wrt benefit generosity η . Compared to the fixed-wage Baily-Chetty formula [\(26\)](#), the only difference is the wage elasticity term (η'). In [Section 3.3](#), we report an average estimate of 0.6 for the elasticity of unemployment duration (η_{-}). This is a lower bound for the fixed-wage behavioral cost of the one-period model.³⁶ How does this compare to available estimates of the wage elasticity to benefits?

³⁵ For a detail proof in the random search model, see model extension 2 in [Chetty \(2008\)](#). [Nekoei and Weber \(2017\)](#) discusses the wage channel in the directed search model. We adopt their model in the online appendix.

³⁶ The fixed-wage behavioral cost ($-\eta_{-}$) depends on the fraction of periods spent employed π . As $\pi \in [0, 1]$, the behavioral cost is larger than the elasticity estimate.

The quasi-experimental literature on wage effects of UI finds (if anything) modest effects on reemployment wages. We analyze twelve elasticity estimates from eight studies (Card et al., 2007; Centeno and Novo, 2009; Johnston and Mas, 2018; Lalivé, 2007; Le Barbanchon, 2016; Nekoei and Weber, 2017; Schmieder et al., 2016; van Ours and Vodopivec, 2008). The average elasticity amounts to $-.028$ and estimates range from $-.16$ to $.017$.³⁷ All studies, except (Nekoei and Weber, 2017), report that more generous UI *decreases* reemployment wages, even though most estimates are not statistically significant at standard levels. Among the studies implying a negative elasticity, Schmieder et al. (2016) has the most precise estimate, which turns out to be statistically significant. Their estimate of the wage elasticity (wrt PBD) is $-.014$. Doubling the UI generosity decreases wages by 1.4% . Nekoei and Weber (2017) is the only study finding a statistically significant positive effect. Their estimate of the wage elasticity (wrt PBD) is $.017$. The two wage elasticity estimates imply a fiscal externality following a $\$1$ UI transfer ranging from $-\$.014$ to $“.017$. Compared to the median estimate of fiscal externality due to disemployment effects ($“.34$), this is one order of magnitude lower. However, the average disemployment estimate may not be the relevant comparison point. For example, Nekoei and Weber (2017) find that the elasticity of nonemployment is also low in their context, so that UI effects on wages can partly compensate for the negative fiscal externality due to disemployment. Their quantification relies on long job duration (high α in the Baily-Chetty formula), and on reemployment wage effects persisting all over the job spell. Two aspects that deserve further research.

Jäger et al. (2020) study the effects of UI replacement rates on *average economy-wide wages*. They implement difference-in-difference designs around four major UI reforms changing benefit levels in Austria. They find that $\$1$ increase in UI benefits leads to $“.01$ dollars increase in wages. The wage effect rises to around $“.11$ for job movers, although the estimate is not statistically significant (see Table IV in Jäger et al. (2020)). Again, this suggests that wage effects are a second-order term of UI behavioral costs.³⁸ Further research on wage effects due to benefit level increases would be helpful to confirm this lesson.

LESSON 3: In the majority of recent studies, more generous UI policy decreases wages imposing further (second-order) behavioral cost to provide UI.

³⁷ The $-.16$ estimate from Johnston and Mas (2018) is an outlier, as the second lowest elasticity equals $-.06$.

³⁸ Lindner and Reizer (2020) is a notable exception where both fiscal externalities of disemployment and of wage effects are of similar order of magnitude. They analyze a benefit front-loading reform in Hungary. It changes the time path of benefits, but keeps constant the overall amount of benefits paid over the potential benefit duration. Accounting for the time path of job finding rate, “the new benefit mechanically increased government spending by around US\\$119 (SE 0.8) per unemployed worker” p.142. Shorter unemployment spells generate positive fiscal externality of $\$77$ (). Higher reemployment wages increase government budget by $\$194$.

3.6.2 UI effects on job separation

Beyond the effects on unemployed job seekers, UI generosity may affect the behavior of employed workers. In theory, when employed workers become eligible to more generous UI in case of job loss, the value of their outside option increases, and job surplus decreases, which can trigger higher separations. In the Baily-Chetty framework, such effects do not contribute to the social value of UI to the extent that separations are the outcome of an optimizing behavior of employed workers.³⁹ Of course, in a general equilibrium framework, there could be extra cost for firms of excess turnover induced by UI rules. Actually, the early literature analyzing UI effects on job separations takes the firms' perspective and asks whether experience-rating in UI contribution rates makes firms internalize the social cost of job loss in their firing decisions (Anderson and Meyer, 1993; Feldstein, 1976; Topel, 1983, 1984).

In this section, we report empirical evidence on the effects of more generous UI on workers separation rates. In the Baily-Chetty framework, separation effects induce fiscal externalities to be accounted for. We do not formalize here the general expression of the fiscal externality and leave it for future research. In a nutshell, when workers react to more generous UI by separating from their firms and claiming benefits, this generates supplementary UI spending to the marginal unemployed workers and tax loss (as employment duration decreases). In practice, the moral hazard cost at the separation margin depends on the precise UI eligibility rules. In many countries, eligibility UI rules require unemployed workers to be involuntary deprived from work, which makes job quitters ineligible to UI.⁴⁰ This limits workers' moral hazard, as employers may be reluctant to pay the extra costs associated to layoffs (minimum severance payments, red-tape, and litigation costs when workers contest dismissals in labor courts). That being said, workers and firms may collude and bargain on the layoff costs internalizing the workers' eligibility for benefits.

The early empirical evidence of UI effects on separations comes from Canada with a series of four papers using the same data from the 1980s (Baker and Rea, 1998; Christofides and McKenna, 1996; Green and Riddell, 1997; Green and Sargent, 1998). To qualify for UI benefits, Canadian workers must work at least 14 weeks over the year before their dismissals. The work experience requirement varies across regions as a function of local unemployment rate. Christofides and McKenna (1996) and Green and Sargent (1998) document spikes in the weekly exit rate from *employment* at the regional experience requirement. This first empirical evidence mirrors spikes found in

³⁹ To account for separation effects, one possibility is to recast the static Baily-Chetty framework such that employment duration is the sum of pre-unemployment and post-unemployment spells.

This formally groups the two margins of adjustment into one.

⁴⁰ In practice, job quitters may become eligible after some waiting period.

the exit rate out of unemployment. While visually appealing, Christofides and McKenna (1996) and Green and Sargent (1998) cannot fully rule out confounding shocks at the regional level (as the source of variation in requirement is driven by local unemployment rate). Green and Riddell (1997) and Baker and Rea (1998) leverage a politically-motivated change in the eligibility rule in 1990 that sets all regions at the maximum 14 week criteria. As there are very few regions at the maximum in 1989, they cannot be used as a precise control group. However, they convincingly show that the spikes follow the requirement change in affected regions (see Panel (a) in [Figure 12](#) which reproduces [Fig. 2](#) from Green and Riddell (1997)).

In [Fig. 13](#), we plot quasi-experimental estimates of separation effects published from the 2000s onwards.⁴¹ We report effects on separation rates of three types of UI variation. In red circles, we plot effects comparing eligible workers vs. ineligible workers for whom any UI claims would be denied (Albanese et al., 2020; Leung and O'Leary, 2020; Van Doornik et al., 2023). They rely on discontinuity in the UI eligibility rules as a function of previous work requirements, eventually coupled with reforms as in Leung and O'Leary (2020). The second and third types of UI variations are at the intensive margin among eligible workers only. Blue squares in [Fig. 13](#) correspond to studies that contrast workers with long vs short Potential Benefit Duration. The green triangle corresponds to the study by Jäger et al. (2020) comparing workers with high vs low replacement rates (had they claimed). Overall, the selected recent quasi-experimental estimates confirm the existence of moderate UI effects that increase separations.

Among prime-age workers (below 50 years old), three studies find statistically significant positive effects (Albanese et al., 2020; Jessen et al., 2023; Van Doornik et al., 2023), and two studies do not reject zero effects (Jäger et al., 2020; Leung and O'Leary, 2020). Albanese et al. (2020) leverages a design with strong credibility. In Italy, workers with above 52 weeks of work experience over the two years before separation are UI eligible if they satisfy a second criteria: they have also worked one extra day *before* the two-year qualification period. Panel (b) in [Fig. 12](#) shows that the layoff hazard rate has spikes at the 52-week threshold for treated workers (satisfying the second criteria), but not for control workers. Studying a reform lowering the minimum work requirement in Brazil, Van Doornik et al. (2023) find strong and precisely

⁴¹ We find two papers providing early evidence in the US: Jurajda (2002) and Light and Omori (2004). While suggestive, they do not leverage quasi-experiments to identify separation effects. Jurajda (2002) analyses past employment duration of a sample of unemployed workers using duration models with unobserved heterogeneity. Light and Omori (2004) leverage variation across states and across year in the overall generosity of benefits, but do not focus on politically-motivated changes. Rebollo-Sanz (2012) confirms the existence of spikes in employment separation at eligibility cutoff in Spain, but does not cast the analysis within a quasi-experimental setting.

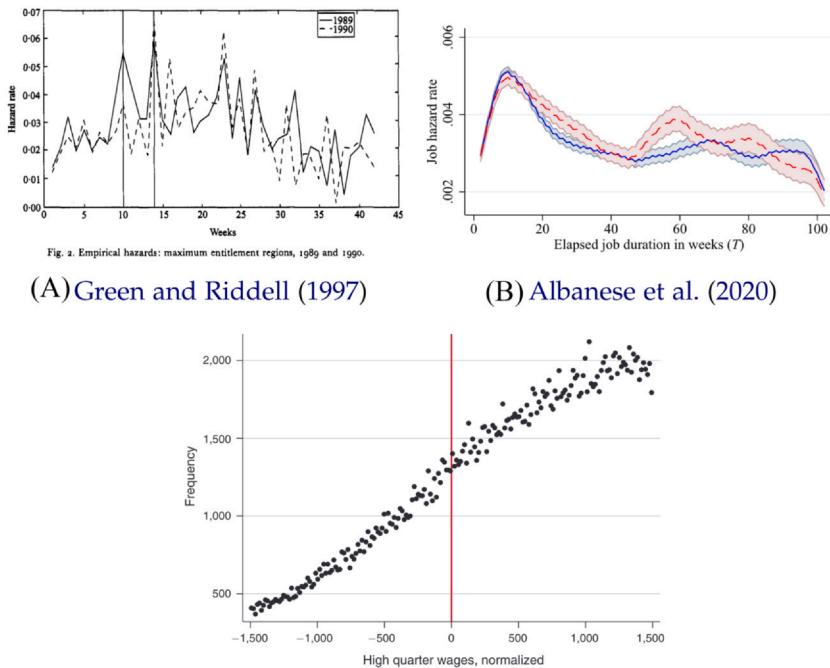


FIGURE 1. DENSITY OF CLAIMS AROUND ELIGIBILITY THRESHOLD

Notes: This figure plots the number of UI claimants in each nonoverlapping \$15 interval of (normalized) high quarter earnings. The vertical line denotes the minimum earnings threshold.

(C) Leung and O'Leary (2020)

FIG. 12 UI effects on job separation. (A) [Green and Riddell \(1997\)](#), (B) [Albanese et al. \(2020\)](#), (C) [Leung and O'Leary \(2020\)](#) **Notes:** The figure reports graphical results from three studies analyzing UI effects on job separation. Panel (A) is Fig. 2 from [Green and Riddell \(1997\)](#). It illustrates the spikes of exit rate from employment at the work experience requirement (10 weeks in 1989 and 14 weeks in 1990). Panel (B) is Fig. 4D from [Albanese et al. \(2020\)](#). It plots the layoff rate as function of job tenure in Italy for two groups of workers. Treated workers (dashed red line) become eligible for UI when they reach the 52 weeks threshold, while control workers are not eligible at any job tenure. Panel (C) is Fig. 1 from [Leung and O'Leary \(2020\)](#). It shows the distribution of high quarter wages for new claimants in the US. On the right-hand side of the vertical red line, claimants are eligible for UI benefits, while on the left-hand side, they are not.

estimated effects on formal employment (with some evidence of substitution towards informal employment). [Jessen et al. \(2023\)](#) study the Polish rules that extend PBD for 6 months in counties with high unemployment rates (compared to the national level). In a RDD design, they estimate that UI eligible workers are 6 % more likely to separate from their employers.⁴² While this effect is

⁴² Note that the 6 % estimate in [Jessen et al. \(2023\)](#) does not capture intertemporal substitution effect around the date when PBD extension is triggered.

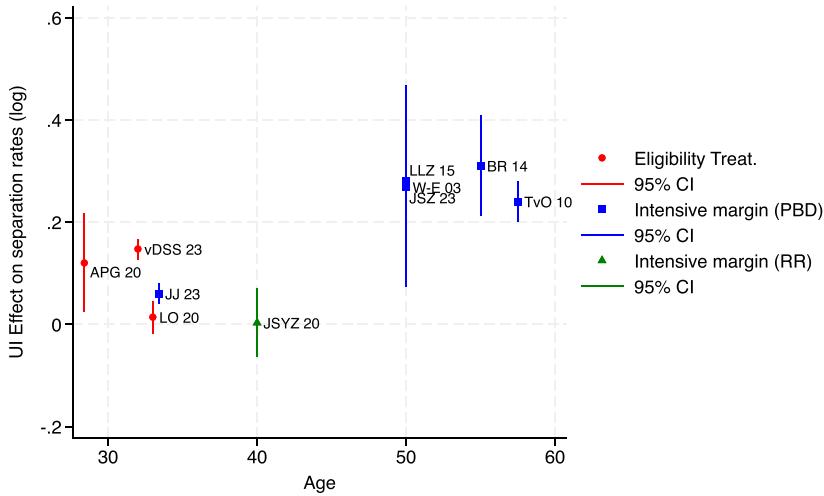


FIG. 13 Recent quasi-experimental estimates of UI effects on workers separation. **Notes:** This figure reports UI effects on separation rates for ten studies (Albanese et al., 2020; Baguelin and Remillon, 2014; Jessen et al., 2023; Jäger et al., 2020, 2023; Lalive et al., 2015; Leung and O'Leary, 2020; Tuit and van Ours, 2010; Van Doornik et al., 2023; Winter-Ebmer, 2003). We plot the main study estimate against the average age in the sample (or the age cutoff for RDD estimates). We distinguish three types of estimates. When the design allows to compare UI eligible workers vs. ineligible workers, estimates are in red (circle markers). When the design compares workers with more or less generous UI benefits (among eligible workers only), we use blue square markers in the case of Potential Benefit Duration contrasts and green triangle markers in the case of Replacement Rate contrasts. Vertical lines correspond to 95 % confidence interval. For Leung and O'Leary (2020) and Baguelin and Remillon (2014), we assume that effects on unemployment inflow rates translate into separation rate effects. The estimates of Winter-Ebmer (2003), Landais (2015) and Jäger et al. (2023) overlap as they use the same DiD design in Austria (REBP).

smaller than the 12–13 % effect in Albanese et al. (2020) and Van Doornik et al. (2023), it is sizeable for a change in benefit generosity at the intensive margin.⁴³ The three previous positive estimates are to be weighed against Leung and O'Leary (2020) who fail to detect manipulation in the US at the past earnings cutoff determining UI eligibility (see Panel (c) in Fig. 12 reproducing their Fig. 1). Note that Leung and O'Leary (2020) analyze unemployment inflow rates rather than separation rates. Interpreting their results as evidence on separation rates implicitly assumes that take-up behaviors do not offset an initial discontinuity in the separation rate at the

⁴³ Hartung et al. (2024) is another study that finds lower transition rates from employment to registered unemployment after the decrease in UI generosity of the Hartz reforms in Germany. As the Hartz reforms are extensive, the authors implement an heterogeneous treatment design around the reform date, which requires stronger identification assumption than the other quasi-experiments in Fig. 13.

qualifying cutoff. This assumption requires further research.⁴⁴ However, the same argument that the absence of manipulation in McCrary (2008) test is implicitly a test of no UI effects on job separation applies beyond the context of Leung and O’Leary (2020) to any RDD studies analyzing UI effects of unemployment duration. We count seven of them in the review by Schmieder and von Wachter (2016) discussed in Section 3.3. This is numerous evidence against UI effects on separations, especially when the underlying UI variation is at the intensive margin. Jäger et al. (2020) also reports a small and insignificant effect of replacement rates on average separation rates in Austria leveraging four large reforms in DiD designs.

The literature is still silent about the sources of the difference in results across the previous studies on middle-age workers. The lack of precise information about UI rules may explain why in some contexts no manipulation is detected, while in other contexts salient reforms make employed workers correctly evaluate their outside options. Beyond information imperfection, the strictness of employment protection and the firms’ separation costs are probably important determinants of the separation elasticity to workers’ outside option. Further research is needed on those explanations.

The UI literature finds stronger separation effects among senior workers (above 50 years old). The evidence comes from large variations in Potential Benefit Duration at the intensive margin (blue squares in Fig. 13). Winter-Ebmer (2003), Landais (2015) and Jäger et al. (2023) all analyze the same 1989 Regional Extended Benefit Period (REBP) reform that extends PBD from 30 to 209 weeks for workers above 50.⁴⁵ Their DiD estimates amount to a separation increase by around 25 %. Tuit and van Ours (2010) and Baguelin and Remillon (2014) find a similar order of magnitude in response to PBD increase from 324 to 405 weeks in the Netherlands and from 200 to 270 weeks in France. One interpretation is that a large PBD increase allows senior workers to bridge the period between job separation and retirement. In such contexts, UI programs implicitly act as early-retirement schemes.⁴⁶ Clear evidence that UI affects job separations via this bridge-to-retirement channel has been found. Recent examples include (Inderbitzin et al., 2016), Kyyrä and Pesola (2020), Riphahn and Schrader (2023), and (Gudgeon et al., 2023).

LESSON 4: The empirical literature of the effects of UI eligibility on job separations is still thin, reporting both positive and zero effects. Separation effects of UI generosity are small for middle-age workers, but can be large for senior workers.

⁴⁴ Anderson and Meyer (1997) report that claiming behavior or take-up depend positively on UI generosity. Note that to falsify the above assumption, one would need strong discontinuity in take up *negatively* correlated with UI eligibility.

⁴⁵ Lalive (2007) and Lalive (2008) analyze the reform effects on unemployed behaviors.

⁴⁶ For example, in France, quarters on covered UI count towards pension eligibility requirements.

The previous studies do not precisely quantify the implied fiscal externality of separation effects. As an illustration, we describe this exercise from [Khoury \(2023\)](#). In France, laid-off workers in firms justifying economic difficulties receive higher benefits (by around 10 %) if their job tenure at separation is above 2 years before 2011 and above 1 year after 2011. [Khoury \(2023\)](#) documents that the 2011 reform does not significantly change the overall number of layoffs, but leads workers and firms to retime layoffs. Namely, bunching estimates show that 10 % of workers who would have been laid off in the month before reaching the one-year cutoff are actually laid off just after the cutoff because of the reform. [Khoury \(2023\)](#) studies the fiscal externality induced by the reform and quantifies two terms. The first term accounts for the behavioral job finding response of laid-off workers who have higher benefits because of their tenure being between one and two years (classical moral hazard of unemployed). The second term captures the extra UI spending on bunchers at the one-year cutoff. Overall, the second term due to separation retiming is marginal compared to the first classical moral hazard term. Further research turning the previous separation effects estimates in [Fig. 13](#) into fiscal externality quantities would be helpful to assess the importance of separation effects for UI design.

In addition to separation effects, standard economic theory predicts that more generous UI would lead workers to exert less effort on the job. The effort reaction is expected to be all the larger that wages do not adjust upwards to support the net value of employment (which seems to be the case given the modest wage effects in the previous section). Empirical research on workers' effort on the job (or productivity) faces a measurement challenge. Workers' effort is not observable in standard dataset (independently of wages). Two recent papers overcome the difficulty by using scanner data from a retail company ([Lusher et al., 2022](#)) or by using sick leave data ([Jäger et al., 2020](#)). [Lusher et al. \(2022\)](#) find that an 18-week PBD extension during the Great Recession in the US leads to a 2 % decrease in cashier scanning speed. [Jäger et al. \(2020\)](#) find with strong statistical power that changes in replacement rates in Austria do not affect time spent on sick leaves of employees. To sum up, the few available estimates point towards modest UI effects on effort on the job.

3.7 Other UI design questions

In this section, we discuss answers to other classical questions on UI design of particular interest for policy-makers. Should replacement rates decrease, increase or stay flat over the unemployment spell? Should UI generosity vary over the business cycle? How does UI interact with other social programs?

3.7.1 *Should benefits decrease / increase over the unemployment spell?*

In the dynamic Baily-Chetty framework analyzed in [Section 3.2](#), we restrict the policy space to constant benefit levels until an exhaustion date when they

drop to zero. In practice, we observe more elaborate designs with time-varying benefits t , that differ across countries. While Belgium and Sweden used to have a constant benefit level that never expires, France implemented in 2021 a decreasing path of benefits where high-wage workers lose 30 % of their unemployment benefits after seven months of unemployment. Should benefits decrease over the unemployment spell?

In their seminal contribution, [Shavell and Weiss \(1979\)](#) give theoretical insights to define the optimal path of UI benefits. They show that declining benefit levels provide powerful incentives for hand-to-mouth unemployed to search for jobs, which significantly reduces moral hazard cost of providing insurance. The intuition follows the job search model already analyzed in [Section 2.2](#). Forward-looking job seekers increase search intensity as the future value of unemployment decreases. Holding constant overall spending on UI (absent any behavioral reactions), front-loading benefits improves workers' welfare by reducing moral hazard costs. However, there are also counteracting arguments supporting an increasing profile. If workers are not consuming hand-to-mouth and have initial wealth at job loss, the insurance agency should delay benefit payments to later periods when the unemployed workers have depleted their wealth. For these workers, marginal utility is higher later in the spell, and a benefit transfer has then higher consumption smoothing value. [Shavell and Weiss \(1979\)](#) conclude that the optimal path of benefits depends on the relative strength of the two channels, and could even result in non-monotonic path (first increasing and then decreasing) as the relative strength may vary over the spell. Following ([Shavell and Weiss, 1979](#)), a large theoretical optimal-contracting literature enriches the environment with more elaborate sources of non stationarity ([Shimer and Werning, 2006, 2008](#)) and broader sets of policy instruments (such as duration dependent wage tax or reemployment bonus in [Hopenhayn and Nicolini \(1997\)](#), [Pavoni \(2009\)](#)). They reach various conclusions from decreasing benefit profiles in [Hopenhayn and Nicolini \(1997\)](#) to slightly increasing profiles in [Shimer and Werning \(2008\)](#). Overall, the theoretical literature emphasizes the importance to introduce an endogenous wedge between consumption and benefits. A direction followed by the recent empirical literature assessing the local optimality of current benefit profiles.

Along those lines, [Kolsrud et al. \(2018\)](#) show that Baily-Chetty formulae apply at each date of the unemployment spell. They consider a small variation of benefits at date t : t . The social value depends on the average marginal utility of unemployed survivors at date t and the corresponding consumption smoothing value. The behavioral costs are due to implied changes in job finding rates over the entire spell. For example, higher benefits later in the spell slow down job finding even of short-term unemployed. [Kolsrud et al. \(2018\)](#) compute the empirical counterparts of the social value and behavioral costs of UI for every month after unemployment entry in Swedish data. They adopt the Consumption-Based approach to estimate the social value of UI along the spell. To identify the elasticity of the exit rate along the spell, they rely on duration-dependent caps on

benefit levels in a RKD design, caps exogeneously shocked by a reform. The rich empirical design allows to estimate the incentive costs of changes in benefits both early versus late in the unemployment spell. As expected, the drop in consumption along the spell translates into higher social value of late benefits. The RKD estimates reveal that incentives costs are larger early in the spell. Consequently, both terms in the Baily-Chetty formulae suggest that increasing locally the benefit profile would increase welfare in Sweden. The fact that the moral hazard channel also leads to benefit back-loading contrasts with the incentive channel emphasized initially in [Shavell and Weiss \(1979\)](#). Surprisingly, estimated behavioral responses to benefit changes early in the spell are more costly than responses to late benefit changes (even though late changes are also found to trigger job finding response early in the spell).

The back-loading recommendation in [Kolsrud et al. \(2018\)](#) contrasts with the positive evaluation of a front-loading reform in Hungary by [Lindner and Reizer \(2020\)](#). The reform changes the time path of benefits, but keeps constant the overall amount of benefits paid over the potential benefit duration. Comparing job finding just before and after the reform date, [Lindner and Reizer \(2020\)](#) find positive effects on exit rates early in the spell. The behavioral reaction generates positive fiscal externality that amounts to 65 % of the mechanical cost of the reform. Positive wage effects imply significant extra positive fiscal externality, even though not statistically significant. Without data on consumption, [Lindner and Reizer \(2020\)](#) cannot assess the social value of the reform.

LESSON 5: The recent empirical evidence challenges recommending benefit schedules with decreasing unemployment benefits over the unemployment spell.

3.7.2 Should benefits vary over the Business Cycle?

Should benefits vary over the Business Cycle? In the United States, federal or state programs provide PBD extension when local unemployment rates cross predefined levels. The US UI system is de facto countercyclical. Its generosity increases in recessions: PBD reached up to 99 weeks during the Great Recession. France is another example with countercyclical UI: since 2022, PBD decreases by 25 % when unemployment rate is below 9 %. On the other hand, many other countries have UI rules without automatic adjustments to business conditions (even though in practice policy-makers may decide discretionary changes in the UI rules depending on labor market conditions). The Baily-Chetty framework allows to assess the pros and cons of cyclical UI rules.

Both the social value and the behavioral costs of UI may a priori vary with business conditions. During recessions, job losers may have lower private savings and thus lower ability to smooth consumption. This would lead to higher insurance value in recessions than in booms. On the behavioral cost side of the Baily-Chetty formula, different theories lead to procyclical or countercyclical elasticities of unemployment duration wrt UI benefits ([Landais et al., 2018a](#)).

One argument for procyclical elasticities is that during recessions vacant jobs are so scarce that job search effort does not significantly improve reemployment prospects. Then unemployment duration elasticities are smaller during downturns. Beyond duration elasticities, the behavioral cost of PBD extensions is determined by the inverse of the share of UI exhaustees (see factor α in Eq. (37)). When labor markets are slack, job findings rates are lower and the share of exhaustees increases mechanically. This second effect of labor market conditions induces procyclical behavioral costs and justifies to increase PBD during downturns. While the theoretical arguments for countercyclical UI are well understood, it is only recently that the UI literature quantifies the cyclical variation in the marginal welfare of benefit increase.

The first important contribution is Schmieder et al. (2012) who find support for countercyclical UI in the data. In Germany, job losers above certain age thresholds have higher PBD, which allows to identify the marginal PBD effect on non-employment duration in RDDs. As the age-specific UI rule remains essentially invariant from the mid-80s to 2005, Schmieder et al. (2012) estimate yearly PBD effects in a consistent way. They find overall low cyclicalities of PBD effects, and, if anything, procyclicalities. Combined with the strong countercyclicalities of UI exhaustees in Germany, Schmieder et al. (2012) conclude that the behavioral cost of UI is procyclical. While they do not have data on the social value of UI to assess its cyclicalities, their empirical evidence points towards countercyclical PBD.

Second, Kroft and Notowidigdo (2016) offer complementary empirical evidence supporting countercyclical UI rules in the US. Kroft and Notowidigdo (2016) analyze consumption data from the PSID as in Gruber (1997). Focusing on within-state time variation in local unemployment rate, they document larger consumption drops when unemployment is higher: -8% in high-unemployment years ($= 1$) vs. -5% in low-unemployment years ($= 0$). Unfortunately, statistical power is low, and such variations are not statistically significant. Other estimates of the social value cyclicalities would be helpful to make progress. In addition, Kroft and Notowidigdo (2016) estimate the unemployment duration elasticity wrt benefit levels (instead of PBD). While PBD changes are correlated with labor market conditions (by US UI design rules), Kroft and Notowidigdo (2016) argue that changes in benefit levels are exogenous and they report that their variations within states over time are not correlated with local unemployment rates. Using within-state time variation, they find that an increase in unemployment by 2.3% points (from 6.2%) attenuates the duration elasticity by 50% (from 0.6 to 0.3). This is a large procyclicalities of duration elasticity.

LESSON 6: The available micro empirical evidence suggests that behavioral costs of UI are lower during recessions.

The Baily-Chetty approach is in partial equilibrium and assumes away externalities and spillovers on uninsured job seekers or on firms and their employees. This assumption may be strong when analyzing UI rules in different labor market conditions. We discuss further macro effects of UI in Section 3.8.

3.7.3 Substitution with other programs from the safety net

After they exhaust their UI benefits, unemployed workers may claim welfare benefits or enter into disability programs. Similarly, job losers who fail to meet the qualification criteria to be eligible for UI benefits may rely on other programs from the safety net. Consequently, changes in UI rules lead to potential substitution effects with other social programs, and the design of UI should take them into account. As emphasized in [Hendren and Sprung-Keyser \(2020\)](#), the UI spillover effects on other programs contribute to the behavioral costs (or savings) part of the UI welfare analysis. If extra spending on UI reduces disability applications, the fiscal externality is positive as the government saves on disability spending. The empirical literature finds suggestive evidence that more generous UI benefit levels decrease applications to Disability Insurance (DI), but effects are imprecisely estimated ([Lindner, 2016](#)). [Mueller et al. \(2016\)](#) do not find meaningful effects of PBD extensions on DI applications during the Great Recession. Using estimates from [Lindner \(2016\)](#), [Hendren and Sprung-Keyser \(2020\)](#) compute that one extra dollar of UI transfer saves \$0.33 in DI spending. The DI spillover fiscal externality is large (about half the median fiscal externality without spillovers reported above). As the initial substitution estimate is imprecisely estimated, there are also large standard errors around the DI fiscal externality and no strong conclusions should be drawn. However, [Hendren and Sprung-Keyser \(2020\)](#) computation illustrates how important spillovers to other programs may be. [Leung and O'Leary \(2020\)](#) studies substitution at the initial eligibility margin. They find that becoming UI eligible at job loss reduces welfare (TANF) receipts by half among low-earnings UI applicants. However overall TANF participation of UI applicants is low, so that transferring one dollar to the unemployed through a softer eligibility requirement saves only \$0.04 in TANF spending.

Finally, the recent empirical UI literature leverages new large and detailed datasets to identify effects on outcomes outside of the labor market, such as health or crime (e.g., [Britto et al., 2022](#)). As providing unemployment insurance typically reduces the occurrence of bad health or crime after job loss, governments may be able to reduce spending in public health programs or in the judicial system. Further research is definitely welcome to better account for those positive spillovers in UI design.

3.8 Micro and macro effects of UI programs

In this section, we discuss effects of UI programs beyond the micro effects on UI beneficiaries. When UI claimants compete for jobs with non beneficiaries, any changes in their job search effort impact the job finding rate of uncovered job seekers. In parallel, UI programs may affect job creation, as they make recruitment more costly for firms or push up equilibrium wages and lower expected profits. The externalities of UI programs on uncovered job seekers and on firms are additional channels affecting the labor market equilibrium.

This section discusses how to account for equilibrium effects when assessing the effects of UI programs on aggregate social welfare and provides recent empirical evidence on their magnitude.

3.8.1 Welfare effects of UI programs in equilibrium

The underlying job search model of the Baily-Chetty approach is in partial equilibrium. It assumes that job finding depends on individual search effort only. To account for externalities and equilibrium effects, we allow individual job finding rates to also depend on aggregate labor market conditions through tightness. To fix ideas, we write the individual job finding rate as:

$$= \theta \quad (59)$$

where ϵ is the individual job search effort and θ is the job finding rate per unit of search effort, which depends positively on the equilibrium labor market tightness (θ). Tightness is defined as the ratio of vacancies posted by firms over aggregate search effort ϵ . In equilibrium, tightness is determined by aggregate labor supply and labor demand. The key idea of the equilibrium analysis is that UI programs affect labor market tightness through the reactions of claimants and of firms. In such random matching models, the decentralized equilibrium is generally inefficient (unless the Hosios condition is satisfied when wages are bargained under Nash, see Pissarides 2000). Consequently, more generous UI programs increase social welfare if they push tightness towards its efficient level (Landais et al., 2018a).

More precisely, Landais et al. (2018a) develop a theory of optimal UI in matching models. They assume random search. Firms post vacancies to recruit, which costs the wages of the employees that they allocate to recruitment activities. Firms have decreasing marginal returns to labor. Those assumptions generate downward sloping labor demand curves in the tightness - employment plane, even when wages are fixed. On the supply side of the market, workers choose search effort. The model has a general mechanism for wage setting (reduced-form, not micro founded). It assumes that wages depend on tightness and the net value of employment for workers. This allows for rigid wages or Nash bargaining models. In such a framework, Landais et al. (2018a) derive the following formula for the optimal UI replacement rate (ϵ):

$$= \underbrace{\frac{1}{\varepsilon} \left[\frac{1}{\theta'} - \frac{1}{\theta} \right]}_{-} \underbrace{\left[-\frac{\varepsilon}{\varepsilon} \right] \left(\theta' \theta \right)}_{-} \quad (60)$$

where ε and ε are the micro and macro elasticities of unemployment wrt UI and θ' is an efficiency term measuring how far the current equilibrium tightness θ is from the efficient level θ . The macro elasticity ε measures the total response of unemployment to a change in UI, when all endogenous reactions are allowed. It takes into account labor demand reactions, and wage

adjustment. The micro elasticity ε measures the change in unemployment due to the reaction of unemployed search effort, holding tightness constant. For the sake of space, we do not detail here the expression of the efficiency term θ , but we note that it is zero when the equilibrium is already efficient ($\theta = \theta^*$).

The first implication of the optimal replacement rate formula is that spillovers matter to the extent that they create a wedge between the micro and macro elasticities. If tightness is not affected by UI generosity, then the micro and macro elasticities are equal. The equilibrium correction disappears from the formula and only the first term remains. It corresponds to the optimal replacement rate from the partial-equilibrium Baily-Chetty formula. It is then sufficient to trade off the consumption smoothing benefits of UI with the incentive cost due to lower search intensity.

When the macro and micro elasticities differ ($\varepsilon \neq \varepsilon^*$), the equilibrium correction matters to the extent that $\theta \neq \theta^*$. For example, if wages are bargained under Nash with bargaining power equal to the elasticity of the matching function (Hosios condition), then the decentralized equilibrium is efficient and the marginal effect of tightness on social welfare is zero. When the baseline tightness is away from θ^* , the efficiency term $\theta - \theta^*$ can be positive or negative depending on whether a tightness increase pushes towards θ^* or resp. away from θ^* . Following up on the same example, under Nash bargaining, the decentralized tightness may be lower or higher than θ^* , depending on the workers bargaining power.

In the end, whether the equilibrium correction increases or decreases welfare depends on the product of both the signs of the efficiency term and of the elasticity wedge factor. When the macro elasticity is greater than the micro elasticity ($\varepsilon > \varepsilon^*$), spillovers reduce tightness following a UI benefit increase. If baseline tightness is inefficiently low ($\theta < 0$), macro effects generate a social welfare cost, and the macro optimal replacement rate is lower than the Baily-Chetty replacement rates. On the contrary, if tightness is inefficiently high ($\theta > 0$), macro effects increase social welfare, pushing upwards the macro optimal replacement rate.⁴⁷ Formula (60) highlights the critical role of the macro-micro elasticity wedge. Following Landais et al. (2018a), a recent but growing empirical UI literature seeks to identify this key parameter. Before reviewing it in the next section, this raises one last theoretical question. The standard search and matching model with Nash bargaining unambiguously predicts that UI lowers labor market tightness (through wage pressures). Which matching models predict the opposite?

Landais et al. (2018a) show that the job rationing model of Michaillat (2012) features lower macro elasticity than micro elasticity. This is due to a *rat-race* channel that limits macro UI effects on unemployment. The intuition can be better conveyed by assuming rigid wages. Let us also assume that wages

⁴⁷ In random search models, tightness is inefficiently high, when workers bargaining power is low.

are relatively high, so that jobs are rationed.⁴⁸ For the sake of the illustration, let us take the extra assumption that the number of jobs is fixed. When an individual unemployed reduces job search effort, this mechanically increases the employment opportunities of competing job seekers. Low-search-effort unemployed go down the queue in front of the jobs, shifting up other candidates in the queue. Consequently, the job finding rate per unit of search increases and so does tightness. Going back to the job rationing model with endogenous job creation, the rat-race effect is stronger when labor demand is less elastic (in the tightness - employment plane).

3.8.2 Empirical evidence on spillovers and macro effects

Following the Great Recession and the theoretical contribution of Landais et al. (2018a), a fast growing UI literature seeks to produce empirical evidence on spillovers and macro effects. The objective is quite different from the micro empirical strand that we reviewed extensively in Section 3.3. Seeking clean identification, the micro literature generally compares treated and untreated unemployed who are as close as possible from one another. Mimicking the experimental paradigm, it looks for individual-level exogenous variations that randomize unemployed in treatment groups with higher UI benefits. One important requirement is then to compare treated and untreated unemployed who search for jobs in the same labor markets in order to hold labor market conditions constant. Consequently, by design, the micro literature does not identify UI effects on aggregate labor market tightness. In addition, while refining the individual-level comparison, the micro method tends to select samples of untreated unemployed who are the most likely to compete with the treated unemployed. As a consequence, the micro elasticity estimates may capture large rat-race effects that only a macro approach would allow to properly account for when concluding about welfare effects. This identification issue is a classical trade off between the CIA and the SUTVA assumptions in the design of credible evaluation.

To quantify the importance of macro effects, the empirical literature develops designs that mimick randomization at the market level. Market-level quasi-experiments identify the macro elasticity of unemployment needed in Formula (60). They also document the nature of spillovers by identifying market-level changes in vacancies and aggregate wages and by identifying UI effects on the job search of non UI claimants in treated markets.

Lalive et al. (2015) find a significant *rat-race* channel in Austria in the late 1980s—early 1990s. They leverage the REBP policy shock that extended PBD by almost 3 years for senior unemployed in 28 of the 100 regions in Austria (micro UI effects of the REBP are previously documented in Lalive (2007, 2008), Winter-Ebmer (2003)). The setting is particularly well suited for

⁴⁸ With high wages, decreasing marginal returns would prevent firms to hire all workers even if tightness is zero.

identifying macro effects, as some regions experience an increase in UI generosity while others are left untreated. Even if treated regions are selected based on their industry composition, the DiD assumptions are reasonably met. The setting is even better suited for documenting market externalities, as some unemployed in treated markets are not eligible for PBD extension because of their exact age or work history. DiD estimates show that ineligible unemployed have lower unemployment duration in treated markets than in non-treated markets (2 to 8 weeks depending on the non-eligible group considered). The spillovers on ineligible workers (and the absence of wage effects) imply that the macro elasticity is 20 % lower than the micro elasticity.

Five studies provide macro UI effect estimates for the US during the Great Recession ([Boone et al., 2021](#); [Chodorow-Reich et al., 2019](#); [Hagedorn et al., 2013](#); [Johnston and Mas, 2018](#); [Marinescu, 2017](#)).

[Marinescu \(2017\)](#) analyzes a state-level dataset on job applications and vacancies from CareerBuilder.com, a very large American online job board, and uses as source of UI variation the PBD extension from the federal EUC and state-level programs. While aggregate job applications decrease by 1 % when PBD increase by 10 %, there is no robust effect on vacancies. The empirical evidence is consistent with the job rationing model with rigid wages, where labor demand is inelastic. The macro-micro wedge amounts to 39 %.

[Johnston and Mas \(2018\)](#) study a large PBD cut in Missouri in 2011 (unanticipated and politically motivated). To obtain the micro elasticity, they contrast the unemployment duration of UI claimants who claim just before or just after the reform date (RDD), which holds tightness constant. They estimate the macro elasticity comparing the evolution of the aggregate unemployment in Missouri to a synthetic control made up of several other US states. Their micro and macro estimates are consistent with the absence of spillovers on unemployed who are not directly affected by the reform (zero macro-micro wedge).⁴⁹

[Chodorow-Reich et al. \(2019\)](#) analyze PBD extension state-level events that occurred because of measurement error in real-time unemployment statistics (observed ex-post thanks to revised unemployment series). This allows to overcome the mechanical link between unemployment and benefit extension in the US during the Great Recession. In their baseline specification, they find that increasing PBD by one month generates at most a 0.02 % point increase in unemployment rate. Similarly, they find small insignificant effects on state-level vacancies and worker earnings. [Chodorow-Reich et al. \(2019\)](#) find smaller macro effects than ([Johnston and Mas, 2018](#)). Assuming that the micro elasticity of [Johnston and Mas \(2018\)](#) applies to the context in [Chodorow-Reich et al. \(2019\)](#), their result implies significant spillovers during the Great Recession.

⁴⁹ In a recent study, [Jessen et al. \(2023\)](#) also find that macro and micro elasticity are equal in Poland. They further show the absence of spillovers on untreated unemployed and of effects on tightness.

[Boone et al. \(2021\)](#) implement another strategy to circumvent the mechanical link between unemployment and benefit extension during the Great Recession in the US: a *border design*. The design compares two adjacent counties in neighboring states that experience different benefit extensions. [Boone et al. \(2021\)](#) rule out that one month increase in PBD decreases the employment to population ratio by more than 0.02. Their result is of similar magnitude as [Chodorow-Reich et al. \(2019\)](#), but much less negative than the first border-design estimates by [Hagedorn et al. \(2013\)](#).⁵⁰

Overall, the US evidence from the Great Recession suggests a limited role for UI extension in the aggregate unemployment increase and the subsequent slow recovery. Except in [Johnston and Mas \(2018\)](#), macro effects are smaller than micro effects because of spillovers on uncovered unemployed. An alternative explanation lies in the stabilization role of UI expansions as they support aggregate demand ([McKay and Reis, 2021](#)). Further research is needed especially on quantifying the aggregate demand externality.

While all the previous studies allow to estimate the micro-macro wedge, they do not inform on the efficiency term and more importantly on its sign. [Landais et al. \(2018b\)](#) develop an identification strategy for the efficiency term based on proxies for recruitment costs (namely the share of workforce dedicated to recruiting).⁵¹ They find that tightness is inefficiently low in recessions and inefficiently high in good times. Together with a macro-micro wedge lower than one, the efficiency term estimates suggest countercyclical UI even when accounting for market-level externalities.

To summarize, our review of the recent literature allows us to draw the following lesson.⁵²

LESSON 7: Recent empirical evidence from seven studies suggests that macro elasticities are not larger than micro elasticities and possibly smaller.

3.9 Discussion

To conclude the section on UI policy design, we put together the different lessons that we draw for UI policies, and discuss other avenues for future research.

After presenting the Baily-Chetty framework and its extensions, we reviewed the recent empirical assessment of the behavioral costs and social value of UI and drew the following lessons:

- 1. LESSON 1: The behavioral costs of providing UI are substantial.**
- 2. LESSON 2: Estimates of the social value of UI differ widely across identification methods. The most recent methods which are robust to risk-aversion assumptions yield significantly higher estimates.**

⁵⁰ Differences come from the use of different data sources and ([Boone et al., 2021](#)) estimating more flexible econometric models.

⁵¹ The efficiency terms also depend on marginal social cost of unemployment: fiscal cost and non-pecuniary cost on unemployed.

⁵² See [Cohen and Ganong \(2024\)](#) a recent review of micro-macro elasticity wedge.

3. **LESSON 3:** In the majority of recent studies, more generous UI policy decreases wages imposing further (second-order) behavioral cost to provide UI.
4. **LESSON 4:** The empirical literature of the effects of UI eligibility on job separations is still thin, reporting both positive and zero effects. Separation effects of UI generosity are small for middle-age workers, but can be large for senior workers.
5. **LESSON 5:** The recent empirical evidence challenges recommending benefit schedules with decreasing unemployment benefits over the unemployment spell.
6. **LESSON 6:** The available micro empirical evidence suggests that behavioral costs of UI are lower during recessions.
7. **LESSON 7:** Recent empirical evidence from seven studies suggests that macro elasticities are not larger than micro elasticities and possibly smaller.

While the Baily-Chetty model and its extensions to market-externalities provide a powerful framework to discuss the design of UI, it relies on optimizing neoclassical agents. In [Section 2](#), we reviewed recent empirical evidence that is more in line with behavioral models with agents whose preferences exhibit reference dependence, present bias, or incorrect beliefs. Welfare effects with this type of agents can differ substantially from the predictions of neoclassical models (for example, [Mueller et al. \(2021\)](#), [Spinnewijn \(2015\)](#) discuss how to account for biased beliefs). Further research would be helpful to cast the optimal UI analysis further into behavioral models.

At least two other strands of the literature on optimal UI make recent advances and are only touched upon in [Section 3](#). First, new data allow to study UI designs in the developing world. This literature discusses classical issues related to informality from a new perspective ([Gerard and Gonzaga, 2021](#); [Liepmann and Pignatti, 2024](#); [Van Doornik et al., 2023](#)). It also delivers first rate empirical evidence about UI effects on crime and domestic violence ([Britto, 2022](#); [Britto et al., 2022](#)). The UI literature in developing countries is also interested in the joint design of severance pay and UI ([Gerard et al., 2024](#)), and in the design of alternative insurance policies such as individual saving accounts ([Hartley et al., 2011](#)).

Second, in the wake of the COVID pandemics (2020–2022), there is a regained interest in exceptional crisis UI policies ([Ganong et al., 2022](#)) and how short-time work policies complement or substitute UI policies (see the corresponding chapter by Cahuc in the same handbook). Short-time work (STW) policies allow firms to reduce working hours of their employees on a temporary basis, while they receive public subsidies to replace their wages. Workers under the STW scheme are in the gray area in-between employment and unemployment. In the US, workers in the gray area in-between employment and unemployment are targeted by another program: partial UI. Partial UI

allows UI claimants to work in low-earnings jobs and still receive some UI benefits. Recently, [Lee et al. \(2021\)](#) and [Le Barbanchon \(2020\)](#) study the design of partial UI.⁵³ Further analysis on those “gray-area” programs would be particularly helpful for UI design.

Another first-order design question is how to complement UI policy with Active Labor Market Policies. We provide answers of how ALMPs work in the next section.

4 Active labor market policies

We have seen in the previous sections that job search and especially the choice of search effort are important parameters determining labor market outcomes. UI programs provide income support that allows unemployed workers to smooth consumption after an income shock but also triggers behavioral responses due to moral hazard. Active labor market policies (ALMPs) have been designed to complement passive benefit programs and address labor market frictions. The policies have three broad strategies to achieve their goals. First, ALMPs aim at reducing moral hazard from benefit receipt by imposing search requirements and monitoring search efforts. Second, they aim at improving the efficiency of individual job search and speeding up the return to employment of unemployed job seekers by providing counseling. Third, they aim at reducing skill mismatch in the labor market by providing skill training to low wage and unemployed workers, allowing them to access better paid jobs and thus improving their labor market outcomes. Following these goals, a large number of different ALMPs have been developed around the world. These programs differ widely depending on the policy objective, the program content, the target population, and in terms of program costs. From policy design, it is extremely important to understand if these programs work and how they achieve their results. Do ALMPs help integrating job seekers in the labor market or improving labor market outcomes of participants? Under which conditions are the programs cost effective? Do programs have unintended externalities for example due to displacement of non-participating job seekers?

In this section, we first review existing meta-analyses and summaries of the ALMP evaluation literature which comprehensively cover studies written until 2014. Then we turn to the most recent contributions that were written over the last 10 years, a period over which we observe a massive expansion of the use of and interest in ALMPs worldwide. We review the recent literature with the aim of highlighting the areas where it has advanced. We organize the review in 10 main lessons drawn from studies that make significant contributions that go beyond standard program evaluation exercises. [Tables 6 and 7](#) provide a list of 37 studies included in our review along with their main features.

⁵³ see [McCall \(1996\)](#) for an early analysis of employment effects of partial UI in the US

We see the review of the ALMP literature in this chapter as a complement to a systematic meta-analysis. The meta-analysis would cover a much larger range of studies to allow a systematic review of program effects found in the literature. But it would have to restrict the focus of studies to a set that fits into a common scheme, for example in terms of program types and outcome variables. Here, we are not so much interested the systematic review of program effects on certain outcomes, as in pointing out novel and promising approaches in program design, research questions regarding the economic and welfare impacts surrounding ALMPs, and contributions in evaluation strategies.⁵⁴ To visualize the overlap between the two approaches, column (1) in [Tables 6 and 7](#) shows an indicator whether the study fits in the ALMP meta-analysis scheme adopted by [Card et al. \(2018\)](#).

4.1 Meta analysis studies

The earliest program evaluation studies of ALMPs go back to the 1970's and 80' in the US ([Ashenfelter, 1987](#)). In Europe interest in the effectiveness of ALMP programs sparked in the mid 1990's at a time when problems with unemployment were high on the political agenda and detailed micro data on program participants and their labor market outcomes increasingly became available. Since then the number of ALMP evaluation studies has been exploding around the world. [Card et al. \(2010\)](#) and [Card et al. \(2018\)](#) performed meta-analyses of studies written from 1995 until 2007 and 2014, respectively. They focus on evaluations of active programs that are targeted at individuals who participate in the program for a limited time. Further, they restrict the analysis to cover studies based on micro data which apply a treatment and control group design with some form of selection correction. The second meta-analysis relies on a cumulative number of 207 studies. [Card et al. \(2018\)](#) categorize programs into five main types. *Job search assistance* (JSA) programs, which provide counseling to job seekers and monitor the search effort of benefit recipients, either in personal meetings with case workers or in group workshops. *Training* programs focus on human capital enhancement either in classroom training or a combination of on-the-job and off-the job training. *Employment subsidies in private sector jobs* aim at bringing job seekers into employment and rely fully on on-the-job human capital enhancement. *Public sector employment programs* create special jobs in the public sector which employ unemployed workers in sectors where they are not competition with the private market. The fifth program type includes programs which combine features of different types, for example training programs with a placement assistance.

To construct the meta-data, the authors extract program effect estimates for 5 main program types and different participant groups over three time horizons,

⁵⁴ In fact, our focus on promising program designs potentially introduces a bias towards examples of successful ALMPs. Our question is not so much what works on average? It rather is, what can work?

estimating short-run effects (in the first year after program completion), medium run (1–2 years after the program), or long run (more than 2 years after the program). In total Card et al. (2018) collect a sample of 857 separate program estimates from all studies. The main outcome variable considered in about 40 % of the collected estimates is the employment rate. Other outcomes are earnings or exit rates from unemployment or into employment. The sample of estimates covers a wide range of countries worldwide. But the majority are from Europe and the US, such that Germany, Denmark, France, and the US make up more than 50 % of the sample. In terms of program types, about 50 % of the estimates evaluate impacts of training programs.

Card et al. (2018) report four main findings from their meta-analysis. First, short run program effects tend to be small but average program effects improve in the medium and longer run. Second, the time profile of program effects varies by program type. While job search assistance programs have relatively stable effects over different horizons, the effects of programs with a human capital component improve strongly over time. Third, there is some evidence of heterogeneity of program effects for different participant groups and of potential gains from matching specific participants to specific program types. But due to small sample sizes these results were not stable. Fourth, program effects vary with cyclical conditions and the authors find that ALMPs have larger impacts in times of low GDP growth or high unemployment.

On the methodological side, Card et al. (2018) find that average effects reported by experimental designs—about 30 % of the sample—are not systematically different from non-experimental estimates. Neither do average effects differ systematically between published and non-published studies. The meta-analysis does not give evidence of other forms of “publication bias” either, which is fairly uncommon in the meta-analysis literature. It appears that researchers or referees did not have a strong preconception that ALMPs necessarily have positive effects or that non-positive findings are uninteresting. Furthermore, many studies were conducted in close collaboration with the administration operating programs which increased the exposure of many different types of findings. Another remarkable insight from the meta-analysis of program estimates is a large dispersion in program effects even among estimates with high levels of precision. Typically, we would assume that as sample sizes increase and estimates become more precise the range of estimates should get closer to the “true” program effect. In the ALMP literature it seems to be hard, however, to nail down a “true” effect. Even within relatively narrow categories of program types there appears to be a large degree of unobserved heterogeneity in program effects that cannot be explained by sampling error. The origin of this heterogeneity is probably due to differences in institutional environments, participant groups, or program implementations.

Limitations of the major part of ALMP evaluation studies written by 2014 were twofold. First, few studies provided a detailed cost-benefit analysis or precise measures of program costs which made it impossible to assess the cost

efficiency of different programs. Second, the evaluation studies focused on partial equilibrium effects comparing the mean outcome in the treatment group with the mean outcome in the control group. At the time, there were little concerns about general equilibrium effects or potential spillover and displacement effects (except [Crépon et al. \(2013a\)](#)). This reduces the external validity of the findings and leaves many questions regarding program scalability unanswered.

Recent reviews of the literature include [Crépon and van den Berg \(2016\)](#) who review evidence from experimental studies and discuss policy relevant and methodological questions that advance the literature. [McCall et al. \(2016\)](#) provide a long run review of the literature on government sponsored vocational education in the US and 5 European countries (UK, Denmark, Sweden, France, Germany) with a focus on training provided via ALMPs. [McKenzie \(2017\)](#) reviews evidence from ALMP evaluations in developing countries.

4.2 Lesson 1: New insights in the role of caseworkers

At the employment office job-seekers interact with caseworkers who provide job search assistance and counseling, monitor search activity, and impose benefit sanctions for non-compliance with search requirements. In addition, caseworkers refer job seekers to ALMP programs. This means that caseworkers potentially have an important role in shaping search success and longer-run outcomes of job-seekers. A number of recent studies provide valuable insights about the role of caseworkers. This literature has overcome two important empirical challenges. Firstly, by accessing novel administrative data that contain information on individual matches between job-seekers and caseworkers. Secondly, by exploiting random assignments of job-seekers to caseworkers either via an RCT design or by exploiting quasi-random assignments in the institutional setting.

Evidence from the US: The use of caseworker resources varies widely across countries. In particular, they are less systematically used in the US than in Europe. Rising unemployment rates at the start of the Great Recession led to increased funding and renewed the interest in so-called *re-employment programs* in the US. The main purpose of this JSA program was to assess benefit eligibility and monitor search effort of new benefit recipients. In addition, some programs also provided job search counseling to unemployed workers with positive eligibility reviews. [Michaelides and Mueser \(2020\)](#) evaluate four programs in three US states that were implemented in an experimental design. Potential participants were randomly assigned to receive letters inviting them to meetings with a caseworker. At the meetings an attendant's eligibility status was assessed and non-eligible benefit recipients were disqualified. If the program had a service component the remaining attendants received counseling services, information about other job search services, and referrals to training programs. Outcome data from administrative records of the employment office

reveal that all programs substantially shortened participants' unemployment benefit receipt durations either because individuals did not attend the meeting or because they were disqualified after the assessment. While programs focused on eligibility monitoring only had short-lived effects, programs with a counseling component also succeeded in increasing employment and earnings of participants over the first year after program assignment.

An evaluation of longer run effects of the most successful re-employment program implemented in Nevada, shows that these employment and earnings effects even persisted over a longer horizon of up to 8 years ([Manoli et al., 2018](#)). Re-employment programs were first expanded during the Great Recession but in many locations they also continued during the post-recessionary period which allows an examination of the effects under different labor market conditions. [Michaelides and Mueser \(2023\)](#) exploit the fact that the program design as well as the random assignment mechanism of the Neveada REA/RES program was more or less unchanged for a nine year period. They show that the beneficial employment and earnings effects can also be found in a tighter labor market. Similarly, [McConnell et al. \(2021\)](#) report positive employment and earnings effects of intensive counseling services in the Adult and Dislocated Worker Programs implemented in the post-recessionary period.

Evidence from the Europe: Many European countries rely more heavily on caseworker meetings and enforce stricter search monitoring of unemployment benefit recipients than the United States. In these systems benefit recipients are required to attend regular meetings with case workers and to keep detailed records of their search activity which is monitored by the caseworker ([Maibom et al., 2023](#)). [Schiprowski \(2020\)](#) examines the effect of caseworker meetings on the exit rate from unemployment. She exploits random variation in caseworker absences in Switzerland and finds that a canceled meeting leads to a 5 % increase in unemployment duration. The impact of a caseworker absence is positively related to caseworker productivity. Missing a meeting with a below median productive caseworker has no impact on unemployment duration but missing a meeting with a highly productive caseworker is strongly detrimental. Another piece of evidence on the effectiveness of caseworker meetings is provided by [Schiprowski et al. \(2024\)](#), who use the SMS data from [DellaVigna et al. \(2022\)](#) to study the dynamics of job search effort around caseworker meetings. They find that search effort increases modestly just before a caseworker meeting and falls back afterwards. Caseworker meetings that are accompanied by a formal agreement on search effort between workers and caseworker are more effective, as are vacancy referrals by caseworkers.

[Cederlöf et al. \(2021\)](#) systematically estimate the value added of caseworkers in Sweden with respect to job finding and job quality. They exploit quasi-random assignment of job seekers to caseworkers by the date of birth within local employment offices. It turns out that caseworker value added has substantial impacts on search outcomes along both outcome dimensions. Replacing the lowest quartile of caseworkers with an average caseworker

would shorten unemployment durations by about 10 %. High value added caseworkers also affect earnings and employment outcomes of job-seekers in the long run. Consistent with the literature on teachers, the Swedish results imply that caseworker value added is multi-dimensional. Different types of caseworkers reduce unemployment durations or increase employment and earnings outcomes. A caseworker characteristic that is strongly correlated with value added is caseworker experience. There are also potential gains from matching caseworkers and job seekers who are similar in terms of labor market experience.

A further strand of the literature investigates the timing and frequency of caseworker meetings. Meetings early in the unemployment spell might have a threat effect of pushing reluctant job seekers off the benefit roll but they are also more costly as they involve a larger number of job seekers some of whom might find jobs without external help. [Homrighausen and Oberfichtner \(2024\)](#) investigate a program in Germany which randomly assigns offers of caseworker meetings to pre-registered job-seekers who have not lost their jobs yet.⁵⁵ While the early meeting invitations increase the number of meetings attended during the first months of the unemployment spell, early meetings are not successful in reducing inflow into unemployment nor do they speed up re-employment.

[Maibom et al. \(2017\)](#) explore the effects of a set of Danish programs which vary the timing and frequency of caseworker meetings in an experimental design. They find that early individual caseworker meetings significantly improve employment outcomes of job seekers. But programs designed to generate a threat effect do not have a significant impact. [Böheim et al. \(2022\)](#) evaluate an experiment which exogenously varies caseworker caseloads at the employment office. They find that caseworkers with lower caseloads schedule more meetings per job seeker and the additional meetings result in small positive employment effects over the next 2 years.

Evidence from programs that are especially designed for disadvantaged workers shows that these workers may benefit from caseworker assistance not only during job search but also from job coaching and career advice *after* a job has been found. In the next section we will discuss how so-called *wraparound support services* which include coaching and follow-up services after program completion lead to more favorable job changes ([Bobonis et al., 2022; Katz et al., 2022](#)).

4.3 Lesson 2: More focus on programs for special groups

Most of the ALMP programs surveyed in the meta-analysis by [Card et al. \(2018\)](#) were available to broadly defined intake groups, such as unemployed, UI benefit recipients, or low income workers. Evaluation studies often estimated treatment effects for different groups to investigate effect heterogeneity

⁵⁵ In Germany workers have to register at the employment office as soon as they receive a layoff notice from their employer or at least 3 months before the termination of a temporary job.

by gender, age and other characteristics. But the needs of individual job seekers are potentially very different and there are limits of one-fits-all programs targeted to broad intake groups. The recent literature presents several studies evaluating programs that were specifically designed for particular target groups. Here we discuss examples of programs for disadvantaged workers, immigrants, and youths.

Disadvantaged workers are individuals with low labor market attachment, low income and low education. [Card et al. \(2018\)](#) find that ALMPs tend to be more successful for participants from disadvantaged groups. In particular, they benefit more JSA programs than registered unemployed. The task of ALMP is to integrate disadvantaged groups in the labor market and to help them find and keep good jobs with career opportunities. Mostly this is achieved by increasing their occupational skills. But in line with the success of search assistance for disadvantaged workers suggested by the meta-analysis, these workers might also benefit from training in soft skills helping them to overcome psychological barriers and improving their work attitudes, behavior and decision making. The literature discusses several approaches to addressing the lack of participants' non-cognitive or cognitive skills. They consist of intensive counseling services, non-cognitive skill training, combinations of skills training and counseling, and wage subsidies.

[Bobonis et al. \(2022\)](#) evaluate the Self-Sufficiency Project (SSP) Plus program that was implemented in a randomized control trial in Canada in the 1990s. The main branch of the program, SSP Regular, offered time limited financial incentives to take up employment in the form of earnings supplements to single parents who were long-term income assistance recipients. In addition to earnings supplements, SSP Plus offered intensive employment support services during and after job search. In particular, program participants were matched with individual caseworkers who pro-actively offered counseling and advice for job search and career advancement over a period of up to four years. Take-up of this service was voluntary for program participants.

[Bobonis et al. \(2022\)](#) evaluate long-run effects of SSP Plus relative to SSP Regular over a 20 year horizon. Their results show that the intensive support services led to substantial and long-lasting earnings gains compared to the regular SSP program where earnings gains faded quickly after the earnings supplements had expired. SSP Plus participants experienced an increase in full-time employment and a decrease in receipt of welfare benefits. Looking into the mechanisms driving these effects, the authors find that the support service helped participants to move up the career ladder towards better paid and more stable jobs. Survey evidence also shows an improvement in non-cognitive skills and measures of grit.

Sectoral employment programs in the US studied by [Katz et al. \(2022\)](#) target young low-wage workers with less than high school education. These programs offer a package of measures combining soft-skills training, occupational skill training and career support services which start with the job

search period and extend to the post-employment period where they provide retention and advancement services. Similar to the Canadian SSP Plus program, sectoral employment programs succeed in persistently raising participant's earnings and job quality in the medium run. While it is challenging to disentangle the contribution of the different program components, the available evidence suggests that wraparound services are essential complements to occupational skill training. Career counseling, in particular, helps job seekers with non-traditional careers get access to jobs in high paying firms.

[Schlosser and Shanahan \(2022\)](#) present an experimental evaluation of *employment circles* in Israel, a program that offers training in soft and occupational skills along with frequent caseworker meetings. The program is targeted at income support recipients a group with low earning and employment prospects, who are, however, required to search for jobs in order to keep their benefits. The authors argue that employment circles can have two potential effects on participants, first a threat effect due to enforcement of program requirements that will mainly reduce welfare recipiency and second, a skill enhancement effect that should lead to increases in employment and labor earnings. The study compares effects among two participant groups, new entrants into income support receipt and long-term support recipients. Evaluation results show that the program increases measures of soft skills such as grit and the motivation to search for jobs and to work in the group of long-term support recipients. This group also gains from the program in terms of employment and labor earnings along with reduced income support receipt. Among new benefit entrants the threat effect of the program dominates, however. While there are only small gains in employment among new entrants assigned to the intensive program, they are more likely to leave the benefit system.

[Kasy and Lehner \(2023\)](#) study a public sector employment subsidy in Austria with the aim of “eradicating long term unemployment”. The program offers 3 years of subsidized employment to long-term unemployed workers. The wage in subsidized jobs is determined by collective bargaining agreements which raise wage earnings of participants above the UI benefit level. Eligible workers were randomly assigned to enter the program in two cohorts which allows evaluating short run program effects. Due to the wage incentives, program take-up is extremely high which in turn implies positive short run employment and earnings effects. Beside labor market outcomes, the focus of the study is on health, job satisfaction and societal well-being of participants. All these outcomes are positively affected by the wage subsidy.

Immigrants and Refugees In many countries immigrants have substantially lower employment rates than natives even many years after arrival ([Brell et al., 2020](#)). Especially low-skilled immigrants from low-income countries or refugee immigrants who relocate involuntarily and have to cope with traumatic experiences face substantial problems with labor market integration. A small literature that is mostly focused on Northern European

countries investigates whether and how these problems can be overcome by specially designed ALMPs. [Arendt et al. \(2022\)](#) and [Foged et al. \(2024\)](#) review the effects of different types of welfare and integration policies, which include ALMP, on the labor market performance of refugees in Denmark. Denmark has admitted refugees over a long period during which the policy environment changed multiple times. Together with detailed longitudinal data this provides an ideal setting of studying these policies.

Generally, the literature finds only moderate effects of ALMP participation on immigrants' or refugees' labor market outcomes. Explanations for the moderate effects relate to the multi-dimensionality of problems faced by participants and to conflicting incentives within the institutional setting ([Arendt et al., 2022](#)). For example, immigrants who are uncertain about their residency status face low incentives to search for jobs or trainees may not be able to fully benefit from training programs because of language problems. The most promising programs are those that closely target program content and the sequence of measures to individual needs. A novel Finnish program combines individualized training plans with intensive language courses for immigrants. [Sarvimäki and Hämäläinen \(2016\)](#) exploit a discontinuity in program eligibility during the program roll-out when eligibility depended on the date of entry into the population register. Based on this design, the authors find that individualized plans are more successful in improving immigrants' earnings in the long run than traditional ALMPs. Another successful approach relies in specialized programs that focus on training newly arrived refugees in target occupations with labor shortages in the local labor market ([Dahlberg et al., 2024; Foged et al., 2022b](#)) which we will discuss in the next section.

The literature is more encouraging, however, on the high value of language courses for the economic integration of immigrants. Generally, selection issues make it difficult to evaluate the effect of language skills on labor market outcomes. But several recent well-identified studies relying on discontinuity designs or randomized assignments document credible and positive effects of intensive language training. In Denmark, [Foged et al. \(2022a\)](#) find that long-run earnings of immigrants increase after participation in intensive language training. These programs also have positive spillover effects on the educational success of the participants' children ([Foged et al., 2023](#)). In France, [Lochmann et al. \(2019\)](#) find positive employment effects of mandatory language courses for immigrants and [Heller and Mumma \(2023\)](#) find long-run earnings gains of language training in the US.

Youths High youth unemployment is a major policy concern in many countries which raises the demand for ALMPs supporting young workers. Effective programs should lead young workers towards successful career tracks especially during in the sensitive transition period between education and labor market entry and thereby avoid long run negative outcomes. However, the evidence regarding youth programs is not particularly encouraging. Based on the meta-analysis results, [Card et al. \(2018\)](#) conclude that programs for youths

are less likely to yield positive impacts. The evidence in the recent literature is mixed regarding the success of programs with the aim of speeding up the school work transition.

[Gelber et al. \(2016\)](#) evaluate the New York City Summer Youth Employment Program which offers short-term public sector jobs to young people aged 14–21 during the summer months. Because the program received more applications than there were available slots, those were randomly assigned via a lottery. Combining application data with tax records, the authors show that the program increases earnings in the short run. But the extra labor market experience has no impact on the probability of college enrollment or on longer run earnings outcomes. However, participation in a summer job is successful in keeping young people out of trouble as it significantly reduces incarceration and mortality.

[Le Barbanchon et al. \(2023b\)](#) exploit a similar lottery design in Uruguay where students between age 16 and 20 apply to a work-study program and get randomly assigned to available seats. The program finances short-term, part-time jobs in public sector enterprises which come with the obligation to remain enrolled in school. To foster attachment to education, program rules also prevent firms from keeping participants in the same job after the program ends. Evaluation results show that the program increases employment and earnings of participants in the first years after the end of the subsidized job. It also succeeds in increasing attachment to education, as participants are more likely to remain in school and perform well in terms of their grades. Not surprisingly, the effect on education is mainly driven by participants from poor households.

[Crépon et al. \(2013a\)](#) evaluate a job search assistance program with intensive counseling that starts during job search and continues once a job is found and which is available for recent university graduates who could not find jobs and had been unemployed for at least 6 months. The results reveal low program take-up rates among individuals randomly assigned to program offers and a small positive employment effect in the short run which, however, fades quickly. Taking into account spillovers to the control group who are not offered the program, the net program effect is negative with less jobs generated than in the absence of the program, see also [Section 4.7](#).

ALMPs in developing countries are typically available for youth who suffer from extremely high rates of joblessness. We discuss available evidence in [Section 4.5](#).

4.4 Lesson 3: Program design takes demand side into account

Traditional ALMP policies focus on workers on the supply side of the market and programs are designed to directly increase job finding rates or to facilitate the workers' access to stable and well-paid jobs. Recent policy approaches increasingly take the demand side into account either by involving local employers actively in the design of training programs or by implementing programs that directly address employers.

Training program design involving employers A comprehensive approach in the design of training programs involves the potential employers of training participants. Surveying local labor demand will allow targeting training efforts towards occupations that are in high demand, offer high starting wages and opportunities for career advancement. Alternatively, employer involvement can be fostered by creating subsidized on-the-job training positions in firms. Evidence on successful training programs with employer involvement comes from programs that are organized locally at the community level and in close cooperation with potential employers. In Latin America, the World Bank and the Inter American Development Bank pushed programs based on the “Chilean model” which is focused on integrating disadvantaged workers in the labor market by designing programs that integrate employer needs and activate employers who provide training jobs ([Ibarrarán and Rosas Shady, 2009](#)).

Exploiting the spatial roll-out over time, [Foged et al. \(2022b\)](#) evaluate *industry packages* an innovative integration program for refugees in Denmark. The program targets training efforts in low-wage occupations with high numbers of unfilled vacancies in the local labor market. After a short training course in one of the target occupations, trainees are matched to jobs with local employers. Compared to standard training programs for immigrants, industry packages result in higher employment rates of participants in the first years after immigration. [Dahlberg et al. \(2024\)](#) evaluate a program with a strong employer involvement that was implemented in a randomized control trial in the labor market of Gothenburg in Sweden. The evaluation finds large positive effects doubling employment rates of participants relative to the control group in the first year after the program.

Sectoral employment programs in the US, surveyed by [Katz et al. \(2022\)](#), are also designed in close collaboration with potential employers. The programs offer short courses with occupational and soft-skills training to low educated and disadvantaged workers. Evaluations across four US sites demonstrate large and persistent gains in employment and earnings over 2 to 6 years after program participation. Overall, the program effects of sectoral employment programs are far more promising than the modest effects that are found in evaluations of traditional US training programs.

While the idea of involving employers more strongly in the design of ALMPs is appealing, there are also a number of concerns and open questions. First, programs which are too closely aligned to specific employers raise concerns of spillover effects as employers receive subsidies for training efforts they would have financed themselves. Second, while these programs speed up employment transitions of participants in the short run they might reduce their job flexibility in the longer run, if they mostly train firm specific skills and make it costly for participants to signal their skills to alternative employers ([Hanushek et al., 2017](#)). In addition, employers might be able to extract rents from trainees and pay lower wages ([Naidu and Sojourner, 2020](#)).

Programs addressed at firms Turning the idea of job search assistance for job-seekers around, the French employment service introduced a free recruitment service for small and medium sized firms. The program offered assistance with vacancy posting, candidate selection, and referrals of job seekers to participating firms. It was implemented in a slack labor market in 2015 at the end of the Great Recession. [Algan et al. \(2020\)](#) evaluate the program effects from a large scale RCT where a set of randomly assigned firms got access to the program. Compared to non-treated control firms, the experimental firms posted more vacancies and increased hiring, which led to a positive gain in net employment at the firm level. The additional hires were mostly low skilled workers in permanent jobs. The interpretation of these findings is that in a slack labor market firms reduce demand for low-skilled positions because of high recruitment costs. The program offers an effective screening service to deal with large number of potential applicants for each vacancy and firms value the high quality information from applicant referrals.

Instead of addressing recruitment costs, hiring credits aim at reducing labor costs to increase labor demand. [Cahuc et al. \(2019\)](#) evaluate a temporary hiring credit that was non-anticipated and targeted towards small firms at the onset of the Great Recession in France in 2009. The program reduced employer social security contributions of new hires with wages close to the minimum wage over a relatively short time period. Exploiting quasi-experimental designs the authors show that the policy succeeded in raising labor demand. The hiring credit substantially increased employment growth and total hours worked in eligible firms. Results from a structural model that allows the comparison of alternative policy scenarios highlight that the success of hiring credits requires a careful policy design. The French scenario in 2009 fulfilled a series of criteria that are crucial for a successful hiring credit. Namely, it involved a temporary subsidy that was not anticipated by firms, was only available during a limited period of time of high unemployment, was targeted towards a small set of firms, and implemented in market with rigid wages due to binding minimum wage floors. Given these criteria, the policy paid for itself, as we will discuss in [Section 4.8](#).

[Bertrand and Crépon \(2021\)](#) investigate a firm side policy in South Africa with the aim of understanding why small firms are not growing in a labor market with high unemployment. Their intervention provides a random sample of small firms with free access to a website with information on labor laws governing hiring and firing regulations. They find that the informational intervention substantially increases employment growth in treated firms and provide evidence that limited knowledge of labor laws can significantly constrain hiring in small firms.

4.5 Lesson 4: Internationalization of ALMP use and evaluations

Developing countries are characterized by high rates of non-employment especially among youths and even among highly educated workers. Firms in

these countries are small to medium sized and they are often reluctant to hire workers. Labor frictions on both sides of the market appear to hinder a more efficient allocation of firms and workers. On the supply side, skill shortages, credit constraints for investments in training, high costs of job search and of signaling skills to employers play an important role. On the demand side, inefficient labor regulation, high costs of training workers on the job, and high costs of screening job applicants and assessing worker skills restrict firm growth. These frictions can potentially be addressed by ALMPs.

[Card et al. \(2018\)](#) include studies from 47 countries world-wide in their meta analysis where the majority of studies and estimates are from Europe and North-America. Studies from low-income countries were mostly from Latin America. The vast majority of studies from low-income countries evaluated training programs which were offered to youth. The program effects were estimated over the short to medium run (i.e. the first 2 years after participation) and the share of evaluations based on experimental designs was lower than in typical high-income countries. In line with [Card et al. \(2018\)](#), a survey by [McKenzie \(2017\)](#) is not very enthusiastic about program impacts in evaluation studies of early ALMPs in developing countries. As the programs tended to be costly and had only moderate effects, McKenzie concludes that ALMPs in developing countries are unlikely to be very effective.

Over the last decade the number of studies from developing countries has exploded in line with the general ALMP literature. We see a significant increase in the quality of evaluations, mostly in well-designed experimental settings with a focus on longer run-outcomes and an increase in the variety of different programs that aim at addressing multiple labor market frictions that are specific to low-income countries. Here, we list some examples of promising approaches:

[Alfonsi et al. \(2020\)](#) implement a two-sided experiment in Uganda, a country with a young population and extremely low levels of youth employment. In the experiment workers are randomized into two treatment groups where they are either offered a seat in a vocational training course or the option to receive an apprenticeship training with a firm. On the other side of the market firms are randomly matched to job applicants with or without training or to applicants with a wage subsidy for apprenticeship training. Experimental participants' labor market outcomes are followed over a four year period. The evaluation results show that take-up is high among workers who are assigned to vocational training courses. Compared with the control group their employment and earnings outcomes improve substantially over the long run. Firms, on the other hand, are very reluctant to hire trained or untrained applicants to whom they are matched. The take-up rate is somewhat higher for applicants with a wage subsidy for apprenticeship training. Compared to the control group workers receiving firm training do better especially in the short run. But in the longer run these employment and earnings advantages fade. [Alfonsi et al. \(2020\)](#) set up a job-ladder model to interpret the findings and

evaluate general equilibrium effects. They conclude that in this market vocational training of young workers is more effective than policies offering incentives for firms to train workers. An important determinant of the labor market success of workers receiving vocational training is a skill certificate that is recognized by employers. This allows workers to effectively signal their skills and facilitates career moves of trained workers.⁵⁶

[Bassi and Nansamba \(2022\)](#) study the effects of signaling non-cognitive skills on labor market outcomes of workers who come out of vocational training. They show that certified skills affect the job matches and labor market outcomes of applicants. At the end of their training program workers are assigned to job interviews with firms. In half of the interviews a certificate of worker non-cognitive skills is randomly revealed to both the applicant and the employer. In this population workers volunteer to participate in vocational training and they are thus positively selected in terms of non-cognitive skills. The authors first evaluate the effect of revealing the signal on worker and firm expectations and find that firm revise their expectations upward while worker expectations are unchanged, which is consistent with positive selection. In terms of employment outcomes, the signal does not increase employment probability relative to the control group. But the signal leads to positive assortative matching of workers with high skills and employers with higher skill demand. In line with this sorting, workers' wages are higher if the signal is revealed and they increase in their revealed skills.

[Carneiro et al. \(2020\)](#) study the effect of a wage subsidy program with training component on labor market outcomes of unemployed workers in Macedonia. Macedonia is one of the poorest countries in Europe with extremely high unemployment and low youth employment rates. For the wage subsidy program, the Macedonian employment office collected job applications from workers and vacancy postings from employers. The officials matched applicants and vacancies and randomly invited half of the applicants assigned to each vacancy to a job interview for a subsidized position. The control group did not have access to one of these jobs. Results show that access to an interview increased the employment probability in the treatment group in the short run and the effects only declined moderately over the 3.5 year horizon of the follow up study. Average treatment effect estimates show that individuals who get access to training in a subsidized job due to the interview have a very high employment probability at the end of the observation period. Survey evidence further shows that both work-related as well as non-cognitive skills increased among treated applicants.

[Abebe et al. \(2021\)](#) study two randomized interventions that aim at helping young unemployed workers finding good jobs in Ethiopia. The first intervention

⁵⁶ In a companion paper [Rasul et al. \(2023\)](#) study the effects of the experimental interventions on job search and expectations of young workers. This is one of the few studies explicitly examining how ALMP affects job search.

is a transport subsidy reducing search cost by offering bus tickets to the city center where a large job board is located. The second intervention is a job search assistance workshop which provides certificates from general skill tests and allows applicants to signal their skills. In the short run both interventions increase the probability of finding stable jobs in the formal sector. But the effect only persists for the job search workshop in the longer run. After 4 years, workshop participants have similar employment rates but substantially higher earnings than control group members. The authors explain this result by an increase in match quality between firms and workers. But the interventions do not create additional jobs.

[Muralidharan et al. \(2023\)](#) evaluate the impact of the National Rural Employment Guarantee Scheme (NREGS) in India, the world's largest scale public employment program. For identification they exploit the spatial roll-out of a high quality payment system which strongly increased access to the program and reduced corruption. To estimate the program effect they compare outcomes in households in regions that implemented the high quality payment system early with households in regions which implemented it 2 years later. They control for potential spatial spillovers from neighboring regions that were treated early. Estimation results show that improving the quality of NREGS implementation reduced poverty and increased income among the rural poor. Thereby, income gains stemmed mainly from increases in employment and private market wages. These results are in line with general equilibrium effects at the level of treated regions where the NREGS created an outside option for poor workers. In a monopsonistic labor market this result induces private sector employers to increase wages and thus leads to a more efficient allocation of labor. Further evidence in support of the interpretation based on imperfect labor markets are rising reservation wages and declining farm earnings and land prices, while production increases. Although NREGS covers only a small share of 4 % of rural employment it thus has a large impact on market wages.

4.6 Lesson 5: Advances in labor market design on online search platforms

Over the last two decades when high speed internet access became widely available, online job boards have substantially transformed formal job search and replaced traditional search channels. By now most vacancies are posted on large job boards and most job seekers search online. For research, search platforms provide a wealth of novel data allowing researchers to closely track agents over time. Linking search platform data to employment registers further allows to observe search outcomes, i.e. which jobs and workers get finally selected. These opportunities have fundamentally transformed research of the search and matching process which used to be treated as a black box for a long time. In addition, online job platforms offer novel opportunities for the design of ALMPs that aim at improving the match between job seekers and vacancies.

The idea is that algorithms can partly substitute caseworker tasks and recommend suitable vacancies to job seekers, which can either speed up job search or broaden search by uncovering job opportunities that the job seeker would have otherwise missed. A clear advantage is that algorithms, once implemented, have negligible marginal costs and offer fascinating prospects of overcoming search and matching frictions. However, new opportunities for market design also raise a series of questions which we will discuss below. For a comprehensive summary of the literature see [Kircher \(2022\)](#).

In a seminal paper [Belot et al. \(2019\)](#) follow a group of job seekers searching on the online platform of the Scottish employment office. The system allows job seekers to enter keywords indicating their desired occupations and job locations. After observing job search choices for a few weeks the researchers randomly selected a treatment group of job seekers who received automated recommendations to search in occupations that are related to their prior choices. The occupation recommendations are generated by a prediction algorithm based on the frequency of occupational transitions observed in the labor market and on similarities in skill requirements between occupations. The objective of the recommendation algorithm is to point out relevant job opportunities that had been overlooked. Compared to a control group of searchers for whom the search interface remained unchanged, job seekers receiving automated recommendations started to search more broadly and also applied to vacancies in a wider range of occupations, especially if they had searched narrowly prior to the intervention. Broader search among the formerly narrow job searchers also led to a significant increase in invitations to job interviews. Interestingly, job seekers transferred the information from online recommendations to their job search activities outside the platform. They received more interview invitations overall, not only from vacancies for which they had applied via the platform. The study was run at a small scale with only 300 participants, which limits the power of finding out about the ultimate job search success and whether broader search also leads to faster transitions into employment and higher quality jobs.

In a follow-up study with a slightly larger sample of long-term unemployed job seekers in England, [Belot et al. \(2022a\)](#) implement the automated personalized occupational advice program in a setting where they can observe employment outcomes in administrative data. The English program has similar effects on the breadth of search and on job applications as the Scottish one. But the experiment also reveals positive effects program effects on the probability that long-term unemployed workers find stable jobs and reach a certain earnings limit. These jobs are generated from searching in a broader set of occupations rather than from increased search effort or search in a larger geographical area.

Both studies treat a relatively small part of the labor market with automated personalized occupational advice, which reduces concerns of displacement effects as it unlikely that treated job seekers get jobs that would have otherwise

gone to workers in the comparison group. However, it remains unknown what would happen once the program is scaled up. Evaluating spillover effects of online job recommendations on other job seekers is the objective of the most recent contributions in the literature, which we will discuss in the next section.

[Le Barbanchon et al. \(2023a\)](#) develop a job recommender system based on a machine learning tool that can be rolled out on the full search platform of the Swedish public employment service. This system generates a personalized short list of most relevant vacancies for each job seeker based on past vacancy views of job seekers with similar search preferences. In contrast to automated occupation recommendations which focus on information that might be overlooked the job recommender diffuses information among similar job seekers. The authors show that that recommender generated vacancies increase the geographical and occupational breadth of job opportunities and have a strong focus on vacancies that are less popular by other job seekers. To assess displacement and congestion effects, [Le Barbanchon et al. \(2023a\)](#) use a clustered 2-sided randomization design, where job seekers as well as vacancies are randomly assigned to treatment and control groups. Treated job seekers receive vacancy recommendations and treated vacancies are shown in the recommendations. In addition, spatial variation is generated by randomly assigning a subgroup of local labor markets to a non-treated super control region. This design allows to study very flexibly the effects of the job recommender on job search and job finding outcomes of various subgroups of job seekers as well as hiring outcomes of vacancies. In terms of job search, findings from the country-wide experiment in 2021–2022 confirm the previous literature. Treated job seekers are more likely to follow the recommendations and to apply to recommended vacancies than to non-recommended ones. In addition, employment rates of treated job seekers increase slightly.

An analysis at vacancy-unemployed pair-level, reveals important reallocation effects of the recommender system. In particular, there is little evidence that job seekers in the control group are crowded out of employment in vacancies that were recommended to treated seekers. Neither is the recommender system driving up competition for recommended vacancies among treated job seekers. The finding of small congestion effects in Sweden is in contrast to results by [Altmann et al. \(2022\)](#), which will be discussed in more detail in the next section.

In the spirit of traditional job search assistance programs, the main aim of automatized recommendations for job seekers is overcoming labor market frictions by improving the rate at which job seekers find jobs and at which vacancies get filled. Ideally, the recommendations should also lead to a more efficient allocation of workers to jobs and improve match quality. Advanced machine learning technologies offer ample opportunities for the development of job recommendation systems and further research will be necessary to investigate the variability of recommendations generated by different systems, their alignment with job seekers' goals, and their impacts on labor market outcomes ([Behaghel et al., 2024; Bied et al., 2023](#)).

4.7 Lesson 6: Growing awareness of spillover or displacement effects

For a long time the comparison of mean outcomes in the treatment group and the control group in a controlled environment with randomized assignment was regarded as the gold standard of ALMP program evaluation. However, results can be misleading if there are spillovers from the treatment to the control group. Early concerns of spillover effects were raised in the context of public sector employment programs which risk subsidizing jobs that would also be created in absence of the program. See [Johnson and Tomola \(1977\)](#), who document that the Public Sector Employment service in the US replaced other jobs that would have filled by local governments.

Regarding training programs, most economists think that spillover effects are relatively limited. In an economy with restricted supply of human capital, trained individuals will compete with workers higher up in the job ladder who have more outside options, which in turn limits the risk of displacements ([Katz et al., 2022](#)). However, search assistance programs or automated job referral programs might create large spillovers by privileging the access to jobs for one group at the cost of the others. This is easily seen in the case of referrals to a specific job vacancy. If the treated job seeker gets the job, the vacancy is no longer available to the job seeker in the control group and the recommendation program just re-orders the job queue.

The first study that addressed this problem seriously was by [Crépon et al. \(2013a\)](#). The authors evaluated a job search assistance program for young unemployed university graduates in France which was implemented across multiple local labor markets. The evaluation design is based on a double randomization strategy. In a first step treatment intensities determining the share of treated job seekers, were randomly assigned across local labor markets. In the second steps eligible job seekers were randomly assigned to the JSA program according to the local treatment intensities. This design allows comparisons of treated and control individual within each region, but also across regions. If individuals in the control group in regions with high treatment intensity have systematically different outcomes than control individuals in regions where the program is not implemented this is indicative of spillover effects.

In their evaluation, [Crépon et al. \(2013a\)](#) find evidence of substantial spillover effects which lead to displacement of workers in the control group. While a comparison of treated and control group outcomes within regions points to a positive employment effect of the JSA program on participants, the comparison of non-treated individuals across regions indicates even stronger displacement effects. The authors conclude that overall more jobs were lost than found. This estimate is compatible with a labor market where vacancies do not adjust to the increase in job search activity induced by the JSA program, at least in the short run. The evidence from France indicates that displacement effects are especially strong in weak labor markets with few vacancies, which further supports this model.

[Cheung et al. \(2023\)](#) replicate the French experiment and confirm the importance of spillover effects. They investigate a JSA program in Sweden which intensifies the frequency of caseworker meetings and job search workshops to which newly unemployed workers are assigned. The experiment is rolled out across 72 employment offices. In a random group of active offices half of the eligible job seekers are assigned to programs and in the comparison group of inactive offices no job seeker is assigned. The experimental results show that in the Swedish case, the net effect of program participation on the probability of leaving unemployment in the first three months of unemployment is positive. But this net effect, i.e. the employment gain of treated job seekers compared to those in inactive regions, is about the same magnitude as the displacement effect on non-treated job seekers. An important mechanism driving the displacement effect are vacancy referrals that are shown earlier to program participants who have more frequent meetings with caseworkers. This indicates that the program increased competition for a fixed number of vacancies and created congestion.

[Altmann et al. \(2022\)](#) implement a large randomized control trial among unemployment benefit recipients in Denmark. Danish job seekers are mandated to set up a search profile specifying their target occupations on the centralized job search platform. UI benefit recipients have to visit this website regularly and keep a record of the jobs they apply for. Based on this information the authors implement an information intervention with three different treatments in a double randomization design. In the first stage municipalities were randomly assigned a treatment intensity. In control regions no job seeker received information treatment and in the remaining regions 60 % and 90 % of job seekers were treated, respectively. In the second stage job-seekers in each region were randomly assigned to a *vacancy treatment* informing them about quantities of vacancies in each of the occupations on their profiles, a *occupational recommendation treatment* referring them to suitable alternative occupations similar to [Belot et al. \(2019\)](#), and a *joint treatment* consisting of both components. This strategy creates a continuous measure of treatment intensity at the level of local labor markets, which takes into account that individuals do not only search and work in their own municipality but may commute to neighboring places.

Regarding search strategies the results by [Altmann et al. \(2022\)](#) confirm [Belot et al. \(2019\)](#) and show that job seekers receiving occupational recommendations broadened their search and job applications. The information on open positions per occupation, in contrast, led to a narrower focus on core occupations, and combining both sources of information canceled out changes in job search such that the joint treatment had no effect on job search. In terms of employment and earnings effects there is large heterogeneity across local labor markets with different treatment intensities. In markets with low treatment intensity both the vacancy and the occupation recommendation treatment have positive effects on hours worked and earnings within 12 months of

treatment, while the joint treatment has smaller effects. In municipalities with high treatment intensity, however, the effects on hours and earnings are small and insignificant. This indicates substantial negative spillover effects of the change in search strategies by treated individuals. Comparing outcomes among untreated and treated individuals across regions shows little evidence that job seekers in the control group are affected by spillovers. But workers receiving the information treatment increasingly compete for the same types of vacancies as treatment intensity goes up. This means that treatment creates congestion effects as all search effort is concentrated on a smaller set of vacancies. These results highlight the importance of carefully designing automated recommendations at a large scale.

4.8 Lesson 7: Discussion of cost effectiveness

To evaluate the overall effectiveness of ALMPs it is essential to compare the gains and the cost of the program. The early ALMP evaluation literature was mostly silent about this comparison and with few exceptions, studies did not report program costs. In their meta-analyses [Card et al. \(2010\)](#) and [Card et al. \(2018\)](#) use information on program duration to proxy for program costs in meta-regressions.

The recent literature is more concerned about the overall welfare implications of ALMPs and by now many studies contain detailed cost benefit analyses. [Table 5](#) summarizes the information on program costs and gains from the studies included in our review. When we compare reported program costs across studies it becomes apparent that there is a large variation as is shown in column (2). Program costs range from a few dollars or euros for short-term job search assistance programs to multiple thousands for long-term training and wage subsidy programs. The variation makes it obvious why a careful analysis of the monetary program gains is important to justify high cost programs.

Studies use a variety of different concepts to evaluate program gains which may be explained by the policy context or by data availability, see column (5). Many analyses consider a cumulative measure of gains over several years ranging from 1 to over 20 years, see column (4). The main differences in the benefit concepts is that some studies consider returns to an individual investment comparing the upfront program participation cost to a measure of total individual earnings gains. Other studies consider returns to a public investment and compare participation costs to measures of social returns in terms of tax revenues and foregone benefit payments.

Interestingly, with the only exception of [Crépon et al. \(2013a\)](#) who report a net loss in jobs in regions where the program is implemented, all studies with detailed cost benefit data in [Table 5](#) report that programs either break even or are financially advantageous for participants or for society as a whole. The examples in our comparison demonstrate that not only low cost programs can be run cost effectively, but also significant investments in human capital or

TABLE 5 Cost and benefit analyses in recent ALMP studies.

Study	Currency	Cost per participant	Cumulative gain	Years of sustained gain	Benefit concept
					(5)
Panel A: Training					
Alfonsi et al. (2020)	USD	510	1246	15	NPV of change in steady state individual earnings
Heller and Mumma (2023)	USD	4492	4374	27	Increased tax revenue
Sarvimäki and Hämäläinen (2016)	Euro	15,000	28,000	10	Cumulative gross income plus foregone benefit payments
Schlosser and Shanan (2022)	NIS	1400	3886	1	Income gain and benefit reduction
Katz et al. (2022)	USD	4459	28,662	5	Cumulative earnings gains
			23,135	38,484	Societal net benefit from participant earnings gains
Hyman (2018)	USD	40,362	43,393	10	NPV of change in individual earnings
Panel B: JSA					
Michaelides and Mueser (2020)	USD	201	775	1	Saving in unemployment benefits
McConnell et al. (2021)	USD	692	6630	2.5	Net income gain of participants
			2187		Increase in tax revenue
Naibom et al. (2017)	Euro	903	2003	4	Discounted net government gain
Bobonis et al. (2022)	CAND	4804	5900	20	Real annual earnings gain

Continued

Table 5 Cost and benefit analyses in recent ALMP studies.—Cont'd

Study	Currency	Cost per participant	Cumulative gain	Years of sustained gain	Benefit concept
	(1)	(2)	(3)	(4)	(5)
Abebe et al. (2021)	USD	18.2	299	1	Earnings gain in year 4
Böhme et al. (2022)	Euro	9	31	1	Earnings gain in year 4
Crépon et al. (2013a)	Euro	390	1075	2	Saving in benefits plus gain in taxes and SS contributions
Kasy and Lechner (2023)	Euro	1600	-12		Jobs created in active regions
Panel C: Public Sector Employment Program					
Kasy and Lechner (2023)	Euro	90,000			Cost of 3 years subsidized job
Panel D: Firm Side Intervention					
Algan et al. (2020)	Euro	145	1277	1	Income in minimum wage job
Bertrand and Crépon (2021)	USD	200	20		Cost per job created
Cahuc et al. (2019)	Euro	700	700	1	Saving in unemployment benefits
Cheung et al. (2023)	Euro	99	0.25		Increase in job finding rate across all job seekers in active offices

Notes: This table presents cost benefit calculations from several studies evaluating ALMP policies. Benefit Concept refers to the definition of program gains applied in the study.

labor market attachment can pay off for individuals or society in the longer run, which is an encouraging message for ALMP. Whether the studies shown in Table 5 are positively selected or whether authors choose to only report favorable cost benefit measures is a question which will have to be resolved in a meta-analysis of the full literature.

A unified analysis that allows a comparison across multiple policies is the marginal value of public funds (MVPF) suggested by [Hendren and Sprung-Keyser \(2020\)](#). Computed in the context of ALMPs, the MVPF compares a measure of the willingness to pay—e.g., the cumulative discounted present value of the net of tax earning gain of participants—to a cost measure given by the upfront program cost plus the fiscal externality for the government. For an example, see the calculation of the MVPFs of Job Corps in [Hendren and Sprung-Keyser \(2020\)](#).

4.9 Lesson 8: Wide range of outcome variables

As can be seen from [Tables 6 and 7](#), significant progress has been made in the ALMP evaluation literature regarding the set of outcome variables that are included in recent studies. A wider set of different outcomes give a more comprehensive picture of the program impacts. [Card et al. \(2018\)](#) restricted the meta-analysis to estimates of program impacts on the probability of employment, because this was the most commonly reported outcome measure. While several studies evaluated effects on exit rates from unemployment only, few included earnings as an outcome measure. This has changed dramatically in studies written over the last 10 years. With very few exceptions, all recent studies report program impacts on employment and earnings. In addition, program effects on measures of job match quality such as occupation and firm type are available in some studies. As we can also see from the tables, a substantial share of studies report program effects on outcomes observed over the longer-run of three years or more after program participation. This is not only the case for studies evaluating effects on training programs, which naturally target long-run outcomes. But also studies evaluating JSA programs are increasingly concerned with longer-run outcomes.

Measures of earnings and other outcomes observed in administrative register data may be imperfect summaries of the value of specific jobs. It is therefore desirable to supplement them with more comprehensive measures from survey data. Several studies provide program effects on measures of cognitive and non-cognitive skills ([Katz et al., 2022; Schlosser and Shanan, 2022](#)), information on health, mortality, well-being, life and job satisfaction and measures of social well-being ([Kasy and Lehner, 2023](#)). Another novelty is the elicitation of measures of job search strategies and search effort. Search outcomes are implicitly observed for programs generating automated job search advice on large search platforms. Other studies elicit information from surveys [?].

TABLE 6 Characteristics ALMP studies: training, employment subsidies, firm programs.

Study	Standard ALMP	Country	Outcome variables	Design	Intake period	Intake group	Long run	Cost benefit	Spillover effects
Panel A: Training									
Alfonsi et al. (2020)	Yes	UG	Employment, earnings, skills	RCT	2013	Youth training applicants	Yes	Yes	Yes
Rasul et al. (2023)	Yes	UG	Job search, expectations, call backs	RCT	2013	Youth training applicants	Yes		
Dahlberg et al. (2024)	Yes	SE	Employment, integration, program participation	RCT	2016–20	Newly arrived refugees			
Foged et al. (2022b)	Yes	DK	Employment, earnings	DD	2008–19	Newly arrived refugees			
Foged et al. (2023)	Yes	DK	Education in second generation	RD		Newly arrived refugees			
Foged et al. (2022a)	Yes	DK	Employment, earnings, occupation, education, criminal outcomes	RD	1996–2003	Refugees	Yes		
Heller and Numma (2023)	Yes	US	Earnings, voter registration	RCT	2008–16	Immigrants	Yes	Yes	
Katz et al. (2022)	Yes	US	Earnings, employment in high wage occupations or industries	RCT	2011–13	Youth, low income	Yes	Yes	
Lochmann et al. (2019)	Yes	FR	Labor force participation, employment, household income	RD	2010	Immigrants			
Sarvimäki and Häniläinen (2016)	Yes	FI	Earnings, benefit receipt, employment, occupational quality, training participation	RD	1990–99	Immigrants	Yes	Yes	
Schlosser and Shanan, (2022)	Yes	IL	Employment, benefit recipiency, non-cognitive skills	RCT	2014	Unemployed		Yes	

Panel B: Employment Subsidies						
	Carneiro et al. (2020)	MK	Employment, cognitive and non-cognitive skills	RCT	2005–2008	Program applicants
Gelber et al. (2016)	Yes	US	Employment, earnings, college enrollment, incarceration, mortality	RCT	Applicants age 14–21	Yes
Kasy and Lehner, (2023)	Yes	AT	Employment, worker well- being, health	RCT	Long term unemployed	Yes
Le Barbançon et al. (2023b)	Yes	UY	Employment, earnings, school enrollment, grades, time use, soft skills	RCT	Students aged 16–20	
Muralidharan et al. (2023)	Yes	IN	Earnings, poverty, wages, employer market power	DD	2010–12	Poor households

Panel C: Firm Programs						
	Algarni et al. (2020)	FR	Vacancy posting, hiring, match quality	RCT	2015	Small firms
Bertrand and Crépon (2021)	No	ZA	Firm level employment, hiring	RCT	2013	Firms
Cahuc et al. (2019)	No	FR	Hiring rate, separation rate, employment growth, hours growth, av. wage	DD, IV	2009	Small firms

Notes: This Table lists the studies evaluating ALMPs included in this review. Standard ALMP refers to the definition by [Card et al. \(2018\)](#), Long run refers to program effects reported for three or more years after program participation, Cost Benefit refers to whether the study includes a cost benefit analysis or information about program costs. Spillovers refers to whether the study addresses spillover or displacement effects.

TABLE 7 Characteristics ALMP studies: job search assistance, automated advice.

Study	Standard ALMP	Country	Outcome variables	Design	Intake period	Intake group	Long run	Cost benefit	Spillover effects
Panel D: Job Search Assistance									
Abebe et al. (2021)	Yes	KE	Job finding, employment, earnings, job satisfaction, search effort	RCT	2014	Youth	Yes	Yes	Yes
Bassi and Nansamba (2022)	No	UG	Employment, matching assortativeness, skills, expectations	RCT	2014	Trainees			
Bobonis et al. (2022)	Yes	CA	Employment, earnings, job quality, non-cognitive skills	RCT	1994-95	Single mothers	Yes	Yes	
Böheim et al. (2022)	No	AT	Number of meetings, job offers, other ALMPs, sanctions, job finding, labor market exit, unemployment exit	RCT	2015	Registered unemployed	Yes	Yes	
Cederlöf et al. (2021)	No	SE	Job finding	RCT	2003-10	Registered unemployed			
Cheung et al. (2023)	Yes	SE	Unemployment	RCT	2015	Newly unemployed	Yes	Yes	
Crépon et al. (2013a)	Yes	FR	Job finding	RCT	2007	Youth long term unemployed	Yes	Yes	
Homrichausen and Oberleitner (2024)	No	DE	Employment, earnings	RCT	2018	Registered unemployed			

Maiboom et al. (2017)	Yes	DK	Employment	RCT	2008	Newly unemployed	Yes	Yes
Manoli et al. (2018)	Yes	US	Benefit receipt, employment, earnings, home-ownership, DI receipt	RCT	2009	Benefit claimants	Yes	
McConnell et al. (2021)	Yes	US	Employment, earnings	RCT	2011–13	Job losers, low income adults	Yes	
Michaelides and Mueser (2020)	Yes	US	Benefit receipt, employment, earnings	RCT	2009	Benefit claimants		
Michaelides and Mueser (2023)	Yes	US	Benefit receipt, employment, earnings, UI exit hazard	RCT	2009, 2015	Benefit claimants	Yes	Yes
Schiprowski (2020)	No	CH	Job finding	DD	2010–12	Registered unemployed		
Panel E: Automated Advice								
Altmann et al. (2022)	No	DK	Job search, employment, earnings [§]	RCT	2019	Benefit claimants		Yes
Belot et al. (2022a)	No	UK	Job search, applications, employment, earnings	RCT	2019	Long term unemployed		
Belot et al. (2022b)	No	UK	Job search/applications, interview invitations	RCT	2013–14	Unemployed		
Le Barbançon et al. (2023a)	No	SWF	Job search, applications, employment	RCT	2021–22	Registered unemployed	Yes	

Notes: See notes for Table 6.

4.10 Lesson 9: What are the mechanisms explaining program effects

Why do programs work? Besides reporting program effects, studies are increasingly interested in this question which certainly is of high relevance for program design. Delving into mechanisms allows us to understand why certain groups can benefit from a program while others cannot or why a program works in one labor market and not in another.

A first approach towards understanding mechanisms driving program effects is an analysis of program effect heterogeneity. Especially heterogeneity over the business cycle or by labor market tightness across local labor markets reveals how the program design interacts with labor market conditions. For example, vacancy referrals and recruitment assistance might be more valuable for job seekers and firms in weak labor markets ([Algan et al., 2020](#); [Altmann et al., 2022](#); [Belot et al., 2019](#); [Cheung et al., 2023](#); [Crépon et al., 2013a](#)) and training might be more effective if it is targeted towards local demand ([Foged et al., 2022b](#); [Katz et al., 2022](#)).

Specific program features can also be relevant for the program success. Several studies highlight that certificates issued by ALMP programs allow workers to signal their skills to employers, which can be crucial for job finding success ([Alfonsi et al., 2020](#); [Bassi and Nansamba, 2022](#); [Katz et al., 2022](#)). Another feature that has received attention in the literature is the timing of counseling services, whether they should be available early in the unemployment spell or later ([Maibom et al., 2017](#)), whether it is beneficial to continue career counseling services once a job is found ([Bobonis et al., 2022](#); [Katz et al., 2022](#)) or whether counseling should already start on the job prior to the transition into unemployment ([Homrighausen and Oberfichtner, 2024](#)).

Another determinant of program success may be non-cognitive skills. They are hard to observe and even harder to train. But available evidence indicates that especially disadvantaged workers with long absences from the labor market benefit from programs that promote their non-cognitive skills ([Bobonis et al., 2022](#); [Schlosser and Shanahan, 2022](#)).

Several recent ALMP evaluation studies successfully introduce structural models that can be used to understand mechanisms driving evaluation outcomes ([Cheung et al., 2023](#); [Crépon et al., 2013a](#)) to simulate alternative policy scenarios ([Cahuc et al., 2019](#)) or to evaluate potential general equilibrium effects ([Alfonsi et al., 2020](#); [Muralidharan et al., 2023](#)).

4.11 Lesson 10: Novel identification strategies

The overview of identification strategies in [Tables 6 and 7](#) shows that RCT has become the standard evaluation methodology of ALMPs, at least among studies written over the last 10 years and reviewed in this chapter. Moreover, important methodological innovations were made in the design of the RCTs. In this chapter, we have discussed experimental designs that randomize at the worker-firm match level ([Abebe et al., 2021](#); [Carneiro et al., 2020](#)) and designs

that randomize both sides of the market (Alfonsi et al., 2020; Le Barbanchon et al., 2023a) in Section 4.5, as well as designs based on double clustered randomization that allows uncovering spillover effects (Altmann et al., 2022; Cheung et al., 2023; Crépon et al., 2013b; Le Barbanchon et al., 2023a) in Section 4.7.

Progress was also made with adaptive targeted treatment assignments in field experiments. Caria et al. (2024) evaluate three different ALMP programs with the aim of integrating Syrian refugees in the Jordanian labor market. They stratify the intake population in 16 strata in order to find out which program works best for which group of participants. To maximizing the precision of treatment effect estimation as well as the welfare of experimental participants an adaptive targeted treatment assignment algorithm is chosen. Outcomes are measured over time such that employment and earnings trajectories are observed for each participant and newly eligible individuals can be included in the randomization over time. Treatment shares of newly entering participants are determined by observed outcomes of earlier cohorts based a Tempered Thompson Algorithm with hierarchical Bayesian updating. Results confirm that while the programs only slightly increase overall employment, the adaptive assignment strategy can improve employment outcomes among certain groups of participants.

Innovations in non-experimental identification involve the transfer of the judge leniency design to variation in caseworker propensity of assigning unemployed workers to ALMP programs. Humlum et al. (2023) show that a large group of benefit recipients in Sweden are as good as randomly assigned to caseworkers based on their day of birth within local employment offices. After isolating quasi-randomly assigned groups the study exploits variation in the propensity of program assignments among caseworkers as an instrumental variable to identify the effect of program participation on employment and earnings outcomes. This identification strategy finds positive employment and earnings effects of training participation in the medium run, which are significantly larger than the effects of wage subsidy programs. Interestingly, OLS estimates of the program effects show the opposite sign, which indicates large negative selection bias as individuals with negative unobserved characteristics are assigned to training programs. Positive program effects are due to individuals who complete training and switch occupations especially into sectors with low levels of offshoring.

An earlier study using a similar IV design to evaluate the earnings effect of Trade Adjustment Assistance programs in the US is Hyman (2018). In this program training courses are available for workers who were displaced by firms suffering from increased import competition or outsourcing. Displaced workers had to apply for program access and the applications were reviewed by case investigators who assess the employer's exposure to import competition. Applications were assigned to case investigators based on caseloads which

generates quasi-random variation. The IV strategy thus exploits variation in approval shares across caseworkers. The estimates indicate large cumulative gains in long-term earnings from retraining displaced workers.⁵⁷

Acknowledgments

We thank the editors, Christian Dustmann and Thomas Lemieux, for giving us the opportunity to contribute to the Handbook of Labor Economics. We thank the participants to the Handbook Workshop at RF Berlin for their helpful comments. We thank Francesco Armillei, David Card, Giulia Giupponi, Peter Ganong, Nathan Hendren, Camille Landais, Amelie Schiprowski, and Pauline Leung for providing feedback on early version of our chapter. Savannah Kochinke provided excellent research assistance.

Appendix A Supporting information

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.addma.2020.101681>.

References

- Abebe, G., Caria, A.S., Fafchamps, M., Falco, P., Franklin, S., Quinn, S., 2021. Anonymity or distance? Job search and labour market exclusion in a growing African city. *Rev. Econ. Stud.* 88, 1279–1310.
- Ahamer, A., Packham, A., 2023. Effects of unemployment insurance duration on mental and physical health. *J. Public Econ.* 226, 104996.
- Albanese, A., Picchio, M., Ghirelli, C., 2020. Timed to say goodbye: does unemployment benefit eligibility affect worker layoffs? *Labour Econ.* 65, 101846.
- Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., et al., 2020. Tackling youth unemployment: evidence from a labor market experiment in Uganda. *Econometrica* 88, 2369–2414.
- Algan, Y., Crépon, B., Glover, D., 2020. Are active labor market policies directed at firms effective? Evidence from a randomized evaluation with local employment agencies. J-PAL working paper.
- Allen, E.J., Dechow, P.M., Pope, D.G., Wu, G., 2017. Reference-dependent preferences: evidence from marathon runners. *Manag. Sci.* 63, 1657–1672 (publisher: INFORMS).
- Altmann, S., Glenny, A.M., Mahlstedt, R., Sebald, A., 2022. The direct and indirect effects of online job search advice. *IZA Discussion Pap.*
- Alvarez, F., Borovičková, K., Shimer, R., 2023. Decomposing duration dependence in a stopping time model. *Rev. Econ. Stud.* rdad109.
- Anderson, P.M., Meyer, B.D., 1993. Unemployment insurance in the united states: layoff incentives and cross subsidies. *J. Labor. Econ.* 11, S70–S95 (publisher: [University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago]).
- Anderson, P.M., Meyer, B.D., 1997. Unemployment insurance takeup rates and the after-tax value of benefits. *Q. J. Econ.* 112, 913–937 (publisher: Oxford University Press).

⁵⁷ Part of the earnings gains might be due to UI benefit extensions available to program participants. The decomposition of the overall effects into contributions from benefits versus training is the subject of ongoing work.

- Arendt, J.N., Dustmann, C., Ku, H., 2022. Refugee migration and the labour market: lessons from 40 years of post-arrival policies in Denmark. *Oxf. Rev. Econ. Policy* 38, 531–556.
- Arni, P., Schiprowski, A., 2019. Job search requirements, effort provision and labor market outcomes. *J. Public Econ.* 169, 65–88.
- Ashenfelter, O., 1987. The case for evaluating training programs with randomized trials. *Econ. Educ. Rev.* 6, 333–338.
- Autor, D.H., Maestas, N., Mullen, K.J., Strand, A., 2015. Does delay cause decay? The effect of administrative decision time on the labor force participation and earnings of disability applicants.
- Baguelin, O., Remillon, D., 2014. Unemployment insurance and management of the older work-force in a dual labor market: evidence from France. *Labour Econ.* 30, 245–264.
- Baily, M.N., 1978. Some aspects of optimal unemployment insurance. *J. Public Econ.* 10, 379–402.
- Baker, M., Rea Jr., S.A., 1998. Employment spells and unemployment insurance eligibility requirements. *Rev. Econ. Stat.* 80, 80–94.
- Baker, S.R., Fradkin, A., 2017. The impact of unemployment insurance on job search: evidence from Google search data. *Rev. Econ. Stat.* 99, 756–768 (publisher: The MIT Press).
- Bassi, V., Nansamba, A., 2022. Screening and signalling non-cognitive skills: experimental evidence from Uganda. *Econ. J.* 132, 471–511.
- Behaghel, L., Dromundo, S., Gurgand, M., Hazard, Y., Zuber, T., 2024. The potential of recommender systems for directing job search: a large-scale experiment. *SSRN Electron. J.*
- Belot, M., Kircher, P., Muller, P., 2019. Providing advice to jobseekers at low cost: an experimental study on online advice. *Rev. Econ. Stud.* 86, 1411–1447.
- Belot, M., Kircher, P., Muller, P., 2022a. Do the long-term unemployed benefit from automated occupational advice during online job search? *SSRN Scholarly Paper* 4178928, Rochester, NY.
- Belot, M., Kircher, P., Muller, P., 2022b. How wage announcements affect job search—a field experiment. *Am. Econ. J.: Macroecon.* 14, 1–67.
- Bertrand, M., Crépon, B., 2021. Teaching labor laws: evidence from a randomized control trial in South Africa. *Am. Econ. J.: Appl. Econ.* 13, 125–149.
- Bied, G., Caillou, P., Crépon, B., Gaillac, C., Pérennes, E., Sebag, M., 2023. Designing labor market recommender systems: how to improve human based search.
- Bobonis, G.J., Bonikowska, A., Oreopoulos, P., Riddell, W.C., Ryan, S.P., 2022. A helping hand goes a long way: long-term effects of counselling and support to welfare program participants. *NBER Working Paper* 30405, National Bureau of Economic Research.
- Böheim, R., Eppel, R., Mahringer, H., 2022. More caseworkers shorten unemployment durations and save costs. Results from a Field Experiment in an Austrian Public Employment Office. *WIFO Working Papers*.
- Boone, C., Dube, A., Goodman, L., Kaplan, E., 2021. Unemployment insurance generosity and aggregate employment. *Am. Econ. J.: Econ. Policy* 13, 58–99.
- Boone, J., van Ours, J.C., 2012. Why is there a spike in the job finding rate at benefit exhaustion? *De. Economist* 160, 413–438.
- Brell, C., Dustmann, C., Preston, I., 2020. The labor market integration of refugee migrants in high-income countries. *J. Econ. Perspect.* 34, 94–121.
- Britto, D.G.C., 2022. The employment effects of lump-sum and contingent job insurance policies: evidence from Brazil. *Rev. Econ. Stat.* 104, 465–482.
- Britto, D.G.C., Pinotti, P., Sampaio, B., 2022. The effect of job loss and unemployment insurance on crime in Brazil. *Econometrica* 90, 1393–1423. (*_eprint_*: <https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18984>).

- Brown, E., Kaufold, H., 1988. Human capital accumulation and the optimal level of unemployment insurance provision. *J. Labor. Econ.* 6, 493–514 (publisher: [University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago]).
- Brown, M., Flinn, C.J., Schotter, A., 2011. Real-time search in the laboratory and the market. *Am. Econ. Rev.* 101, 948–974.
- Browning, M., Crossley, T.F., 2001. The life-cycle model of consumption and saving. *J. Econ. Perspect.* 15, 3–22.
- Cahuc, P., Carcillo, S., Le Barbanchon, T., 2019. The effectiveness of hiring credits. *Rev. Econ. Stud.* 86, 593–626.
- Caliendo, M., Cobb-Clark, D.A., Uhlendorff, A., 2015. Locus of control and job search strategies. *Rev. Econ. Stat.* 97, 88–103.
- Card, D., Chetty, R., Weber, A., 2007. Cash-on-hand and competing models of intertemporal behavior: new evidence from the labor market. *Q. J. Econ.* 122, 1511–1560.
- Card, D., Johnston, A., Leung, P., Mas, A., Pei, Z., 2015a. The effect of unemployment benefits on the duration of unemployment insurance receipt: new evidence from a regression Kink Design in Missouri, 2003–2013. *Am. Econ. Rev.* 105, 126–130.
- Card, D., Kluev, J., Weber, A., 2010. Active labour market policy evaluations: a meta-analysis. *Econ. J.* 120, F452–F477.
- Card, D., Kluev, J., Weber, A., 2018. What works? A meta analysis of recent active labor market program evaluations. *J. Eur. Econ. Assoc.* 16, 894–931.
- Card, D., Lee, D.S., Pei, Z., Weber, A., 2015b. Inference on causal effects in a generalized regression kink design. *Econometrica* 83, 2453–2483 (publisher: [Wiley, The Econometric Society]).
- Card, D., Levine, P.B., 2000. Extended benefits and the duration of UI spells: evidence from the New Jersey extended benefit program. *J. Public Econ.* 78, 107–138.
- Caria, A.S., Gordon, G., Kasy, M., Quinn, S., Shami, S.O., Teytelboym, A., 2024. An adaptive targeted field experiment: job search assistance for refugees in Jordan. *J. Eur. Econ. Assoc.* 22, 781–836.
- Carneiro, P., Armand, A., Tagliati, F., Xia, Y., 2020. Can subsidized employment tackle long-term unemployment? Experimental evidence from North Macedonia. CEPR Discussion Paper 15192, C.E.P.R. Discussion Papers.
- Cederlöf, J., Söderström, M., Vikström, J., 2021. What makes a good caseworker? Working Paper 2021:9, Working Paper.
- Centeno, M., Novo, A.A., 2009. Reemployment wages and UI liquidity effect: a regression discontinuity approach. *Portuguese Econ. J.* 8, 45–52.
- Chetty, R., 2006. A general formula for the optimal level of social insurance. *J. Public Econ.* 90, 1879–1901.
- Chetty, R., 2008. Moral hazard versus liquidity and optimal unemployment insurance. *J. Political Econ.* 116, 173–234 (publisher: The University of Chicago Press).
- Chetty, R., Looney, A., 2006. Consumption smoothing and the welfare consequences of social insurance in developing economies. *J. Public Econ.* 90, 2351–2356.
- Cheung, M., Egebark, J., Forslund, A., Laun, L., Rödin, M., Vikström, J., 2023. Does job search assistance reduce unemployment? Evidence on displacement effects and mechanisms. *J. Labor. Econ.*, 726384.
- Chodorow-Reich, G., Coglianese, J., Karabarbounis, L., 2019. The macro effects of unemployment benefit extensions: a measurement error approach. *Q. J. Econ.* 134, 227–279.
- Chou, C., Shi, R., 2021. What time use surveys can (and cannot) tell us about labor supply. *J. Appl. Econ.* 36, 917–937.

- Christofides, L.N., McKenna, C.J., 1996. Unemployment insurance and job duration in Canada. *J. Labor. Econ.* 14, 286–312 (publisher: [University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago]).
- Cochrane, J.H., 1991. A simple test of consumption insurance. *J. Political Econ.* 99, 957–976 (publisher: University of Chicago Press).
- Cohen, J.P., Ganong, P., 2024. Disemployment effects of unemployment insurance: a meta-analysis.
- Cohen, J.P., Johnston, A.C., Lindner, A.S., 2023. Skill depreciation during unemployment: evidence from panel data.
- Coles, M., Petrongolo, B., 2008. A test between stock-flow matching and the random matching function approach. *Int. Econ. Rev.* 49, 1113–1141. (*_eprint:*). (<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2354.2008.00508.x>).
- Cooper, M., Kuhn, P., 2020. Behavioral job search. *Handbook of Labor, Human Resources and Population Economics*. Springer International Publishing, Cham, pp. 1–22.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., Zamora, P., 2013a. Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Q. J. Econ.* 128, 531–580.
- Crépon, B., van den Berg, G.J., 2016. Active labor market policies. *Annu. Rev. Econ.* 8, 521–546.
- Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., Zamora, P., 2013b. Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Q. J. Econ.* 128, 531–580.
- Dahlberg, M., Egebark, J., Vikman, U., Özcan, G., 2024. Labor market integration of refugees: RCT evidence from an early intervention program in Sweden. *J. Econ. Behav. Organ.* 217, 614–630.
- Davis, S.J., 2011. Comment and discussion of ‘job search, emotional well-being, and job finding in a period of mass unemployment: evidence from high frequency longitudinal data’. *Brook. Pap. Econ. Act.* 58–70.
- DellaVigna, S., Heining, J., Schmieder, J.F., Trenkle, S., 2022. Evidence on job search models from a survey of unemployed workers in Germany. *Q. J. Econ.* 137, 1181–1232.
- DellaVigna, S., Lindner, A., Reizer, B., Schmieder, J.F., 2017. Reference-dependent job search: evidence from Hungary. *Q. J. Econ.* 132, 1969–2018.
- DellaVigna, S., Paserman, M., 2005. Job search and impatience. *J. Labor. Econ.* 23, 527–588 (publisher: The University of Chicago Press).
- Deschacht, N., Vansteenkiste, S., 2021. The effect of unemployment duration on reservation wages: evidence from Belgium. *Labour Econ.* 71, 102010.
- Devine, T.J., 1991. *Empirical Labor Economics the Search Approach*. Oxford University Press, New York.
- Devine, T.J., Kiefer, N.M., 1993. The empirical status of job search theory. *Labour Econ.* 1, 3–24.
- Edin, P.-A., Gustavsson, M., 2008. Time out of work and skill depreciation. *ILR Rev.* 61, 163–180 (publisher: SAGE Publications Inc).
- Eriksson, S., Rooth, D.-O., 2014. Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *Am. Econ. Rev.* 104, 1014–1039.
- Faberman, R.J., Kudlyak, M., 2019. The intensity of job search and search duration. *Am. Econ. J.: Macroecon.* 11, 327–357 (publisher: American Economic Association).
- Fallick, B., Haltiwanger, J.C., McEntarfer, E., Staiger, M., 2021. Job displacement and job mobility: the role of joblessness.
- Farber, H.S., Herbst, C.M., Silverman, D., von Wachter, T., 2019. Whom do employers want? The role of recent employment and unemployment status and age. *J. Labor. Econ.* 37, 323–349 (publisher: The University of Chicago Press).

- Feldstein, M., 1976. Temporary layoffs in the theory of unemployment. *J. Political Econ.* 84, 937–957 (publisher: University of Chicago Press).
- Feldstein, M., Poterba, J., 1984. Unemployment insurance and reservation wages. *J. Public. Econ.* 23, 141–167.
- Field, E., Garlick, R., Subramanian, N., Vyborny, K., 2023. Why don't jobseekers search more? Barriers and returns to search on a job matching platform*. Institution: American Economic Association.
- Flemming, J.S., 1978. Aspects of optimal unemployment insurance: search, leisure, savings and capital market imperfections. *J. Public Econ.* 10, 403–425.
- Flinn, C., Heckman, J., 1982. New methods for analyzing structural models of labor force dynamics. *J. Econ.* 18, 115–168.
- Fluchtmann, J., Glenny, A.M., Harmon, N., Maibom, J., 2023. Unemployed job search across people and over time: evidence from applied-for jobs. *Journal of Labor Economics.* The University of Chicago Press.
- Foged, M., Hasager, L., Peri, G., 2024. Comparing the effects of policies for the labor market integration of refugees. *J. Labor. Econ.* 42, S335–S377.
- Foged, M., Hasager, L., Peri, G., Arendt, J.N., Bolvig, I., 2022a. Language training and refugees' integration. *Rev. Econ. Stat.* 1, 41.
- Foged, M., Hasager, L., Peri, G., Arendt, J.N., Bolvig, I., 2023. Intergenerational spillover effects of language training for refugees. *J. Public Econ.* 220, 104840.
- Foged, M., Kreuder, J., Peri, G., 2022b. Integrating refugees by addressing labor shortages? A policy evaluation. NBER Working Paper 29781, National Bureau of Economic Research.
- Fujita, S., Moscarini, G., 2017. Recall and unemployment. *Am. Econ. Rev.* 107, 3875–3916.
- Ganong, P., Greig, F., Noel, P., Sullivan, D., Vavra, J., 2022. Spending and job-finding impacts of expanded unemployment benefits: evidence from administrative micro data. Tech. Rep. w30315, National Bureau of Economic Research, Cambridge, MA.
- Ganong, P., Noel, P., 2019. Consumer spending during unemployment: positive and normative implications. *Am. Econ. Rev.* 109, 2383–2424.
- Gelber, A., Isen, A., Kessler, J.B., 2016. The effects of youth employment: evidence from New York City lotteries. *Q. J. Econ.* 131, 423–460.
- Gerard, F., Gonzaga, G., 2021. Informal labor and the efficiency cost of social programs: evidence from unemployment insurance in Brazil. *Am. Econ. J.: Econ. Policy* 13, 167–206.
- Gerard, F., Gonzaga, G., Naritomi, J., 2024. Job displacement insurance in developing countries. In: Hanna, R., Olken, B. (Eds.), *The Handbook on Social Protection: Local Solutions for Global Poverty*.
- Gerard, F., Naritomi, J., 2021. Job displacement insurance and (the lack of) consumption-smoothing. *Am. Econ. Rev.* 111, 899–942.
- Green, D.A., Riddell, W.C., 1997. Qualifying for unemployment insurance: an empirical analysis. *Econ. J.* 107, 67–84.
- Green, D.A., Sargent, T.C., 1998. Unemployment insurance and job durations: seasonal and non-seasonal jobs. *Can. J. Econ. / Rev. Canadienne d'Economique* 31, 247–278 (publisher: [Wiley, Canadian Economics Association]).
- Gronau, R., 1971. Information and frictional unemployment. *Am. Econ. Rev.* 61, 290–301 (publisher: American Economic Association).
- Gruber, J., 1997. The consumption smoothing benefits of unemployment insurance. *Am. Econ. Rev.* 87, 192–205 (publisher: American Economic Association).
- Gudgeon, M., Guzman, P., Schmieder, J.F., Trenkle, S., Ye, H., 2023. When institutions interact: how the effects of unemployment insurance are shaped by retirement policies.

- Hagedorn, M., Karahan, F., Manovskii, I., Mitman, K., 2013. Unemployment benefits and unemployment in the great recession: the role of macro effects.
- Hanushek, E.A., Schwerdt, G., Woessmann, L., Zhang, L., 2017. General education, vocational education, and labor-market outcomes over the lifecycle. *J. Hum. Resour.* 52, 48–87.
- Hartley, G.R., van Ours, J.C., Vodopivec, M., 2011. Incentive effects of unemployment insurance savings accounts: evidence from Chile. *Labour Econ.* 18, 798–809.
- Hartung, B., Jung, P., Kuhn, M., 2024. Unemployment insurance reforms and labor market dynamics.
- Heller, B.H., Mumma, K.S., 2023. Immigrant integration in the United States: the role of adult English language training. *Am. Econ. J.: Econ. Policy* 15, 407–437.
- Hendren, N., 2017. Knowledge of future job loss and implications for unemployment insurance. *Am. Econ. Rev.* 107, 1778–1823.
- Hendren, N., Sprung-Keyser, B., 2020. A unified welfare analysis of government policies. *Q. J. Econ.* 135, 1209–1318.
- Hernandez Martinez, V., Liu, K., Grice, R., 2023. Estimating duration dependence on re-employment wages when reservation wages are binding.
- Holmlund, B., Lundborg, P., 1999. Wage bargaining, union membership, and the organization of unemployment insurance. *Labour Econ.* 6, 397–415.
- Homrichausen, P., Oberfichtner, M., 2024. Do caseworker meetings prevent unemployment? Evidence from a field experiment. *SSRN Electron. J.*
- Hopenhayn, H.A., Nicolini, J.P., 1997. Optimal unemployment insurance. *J. Political Econ.* 105, 412–438 (publisher: The University of Chicago Press).
- Huang, P.-C., Yang, T.-T., 2021. The welfare effects of extending unemployment benefits: evidence from re-employment and unemployment transfers. *J. Public Econ.* 202, 104500.
- Humlum, A., Munch, J.R., Rasmussen, M., 2023. What works for the unemployed? Evidence from quasi-random caseworker assignments. *IZA Discussion Paper* 16033, Institute of Labor Economics (IZA).
- Hyman, B.G., 2018. Can displaced labor be retrained? Evidence from quasi-random assignment to trade adjustment assistance. *Working Pap.*
- Ibarrarán, P., Rosas Shady, D., 2009. Evaluating the impact of job training programmes in Latin America: evidence from IDB funded operations. *J. Dev. Effectiveness* 1, 195–216.
- Inderbitzin, L., Staubli, S., Zweimüller, J., 2016. Extended unemployment benefits and early retirement: program complementarity and program substitution. *Am. Econ. J.: Econ. Policy* 8, 253–288.
- Jessen, J., Jessen, R., Galecka-Burdziak, E., Góra, M., Kluge, J., 2023. The micro and macro effects of changes in the potential benefit duration. *SSRN Electron. J.*
- Johnson, G.E., Tomola, J.D., 1977. The fiscal substitution effect of alternative approaches to public service employment policy. *J. Hum. Resour.* 12, 3–26.
- Johnston, A.C., Mas, A., 2018. Potential unemployment insurance duration and labor supply: the individual and market-level response to a benefit cut. *J. Political Econ.* 126, 2480–2522 (publisher: The University of Chicago Press).
- Jurajda, S., 2002. Estimating the effect of unemployment insurance compensation on the labor market histories of displaced workers. *J. Econom.* 108, 227–252.
- Jäger, S., Schoefer, B., Young, S., Zweimüller, J., 2020. Wages and the value of nonemployment. *Q. J. Econ.* 135, 1905–1963.
- Jäger, S., Schoefer, B., Zweimüller, J., 2023. Marginal jobs and job surplus: a test of the efficiency of separations. *Rev. Economic Stud.* 90, 1265–1303.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica ((pre-1986))* 47, 263–291 num Pages: 29 Place: Evanston, United Kingdom Publisher: Blackwell Publishing Ltd.

- Kasy, M., Lehner, L., 2023. Employing the unemployed of marienthal: evaluation of a guaranteed job program. SSRN Scholarly Paper 4431385, Rochester, NY.
- Katz, L.F., Meyer, B.D., 1990. Unemployment insurance, recall expectations, and unemployment outcomes. *Q. J. Econ.* 105, 973–1002.
- Katz, L.F., Roth, J., Hendra, R., Schaberg, K., 2022. Why do sectoral employment programs work? Lessons from work advance. *J. Labor. Econ.* 40, S249–S291.
- Khoury, L., 2023. Unemployment benefits and redundancies: incidence and timing effects. *J. Public Econ.* 226, 104984.
- Kircher, P., 2022. Schumpeter Lecture 2022: Job Search in the 21St Century. *J. Eur. Econ. Assoc.* 20, 2317–2352.
- Koenig, F., Manning, A., Petrongolo, B., 2016. Reservation wages and the wage flexibility puzzle.
- Kolsrud, J., Landais, C., Nilsson, P., Spinnewijn, J., 2018. The optimal timing of unemployment benefits: theory and evidence from Sweden. *Am. Econ. Rev.* 108, 985–1033.
- Kroft, K., Lange, F., Notowidigdo, M.J., 2013. Duration dependence and labor market conditions: evidence from a field experiment*. *Q. J. Econ.* 128, 1123–1167.
- Kroft, K., Lange, F., Notowidigdo, M.J., Katz, L.F., 2016. Long-term unemployment and the great recession: the role of composition, duration dependence, and nonparticipation. *J. Labor. Econ.* 34, S7–S54 (publisher: The University of Chicago Press).
- Kroft, K., Notowidigdo, M.J., 2016. Should unemployment insurance vary with the unemployment rate? Theory and evidence. *Rev. Econ. Stud.* 83, 1092–1124.
- Krueger, A.B., Cramer, J., Cho, D., 2014. Are the long-term unemployed on the margins of the labor market? *Brook. Pap. Econ. Act.* 2014, 229–299 (publisher: Johns Hopkins University Press).
- Krueger, A.B., Meyer, B.D., 2002. Chapter 33 Labor supply effects of social insurance. *Handbook of Public Economics* 4. Elsevier, pp. 2327–2392.
- Krueger, A.B., Mueller, A.I., 2010. Job search and unemployment insurance: new evidence from time use data. *J. Public Econ.* 94, 298–307.
- Krueger, A.B., Mueller, A.I., 2011. Job search, emotional well-being, and job finding in a period of mass unemployment: evidence from high frequency longitudinal data. *Brookings Papers on Economic Activity*. Brookings Institution Press, pp. 1–81.
- Krueger, A.B., Mueller, A.I., 2012. The lot of the unemployed: a time use perspective. *J. Eur. Econ. Assoc.* 10, 765–794 (publisher: Oxford University Press).
- Krueger, A.B., Mueller, A.I., 2016. A contribution to the empirics of reservation wages. *Am. Econ. J.: Econ. Policy* 8, 142–179.
- Kyyrä, T., Pesola, H., 2020. Long-term effects of extended unemployment benefits for older workers. *Labour Econ.* 62, 101777.
- Kőszegi, B., Loewenstein, G., Murooka, T., 2022. Fragile self-esteem. *Rev. Econ. Stud.* 89, 2026–2060.
- Kőszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. *Q. J. Econ.* 121, 1133–1165.
- Laibson, D., 1997. Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112, 443–478.
- Lalive, R., 2007. Unemployment benefits, unemployment duration, and post-unemployment jobs: a regression discontinuity approach. *Am. Econ. Rev.* 97, 108–112.
- Lalive, R., 2008. How do extended benefits affect unemployment duration? A regression discontinuity approach. *J. Econ.* 142, 785–806.
- Lalive, R., Landais, C., Zweimüller, J., 2015. Market externalities of large unemployment insurance extension programs. *Am. Econ. Rev.* 105, 3564–3596 (publisher: American Economic Association).

- Landais, C., 2015. Assessing the welfare effects of unemployment benefits using the regression kink design. *Am. Econ. J.: Econ. Policy* 7, 243–278.
- Landais, C., Michaillat, P., Saez, E., 2018a. A macroeconomic approach to optimal unemployment insurance: theory. *Am. Econ. J.: Econ. Policy* 10, 152–181.
- Landais, C., Michaillat, P., Saez, E., 2018b. A macroeconomics approach to optimal unemployment insurance: applications. *Am. Econ. J.: Econ. Policy* 10, 182–216.
- Landais, C., Spinnewijn, J., 2021. The value of unemployment insurance. *Rev. Econ. Stud.* 88, 3041–3085.
- Le Barbanchon, T., 2016. The effect of the potential duration of unemployment benefits on unemployment exits to work and match quality in France. *Labour Econ.* 42, 16–29.
- Le Barbanchon, T., 2020. Taxes today, benefits tomorrow.
- Le Barbanchon, T., Hensvik, L., Rathelot, R., 2023a. How can AI improve search and matching? Evidence from 59 million personalized job recommendations. SSRN Scholarly Paper 4604814, Rochester, NY.
- Le Barbanchon, T., Rathelot, R., Roulet, A., 2019. Unemployment insurance and reservation wages: evidence from administrative data. *J. Public Econ.* 171, 1–17.
- Le Barbanchon, T., Rathelot, R., Roulet, A., 2021. Gender differences in job search: trading off commute against wage. *Q. J. Econ.* 136, 381–426.
- Le Barbanchon, T., Ubfal, D., Araya, F., 2023b. The effects of working while in school: evidence from employment lotteries. *Am. Econ. J.: Appl. Econ.* 15, 383–410.
- Lee, D.S., Leung, P., O'Leary, C.J., Pei, Z., Quach, S., 2021. Are sufficient statistics necessary? Nonparametric measurement of deadweight loss from unemployment insurance. *J. Labor. Econ.* 39, S455–S506 (publisher: The University of Chicago Press).
- Lentz, R., 2009. Optimal unemployment insurance in an estimated job search model with savings. *Rev. Econ. Dyn.* 12, 37–57.
- Leung, P., O'Leary, C., 2020. Unemployment insurance and means-tested program interactions: evidence from administrative data. *Am. Economic Journal: Economic Policy* 12, 159–192.
- Lichter, A., Schiprowski, A., 2021. Benefit duration, job search behavior and re-employment. *J. Public Econ.* 193, 104326.
- Liepmann, H., Pignatti, C., 2024. Welfare effects of unemployment benefits when informality is high. *J. Public Econ.* 229, 105032.
- Light, A., Omori, Y., 2004. Unemployment insurance and job quits. *J. Labor. Econ.* 22, 159–188 (publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago]).
- Lindner, A., Reizer, B., 2020. Front-loading the unemployment benefit: an empirical assessment. *Am. Econ. J.: Appl. Econ.* 12, 140–174.
- Lindner, S., 2016. How do unemployment insurance benefits affect the decision to apply for social security disability insurance? *J. Hum. Resour.* 51, 62–94 (publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System]).
- Lippman, S.A., McCall, J.J., 1976. Job search in a dynamic economy. *J. Econ. Theory* 12, 365–390.
- Lochmann, A., Rapoport, H., Speciale, B., 2019. The effect of language training on immigrants' economic integration: empirical evidence from France. *Eur. Economic Rev.* 113, 265–296.
- Lopes, M.C., 2022. A review on the elasticity of unemployment duration to the potential duration of unemployment benefits. *J. Econ. Surv.* 36, 1212–1224. (_eprint). (<https://onlinelibrary.wiley.com/doi/pdf/10.1111/joes.12479>).
- Lusher, L., Schnorr, G.C., Taylor, R.L.C., 2022. Unemployment insurance as a worker indiscipline device? Evidence from scanner data. *Am. Econ. J.: Appl. Econ.* 14, 285–319.

- Machin, S., Manning, A., 1999. Chapter 47 The causes and consequences of longterm unemployment in Europe. *Handb. Labor. Econ.* 3. Elsevier, pp. 3085–3139.
- Maibom, J., Harmon, N., Glenny, A., Fluchtmann, J., 2023. Unemployed job search across people and over time: evidence from applied-for jobs. *J. Labor Econ.*
- Maibom, J., Roshholm, M., Svarer, M., 2017. Experimental evidence on the effects of early meetings and activation. *Scand. J. Econ.* 119, 541–570.
- Manoli, D.S., Michaelides, M., Patel, A., 2018. Long-term and heterogeneous effects of job-search assistance. NBER Working Paper 24422, National Bureau of Economic Research.
- Marinescu, I., 2017. The general equilibrium impacts of unemployment insurance: evidence from a large online job board. *J. Public Econ.* 150, 14–29.
- Marinescu, I., Skandalis, D., 2021. Unemployment insurance and job search behavior. *Q. J. Econ.* 136, 887–931.
- Massenkoff, M., 2023. Job search and unemployment insurance: new evidence from claimant audits.
- McCall, B., Smith, J., Wunsch, C., 2016. Chapter 9 – Government-sponsored vocational education for adults. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), *Handbook of the Economics of Education* 5. Elsevier, pp. 479–652.
- McCall, B.P., 1996. Unemployment insurance rules, joblessness, and part-time work. *Econometrica* 64, 647–682 (publisher: [Wiley, Econometric Society]).
- McCall, J.J., 1970. Economics of information and job search. *Q. J. Econ.* 84, 113–126 (publisher: Oxford University Press).
- McConnell, S., Schochet, P.Z., Rotz, D., Fortson, K., Burkander, P., Mastri, A., 2021. The effects of employment counseling on labor market outcomes for adults and dislocated workers: evidence from a nationally representative experiment. *J. Policy Anal. Manag.* 40, 1249–1287.
- McCrory, J., 2008. Manipulation of the running variable in the regression discontinuity design: a density test. *J. Econom.* 142, 698–714.
- McGee, A.D., 2015. How the perception of control influences unemployed job search. *ILR Rev.* 68, 184–211 (publisher: SAGE Publications Inc.).
- McKay, A., Reis, R., 2021. Optimal automatic stabilizers. *Rev. Econ. Stud.* 88, 2375–2406.
- McKenzie, D., 2017. How effective are active labor market policies in developing countries? A critical review of recent evidence. *World Bank. Res. Observer* 32, 127–154.
- Meyer, B.D., 1990. Unemployment insurance and unemployment spells. *Econometrica* (1986–1998) 58, 757 num Pages: 26 Place: Evanston, United Kingdom Publisher: Blackwell Publishing Ltd.
- Meyer, B.D., 2002. Unemployment and workers' compensation programmes: rationale, design, labour supply and income support. *Fisc. Stud.* 23, 1–49 (publisher: Wiley).
- Michaelides, M., Mueser, P., 2020. The labor market effects of US reemployment policy: lessons from an analysis of four programs during the great recession. *J. Labor. Econ.* 38, 1099–1140.
- Michaelides, M., Mueser, P., 2023. The employment and displacement effects of job counselling: evidence from the US unemployment insurance system. Available at SSRN 4331161.
- Michaillat, P., 2012. Do matching frictions explain unemployment? Not in bad times. *Am. Econ. Rev.* 102, 1721–1750.
- Moffitt, R., 1985. Unemployment insurance and the distribution of unemployment spells. *J. Econ.* 28, 85–101.
- Mortensen, D.T., 1970. Job search, the duration of unemployment, and the Phillips curve. *Am. Econ. Rev.* 60, 847–862 (publisher: American Economic Association).
- Mortensen, D.T., 1986. Chapter 15 Job search and labor market analysis. *Handbook of Labor Economics*, vol. 2. Elsevier, pp. 849–919.

- Mueller, A.I., Rothstein, J., von Wachter, T.M., 2016. Unemployment insurance and disability insurance in the great recession. *J. Labor. Econ.* 34, S445–S475 (publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago]).
- Mueller, A.I., Spinnewijn, J., 2023. The nature of long-term unemployment: predictability, heterogeneity and selection.
- Mueller, A.I., Spinnewijn, J., Topa, G., 2021. Job seekers' perceptions and employment prospects: heterogeneity, duration dependence, and bias. *Am. Econ. Rev.* 111, 324–363.
- Muralidharan, K., Niehaus, P., Sukhtankar, S., 2023. General equilibrium effects of (improving) public employment programs: experimental evidence from India. *Econometrica* 91, 1261–1295.
- Naidu, S., Sojourner, A., 2020. Employer power and employee skills.
- Nekoei, A., Weber, A., 2015. Recall expectations and duration dependence. *Am. Econ. Rev.* 105, 142–146.
- Nekoei, A., Weber, A., 2017. Does extending unemployment benefits improve job quality? *Am. Econ. Rev.* 107, 527–561.
- Neumark, D., 2018. Experimental research on labor market discrimination. *J. Econ. Lit.* 56, 799–866.
- Nunley, J.M., Pugh, A., Romero, N., Seals, R.A., 2017. The effects of unemployment and underemployment on employment opportunities: results from a correspondence audit of the labor market for college graduates. *ILR Rev.* 70, 642–669 (publisher: SAGE Publications Inc).
- O'Donoghue, T., Rabin, M., 1999. Doing it now or later. *Am. Econ. Rev.* 89, 103–124.
- Paserman, M.D., 2008. Job search and hyperbolic discounting: structural estimation and policy evaluation*. *Econ. J.* 118, 1418–1452. (_eprint). (<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0297.2008.02175.x>).
- Pavoni, N., 2009. Optimal unemployment insurance, with human capital depreciation, and duration dependence. *Int. Econ. Rev.* 50, 323–362. (_eprint). (<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2354.2009.00532.x>).
- Potter, T., 2021. Learning and job search dynamics during the Great Recession. *J. Monetary Econ.* 117, 706–722.
- Rasul, I., Bassi, V., Bandiera, O., Burgess, R., Vitali, A., Sulaiman, M., 2023. The search for good jobs: evidence from a six-year field experiment in Uganda. *J. Labor. Econ.*
- Rebolledo-Sanz, Y., 2012. Unemployment insurance and job turnover in Spain. *Labour Econ.* 19, 403–426.
- Riphahn, R.T., Schrader, R., 2023. Reforms of an early retirement pathway in Germany and their labor market effects. *J. Pension. Econ. Financ.* 22, 304–330.
- Rogerson, R., Shimer, R., 2011. Chapter 7 – Search in macroeconomic models of the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics* 4. Elsevier, pp. 619–700.
- Rothstein, J., 2011. Unemployment insurance and job search in the great recession.
- Rothstein, J., Valletta, R.G., 2017. Scraping by: income and program participation after the loss of extended unemployment benefits. *J. Policy Anal. Manag.* 36, 880–908. (_eprint). (<https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.22018>).
- Sarvimäki, M., Hämäläinen, K., 2016. Integrating immigrants: the impact of restructuring active labor market programs. *J. Labor. Econ.* 34, 479–508.
- Schiprowski, A., 2020. The role of caseworkers in unemployment insurance: evidence from unplanned absences. *J. Labor. Econ.* 38, 1189–1225.
- Schiprowski, A., Schmidtke, J., Schmieder, J., Trenkle, S., 2024. The effects of unemployment insurance caseworkers on job search effort. *AEA Pap. Proc.* 114, 567–571.
- Schlosser, A., Shanahan, Y., 2022. Fostering soft skills in active labor market programs: evidence from a large-scale RCT. *SSRN Electron. J.*

- Schmieder, J.F., von Wachter, T., 2016. The effects of unemployment insurance benefits: new evidence and interpretation (eprint). *Annu. Rev. Econ.* 8, 547–581. <https://doi.org/10.1146/annurev-economics-080614-115758>
- Schmieder, J.F., von Wachter, T., Bender, S., 2012. The effects of extended unemployment insurance over the business cycle: evidence from regression discontinuity estimates over 20 years. *Q. J. Econ.* 127, 701–752.
- Schmieder, J.F., von Wachter, T., Bender, S., 2016. The effect of unemployment benefits and nonemployment durations on wages. *Am. Econ. Rev.* 106, 739–777.
- Shavell, S., Weiss, L., 1979. The optimal payment of unemployment insurance benefits over time. *J. Political Econ.* 87, 1347–1362 (publisher: University of Chicago Press).
- Shimer, R., Werning, I., 2006. On the optimal timing of benefits with heterogeneous workers and human capital depreciation.
- Shimer, R., Werning, I., 2007. Reservation wages and unemployment insurance. *Q. J. Econ.* 122, 1145–1185 (publisher: Oxford University Press).
- Shimer, R., Werning, I., 2008. Liquidity and insurance for the unemployed. *Am. Econ. Rev.* 98, 1922–1942 (publisher: American Economic Association).
- Spinnewijn, J., 2015. Unemployed but optimistic: optimal insurance design with biased beliefs. *J. Eur. Econ. Assoc.* 13, 130–167.
- Stephens, M., 2001. The long-run consumption effects of earnings shocks. *Rev. Econ. Stat.* 83, 28–36 (publisher: The MIT Press).
- Stigler, G.J., 1961. The economics of information. *J. Political Econ.* 69, 213–225 (publisher: University of Chicago Press).
- Stigler, G.J., 1962. Information in the labor market. *J. Political Econ.* 70, 94–105 (publisher: University of Chicago Press).
- Srotz, R.H., 1955. Myopia and inconsistency in dynamic utility maximization. *Rev. Econ. Stud.* 23, 165–180 (publisher: [Oxford University Press, Review of Economic Studies, Ltd.]).
- Topel, R.H., 1983. On layoffs and unemployment insurance. *Am. Econ. Rev.* 73, 541–559 (publisher: American Economic Association).
- Topel, R.H., 1984. Experience rating of unemployment insurance and the incidence of unemployment. *J. Law Econ.* 27, 61–90 (publisher: [University of Chicago Press, Booth School of Business, University of Chicago, University of Chicago Law School]).
- Tuit, S., van Ours, J.C., 2010. How changes in unemployment benefit duration affect the inflow into unemployment. *Econ. Lett.* 109, 105–107.
- Uusitalo, R., Verho, J., 2010. The effect of unemployment benefits on re-employment rates: evidence from the Finnish unemployment insurance reform. *Labour Econ.* 17, 643–654.
- Van Doornik, B., Schoenherr, D., Skrastins, J., 2023. Strategic formal layoffs: unemployment insurance and informal labor markets. *Am. Econ. J.: Appl. Econ.* 15, 292–318.
- van Ours, J.C., Vodopivec, M., 2008. Does reducing unemployment insurance generosity reduce job match quality? *J. Public Econ.* 92, 684–695.
- Winter-Ebmer, R., 2003. Benefit duration and unemployment entry: a quasi-experiment in Austria. *Eur. Econ. Rev.* 47, 259–273.
- Wright, R., Kircher, P., Julien, B., Guerrieri, V., 2021. Directed search and competitive search equilibrium: a guided tour. *J. Econ. Lit.* 59, 90–148.

Chapter 7

Families, public policies, and the labor market[☆]

Gordon Dahl^{a,*} and Katrine V. Loken^b

^aUniversity of California, San Diego, ^bNorwegian School of Economics

*Corresponding author. e-mail address: gdahl@ucsd.edu

Chapter Outline

1 Introduction	581	7 Gender inequality	604
2 Conceptual framework	582	7.1 Rates and trends	604
3 Public policies in OECD countries	585	7.2 Public policies and gender inequality	605
4 Fertility	586	8 Child outcomes	606
4.1 Rates and trends	586	8.1 Public policies and child outcomes	606
4.2 Public policies affecting fertility	587	9 Norms and spillovers	609
5 Marriage, divorce, and cohabitation	591	9.1 Public policies and norms and spillovers	609
5.1 Rates and trends	591	10 Lessons learned and avenues for future research	610
5.2 Public policies affecting marriage, divorce, and cohabitation	594	10.1 Short summary	610
6 Family labor supply	596	10.2 Avenues for future research	611
6.1 Rates and trends	596	References	611
6.2 Public policies and family labor supply	598		

1 Introduction

The past two decades has seen a changing landscape of how families and the labor market interact. There have been sweeping changes in fertility, marriage, cohabitation, divorce, child-rearing, and the allocation of men's and women's time to paid and unpaid work. This chapter explores the role of government policies in influencing these trends as they relate to the labor market.

Since the last Handbook of Labor Economics appeared roughly 15 years ago, there has been a sharp increase in the availability and use of administrative data, particularly in Europe, but also in select countries worldwide. This has coincided with a continuing rise in the number of young, talented labor economists worldwide. At the same time, there has been a continuing evolution and refinement of

☆ We thank Christian Cervellera for excellent research assistance.

convincing research designs in applied econometrics. Combined, these recent developments have spurred a new generation of research on how public policies affect families. In this chapter, we review this literature.

To ground the discussion, we begin by laying out a conceptual framework. We outline the incentives and costs which couples and families face when making labor market decisions. This in turn provides a frame of reference for understanding the interaction of public policies and key family outcomes. We then provide a brief overview of the policy landscape across OECD countries.

The ensuing discussion is organized into six topical areas: (i) fertility, (ii) marriage, divorce, and cohabitation, (iii) family labor supply, (iv) gender inequality, (v) child outcomes, and (vi) norms and spillovers. In each section, we first present descriptive statistics on these outcomes across OECD countries, when possible, for 2019. Although data for many countries is available for later periods, we focus on 2019 since it is prior to the covid pandemic, and this chapter is not about the effect of pandemic-era policies. We also examine trends over the same set of OECD countries over time, focusing on recent patterns for the period from 2001 to 2019 insofar as the data is available.

We classify countries into six distinct groups: Northern Europe (Iceland, Denmark, Sweden, Norway, and Finland), Central Europe (Switzerland, Netherlands, Luxembourg, Ireland, Germany, France, Belgium, and Austria), Southern Europe (Spain, Portugal, Malta, Italy, Greece, and Cyprus), Eastern Europe (Slovenia, Slovak Republic, Romania, Poland, Lithuania, Latvia, Hungary, Estonia, Czechia, Croatia, and Bulgaria), Anglo-Saxon countries (United States, United Kingdom, New Zealand, Canada, and Australia), and a miscellaneous category (Turkey, Mexico, Korea, Japan, Israel, Costa Rica, Colombia, and Chile).

After laying out the recent rates and trends for a topic area, we delve into the existing literature on public policies related to that topic area. Some public policies have the potential to affect more than one topic area, in which case the public policy will be discussed in more than one section. In each section, we synthesize the general findings into a collective discussion, tying them back to the conceptual framework we lay out.

The chapter concludes by offering some thoughts on the general lessons learned, as well as emerging trends and avenues for future research.

2 Conceptual framework

We begin by outlining the conceptual underpinnings behind family decisions which interact with the labor market, and which could be affected by public policies. Modern families can take many forms: married couples, cohabiting partners, families with and without children, and joint versus solo parenting. For simplicity of notation, in this section, the main actors are taken to be a man and a woman, although we recognize that other gender mixes and family arrangements are also possible. The framework is intentionally general, with the point of highlighting how government policies could affect families.

A starting point is the unitary model, where partnered couples maximize a joint utility function. This work was pioneered by [Mincer \(1962\)](#) and formalized by [Ashenfelter and Heckman \(1974\)](#). These papers recognized that analyzing labor force decisions for men and women in a family context depends not only on market wages and each individual's demand for leisure, but also on joint leisure and home production of goods and services for the family. They highlighted that one of the most important elements of home production in a family is investments in the raising of children.

The key elements of the unitary model are:

- Partnered family utility function: $U(h_m, h_f, n_m, n_f, x_m, x_f, x_j, x_m^h, x_f^h, x_j^h, C)$
- Unpartnered utility functions: $U(h_i, n_i, x_i, x_i^h, C)$, $i = m, f$
- Housework production function: $x_i^h(n_m^h, n_f^h)$, $i = m, f, j$
- Child production function: $C(n_m^c, n_f^c, x_c, P)$

where the subscripts m , f , c , and j denote males, females, children, and households, respectively. The inputs into the functions are hours worked in the labor market (h_i), hours spent in unpaid work (n_i) – divided into child rearing (n_i^c) and housework (n_i^h), purchased goods (x_i) – which can be private (m, f, c) or joint (j), home produced goods (x^h) – which can be private or joint, and an indicator for whether both partners are present in the household (P).

The budget constraints in this model can be written as:

- Partnered financial budget constraint: $p_m x_m + p_f x_f + p_j x_j + p_c x_c < w_m h_m + w_f h_f + G - t(w_m h_m, w_f h_f)$
- Unpartnered financial budget constraints: $p_i x_i + p_c x_c < w_i h_i + G - t(w_i h_i)$, $i = m, f$
- Time budget constraints: $T_i = h_i + n_i + l_i$, $i = m, f$

where p_i are market prices, w_i are market wages, T_i is total time available, l_i is leisure, G is government transfers (which can vary with marriage and children), and $t(\cdot)$ is the tax function (which can vary with marriage and children).

As recognized early on by [Becker \(1973\)](#), a limitation of the unitary model is that partners maximize a joint utility function which does not distinguish by source of income. This household aggregation implies that who works in the labor market versus at home does not influence the types of goods purchased or the amount of investment in children.

Subsequent work developed what is known as the collective model. In this model, the man and the woman are not completely altruistic and do not necessarily have the same decision-making weights for consumption choices. Instead, factors such as the amount of income each party earns in the market influence bargaining power within a marriage. These models became the standard starting in the 1990s, with influential work by [Chiappori \(1992\)](#), [Lundberg and Pollak \(1993\)](#), [Fortin and Lacroix \(1997\)](#), and [Browning and Chiappori \(1998\)](#). Empirical work is generally more consistent with the collective model compared to the unitary

model. While there have been further refinements to this workhorse model in the past two decades, the core logic remains: bargaining power matters for the labor supply of each partner, individual and joint consumption choices, investments in children, and the value of marriage.

While we will not detail the refinements to the collective model in this chapter, we draw the reader's attention to one key insight which has been developed in the past two decades. This is the recognition that changing gender norms, which would show up as changes in utility functions, have the potential to affect families. For example, traditional beliefs about men being the primary breadwinner and women staying home to rear children have started to give way to more gender-equal preferences and norms. This alters the benefits and costs of marriage and children, as well as changes the labor market investments of men and women. We refer the interested reader to the chapter in this handbook by Olivetti, Petrongolo, and Pan which goes into more details. An interesting question we will return to in this chapter is whether government policies can affect such norms.

Regardless of a model's specific formulation, family decision-making models highlight the interdependent nature of partners' decisions and have implications for both labor market and family outcomes. To summarize a few of these, marriage and cohabitation surpluses arise due to joint leisure complementarities, joint consumption goods, and specialization in market versus unpaid work. Marriage (and cohabitation) further affect fertility decisions and the child production function. Divorce naturally occurs in these models when the surplus from marriage turns negative. Marital breakup can occur even if the total possible surplus in the partnership is positive, but the couple cannot reach an agreement on the division of time use, consumption portfolio, or child investments. In turn, the decision on whether to have children, and how many, depends on preferences and costs.

The general model also highlights that child rearing requires inputs of both time and money, but that not all child inputs can be purchased in the market. Family labor supply interacts with these decisions, and relative wage rates and comparative advantage can lead to corner solutions where one partner specializes in home production or child rearing. Staying home to raise a child involves not only an immediate cost, but also a potential loss in human capital if it depreciates due to reduced investments in one's professional career.

We have been intentionally general about the functional forms for utility, the child production function, and the exact nature of household bargaining. Our goal in presenting a simple conceptual framework is instead to highlight the factors which uniquely affect family decisions, so that we can in turn discuss how these choices are impacted by public policies.

To provide just one example, both the unitary and collective models predict some degree of specialization in home versus market work based on comparative advantage. This feature could matter for government policies, such as the tax code. Taxing earnings at the individual versus household level will have different impacts on male and female labor supply for partnered couples, and could influence the decision to marry and have children.

More generally, there are a wide variety of policies available to governments which impact families. These include subsidies for specific inputs or goods (e.g., parental leave, child care subsidies), cash transfers (fertility bonuses, child tax credits), and taxes (individual versus joint taxation, tax treatment of children, lone mother tax credits). We discuss the predictions of the conceptual model as we discuss each of these issues in the sections which follow. Laws regarding marriage, divorce, custody, alimony, and abortion are also important considerations, but not covered in this chapter.

3 Public policies in OECD countries

Before we begin our review of how different public policies affect outcomes, we first briefly document the landscape for some of these policies across OECD countries in two graphs.

[Fig. 1](#) shows per capita social expenditure on children in early childhood as of 2019 for our OECD countries. Amounts have been converted to US dollars and are purchasing power parity adjusted to make them comparable. Almost all

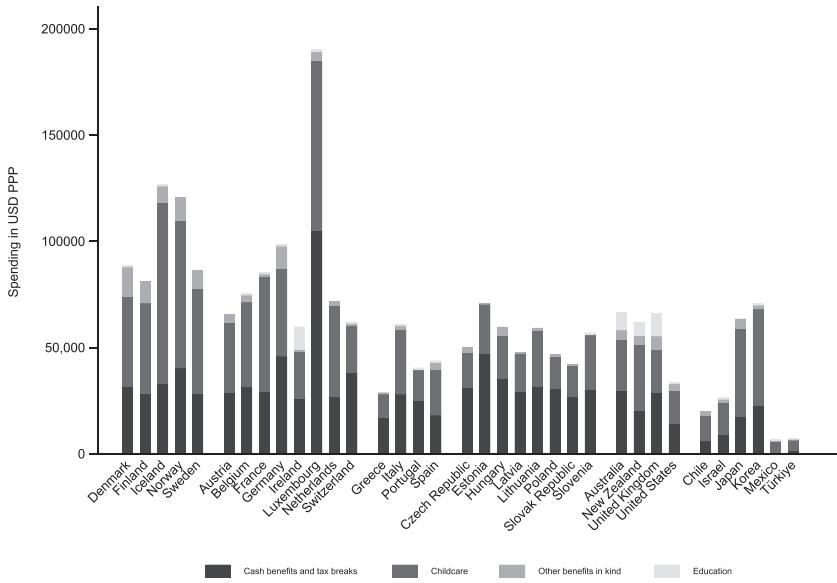


FIG. 1 Per capita social expenditure for children across OECD countries, 2019. [Notes: The source is the OECD family database, PF1.6.C. Countries are classified into six distinct groups. Northern Europe includes Iceland, Denmark, Sweden, Norway, and Finland; Central Europe includes Switzerland, Netherlands, Luxembourg, Ireland, Germany, France, Belgium, and Austria; Southern Europe includes Spain, Portugal, Malta, Italy, Greece, and Cyprus; Eastern Europe includes Slovenia, Slovak Republic, Romania, Poland, Lithuania, Latvia, Hungary, Estonia, Czechia, Croatia, and Bulgaria; Anglo-Saxon countries include United States, United Kingdom, New Zealand, Canada, and Australia; the miscellaneous category include Turkey, Mexico, Korea, Japan, Israel, Costa Rica, Colombia, and Chile.]

countries have significant spending on cash benefits/tax breaks and child care. In most countries, the amount spent on childcare is roughly equivalent to cash benefits/tax breaks. The largest expenditures are found in the Nordic and Central European countries, with Luxembourg being a clear outlier. Southern and Eastern European countries, Anglo-Saxon countries, and Japan and Korea spend considerably less, but still substantial amounts. A few countries, such as Chile, Israel, Mexico and Turkey, spend very little.

Fig. 2 shows the average duration of paid family leave entitlements as of 2022. Note that the graph does not show variation in replacement rates, which can vary from very low to full replacement of wages across different countries. All countries, with the exception of the US, have some mandated parental leave at the federal level. There is substantial variation across countries, with longer leaves on average in Northern European, Central, and Eastern European countries. Finland, Hungary, and the Slovak Republic stand out as having especially long leave durations, while the Netherlands, Switzerland, Australia, Colombia, Costa Rica, Korea, Mexico, and Turkey have particularly short durations. Another difference across countries is the amount of leave reserved specifically for mothers or fathers versus shareable leave.

4 Fertility

4.1 Rates and trends

Fig. 3 displays fertility rates across OECD countries in 2019, revealing that all but one country (Israel) is below the replacement threshold of 2.1 children per

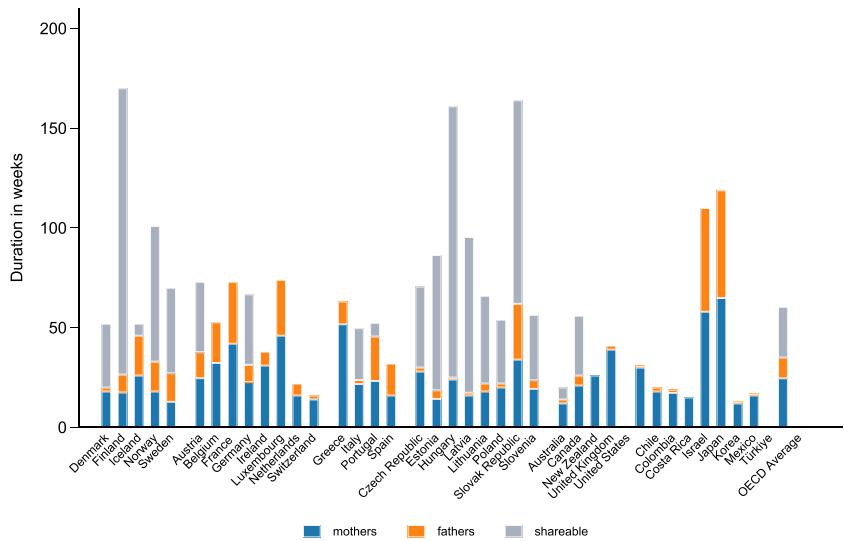


FIG. 2 Reserved and shareable paid family leave entitlements across OECD countries, 2019]. [Notes: The source is the OECD Family Database, PF2.1. See Fig. 1 for the classification of countries into groups.]

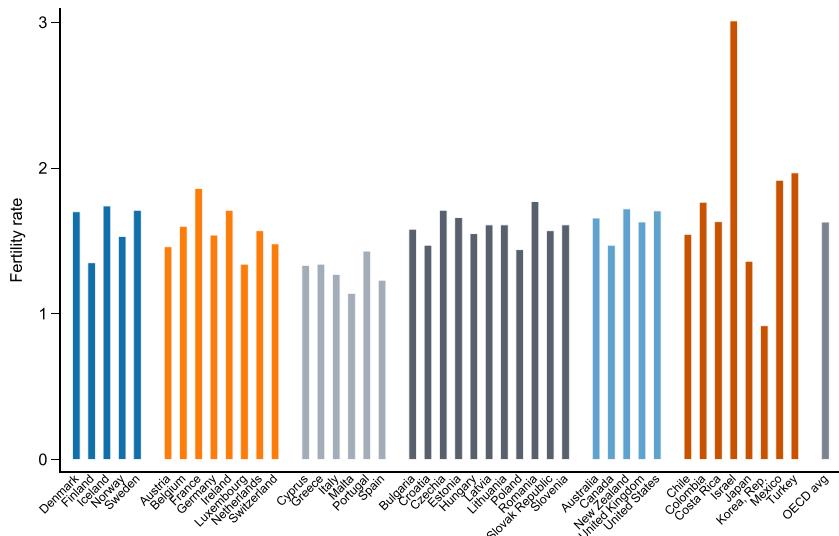


FIG. 3 Fertility rates across OECD countries, 2019. [Notes: World Bank data. See Fig. 1 for the classification of countries into groups.]

woman. Of particular concern are the notably low rates observed in some East Asian and South European countries. Fig. 4 documents time trends for six distinct geographic areas: Central, Eastern, Northern and Southern Europe as well as the UK and the US. In all country groups, there have been significant declines in fertility between 1973 to 2019. At the start of the period, most country groups were near the replacement rate, but by 2019, fertility had fallen to 1.7 children per woman or less. A shrinking population of working-age individuals has implications for a wide array of outcomes, including labor productivity, gross domestic product, tax revenue, and retirement programs.

One plausible factor contributing to declining fertility is the concurrent increase in marriage ages (age at first birth has likewise increased in most countries). There is substantial variation in the age at first marriage across our six OECD country groups, as shown in Fig. 5. The highest marriage ages are found in Northern Europe (men: age 35, women: age 33) and the lowest are found in the US (men: age 29, women: age 27). Despite the differences in levels, the trends have been remarkably similar across OECD countries between 2002 and 2019, with the age at first marriage rising by roughly 2 years for both genders.

These universal trends prompt the question: What influence have various public policies had in both contributing to and mitigating the decline in fertility?

4.2 Public policies affecting fertility

Looking back to our conceptual framework, the decision on whether to have children, and how many, depends on both utility and costs. Public policies



FIG. 4 Fertility trends, 1973–2019. [Notes: World Bank data. See Fig. 1 for the classification of countries into groups.]

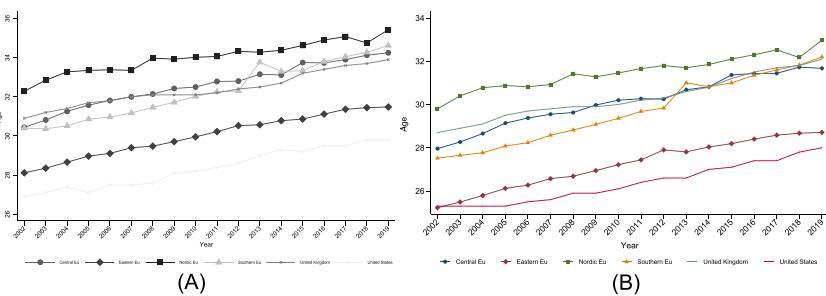


FIG. 5 Age at first marriage trends by gender, 2002–2019. [Notes: The source is the OECD family database, SF3.1. See Fig. 1 for the classification of countries into groups.]

could impact fertility by changing the incentives to raise a child. Policies could operate both indirectly by changing the benefits and costs of time spent on child rearing versus in the labor market and more directly by affecting the household budget constraint with cash transfers or family-friendly tax policies.

Navigating the empirical literature on the impact of public policies on fertility presents a challenge. While methods for causal identification works well for uncovering short-term and immediate responses, fertility decisions are inherently long-term and may not promptly react to policy changes. In addition the majority of studies we have identified look at higher order fertility as identification is based on comparing similar mothers before and after reforms

for those that already have made the decision to have a first child (notable exceptions are [Adda et al. \(2017\)](#) and [Raute \(2019\)](#)). Bearing these caveats in mind, we now dive into the literature on public policies and their influence on fertility.

First we review studies on the effects of parental leave and child care on fertility. Then we move on to the literature which analyzes various cash and near-cash policies.

In-kind subsidies. A seminal study on parental leave and its impact on fertility is [Lalive and Zweimüller \(2009\)](#). This research focuses on a reform in Austria that extended parental leave from one to two years for children born after a specific date. They find that mothers who gave birth to their first child immediately after the reform went on to have a second child at higher rates compared to mothers who had their first child before the reform. However, it is important to note that this effect primarily pertains to the timing of births and higher order births, and does not conclusively identify the impact on overall completed fertility.

Other papers examining the relationship between parental leave policies and higher-order fertility include [Cannonier \(2014\)](#), which evaluates the Family Care Act in the US; [Bassford and Fisher \(2020\)](#), which investigates the introduction of 18 weeks of leave in Australia and its effects on fertility intentions; and [Ang \(2015\)](#), which explores similar dynamics in Canada. These studies generally corroborate the findings of [Lalive and Zweimüller \(2009\)](#), showing positive effects on fertility intentions or birth timing. On the other hand, [Dahl et al. \(2016\)](#) analyses a series of expansions in paid maternity leave in Norway and finds little effect on completed fertility. [Kluve and Schmitz \(2018\)](#) and [Cygan-Rehm \(2016\)](#) analyze a German reform and find no effect, or even a small negative effect, on subsequent birth timing.

A recent contribution to the literature by [Raute \(2019\)](#) makes a significant advancement. Using German data, the study is able to identify effects for the extensive margin in the medium run. The analysis reveals positive effects on fertility decisions up to five years after the implementation of reforms in earnings-related maternity benefits, in particular for highly educated and higher-earning women. This highlights that parental leave policies can indeed influence fertility decisions over the medium term, offering valuable insights into their broader impact beyond possible effects on higher-order births.

There has also been work on paternity quotas and fertility, focusing on higher-order births. [Kotsadam and Finseraas \(2011\)](#); [Cools et al. \(2015\)](#); [Dahl et al. \(2016\)](#) and [Hart et al. \(2022\)](#) found no effect of paternity leave extensions on fertility in Norway, and [Bartel et al. \(2018\)](#) finds no effect for California. Contrary to these null effects, [Farré and González \(2019\)](#) shows that two weeks of paid paternity leave in Spain led to delays in subsequent fertility up to six years after the birth of the first child. The effects are driven by two channels. First, fathers' increased their involvement in childcare, which led to higher labor force attachment among mothers, and hence possibly higher opportunity costs for mothers to

have an additional child. Second, men reported lower desired fertility after the reform, possibly due to their increased awareness of the costs of childrearing, or to a shift in preferences from child quantity to quality.

In comparison, the literature on child care access and fertility is scarce. A well-identified paper from Germany using expansions of child care for children below age three finds positive effects of child care on fertility, especially along the intensive margin (Bauernschuster et al., 2016). Mörk et al. (2013) studies changes in child care subsidies in Sweden and finds some effect on the timing of birth but no effect on completed fertility. Wood and Neels (2019) finds effects for first births, but not for higher order births, from an expansion in child care for children below age three in Belgium. Wang (2022) estimates the joint decisions of fertility and female labor supply using two recent German reforms affecting parental leave and public child care. The paper finds positive effects on fertility, but with parental leave only impacting higher educated women, while the child care reform affected all women. They find the effect is much larger under individual taxation than under a joint taxation regime. This underscores the importance of studying public policies jointly and not only in isolation.

In sum, the literature on the role of in-kind transfers in influencing fertility indicates that certain programs have the potential to make an impact. The effects are heterogeneous across countries and groups of people, and are generally observed more for higher-order births. This does not mean that policies cannot affect first births and completed fertility, but rather reflects that there are few well-identified studies for women who have not yet started their fertility cycles. Since the estimated effects are relatively small (and several studies find no effect), we conclude that in-kind benefits alone are unlikely to be an effective policy tool for addressing low fertility rates. Nevertheless, when combined with other incentives in the labor market, such as tax breaks and cash transfers, in-kind benefits could potentially have a more significant impact.

Cash and near-cash transfers. Cash and financial incentives are also active policy tools used in many countries and there are several papers studying their effect on fertility. Adda et al. (2017) simulate the impact of a pronatalist cash transfer and find large short-term effects on fertility but smaller long-run effects which are concentrated among younger women. This paper is one of the few that look at extensive-margin fertility effects for those that have not yet had children. Laroque and Salanié (2014) and Haan and Wrohlich (2011) exploit cross-sectional variation in financial incentives resulting from the French and German tax and transfer systems, driven by differences in household characteristics. Both papers find sizable fertility effects of a simulated, universal child subsidy. Other related papers exploit variation in universal child subsidies for third (or higher order) children in Quebec (Milligan, 2005) and Israel (Cohen et al., 2013). Both papers find a strong pronatal effect. Riphahn and Wiynck (2017) assess the effects of a child benefit reform in Germany and find a modest positive effect, but only for second-order births in higher-income households.

[Sandner and Wiynck \(2023\)](#) study the fertility response to cutting child-related welfare benefits and find a negative effect on higher order births. A more recent paper by [Dachille et al. \(2023\)](#) evaluates an Italian 2019 reform that guaranteed minimum income. Even though the program was targeted to combat poverty, they also find positive effects on fertility. [Elmallakh \(2023\)](#) studies fertility responses to a 2014 French reform restricting access to child allowances. She finds negative effects on fertility, especially among high income individuals. [Yonzan et al. \(2024\)](#) find positive fertility effects in households receiving more dividend transfers after a policy change to the Alaska Permanent Fund. [González and Trommlerová \(2023\)](#) studies fertility responses to a cash transfer program in Spain. They examine the introduction and cancellation of the program and find a larger fertility decline after the cancellation.

In contrast, [Reader et al. \(2022\)](#) finds very small to no effect on fertility after a reduction of child benefits in the UK. This targeted cut affected low income families and is therefore in line with other studies which find larger fertility effects for high income families. A similar conclusion arises in [Aizer et al. \(2024\)](#) finding no long-term fertility effects from increased cash transfers to the poor in the US.

All in all, the literature on cash transfers suggest that there are potential fertility effects of providing families cash. However the effects sizes are often relatively modest. The literature typically finds larger effects for high income families, which suggests lower-income families face different economic or social constraints. As with in-kind public policies and fertility, most of the evidence for cash transfers is for higher order births.

5 Marriage, divorce, and cohabitation

5.1 Rates and trends

[Fig. 6](#) illustrates the substantial variation in marriage rates across OECD countries, with Southern European nations exhibiting some of the lowest rates and Eastern European countries tending toward the highest rates. The US also stands out with relatively high marriage rates. Meanwhile, [Fig. 7](#) illustrates relatively minor changes in marriage rates for most OECD country groups from 2002 to 2019, with some movement up and down. Two exceptions are the US and the UK, both of which have experienced sizable declines. Turning to divorce trends in [Fig. 8](#), we observe relatively stable rates between 2002 and 2019 across most OECD country groups. The exceptions are again the US and the UK, both of which have seen divorce rates drop in tandem with the decline in marriage.

In [Fig. 9](#) we complete the picture on partnerships by plotting cohabitation rates for families with children. The Nordic countries have the highest rate of cohabitation, with roughly 20 % of children living with cohabiting partners – a rate which has remained constant between 2005 and 2018. The US and

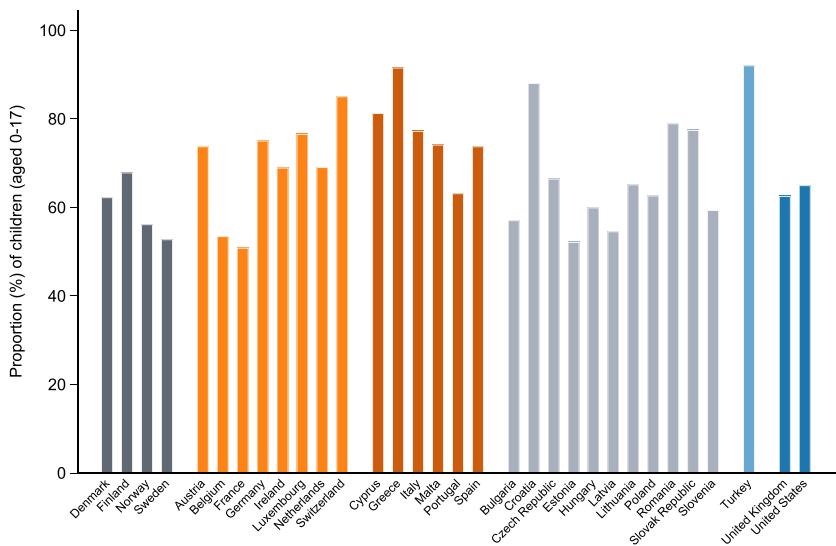


FIG. 6 Marriage rates across OECD countries, 2019. [Notes: The source is the OECD family database, SF3.1. See Fig. 1 for the classification of countries into groups.]

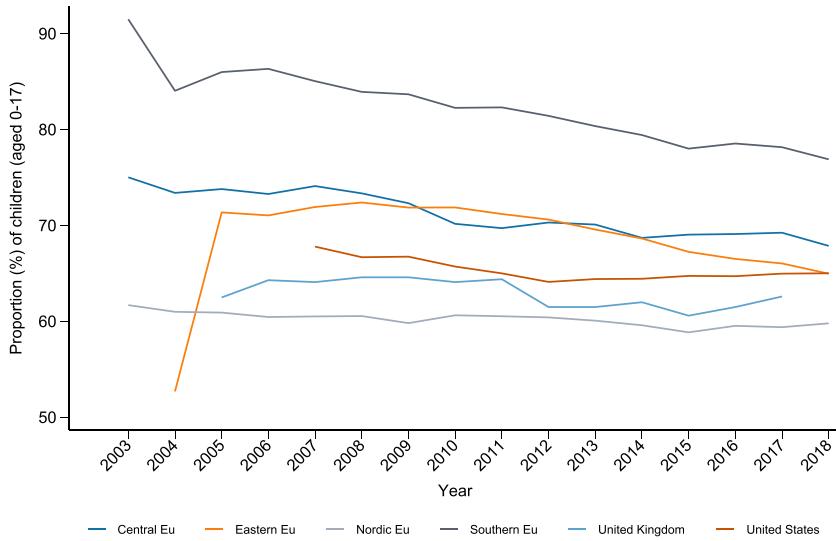


FIG. 7 Marriage trends, 2002–2019. [Notes: The source is the OECD family database, SF3.1. See Fig. 1 for the classification of countries into groups.]

Southern Europe have the lowest rates of cohabitation. Central, Eastern, and Southern Europe, along with the UK have all seen rises in cohabitation over time.

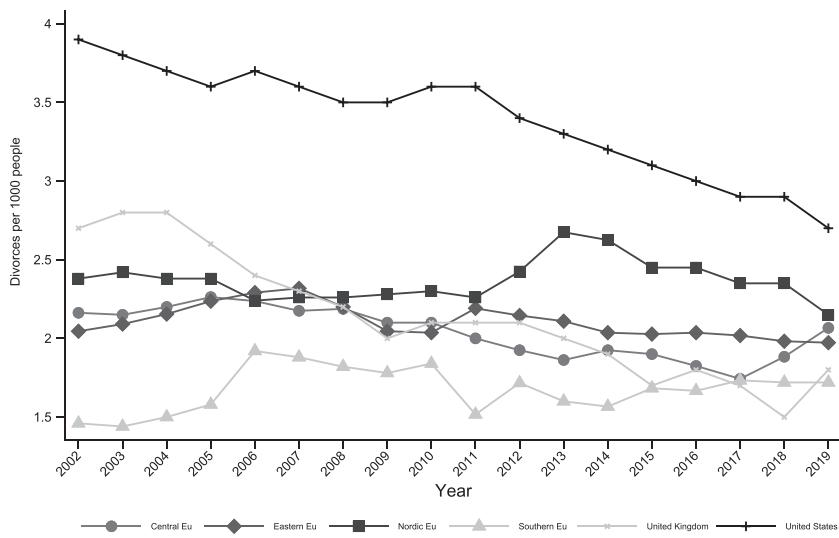


FIG. 8 Divorce trends, 2002–2019. [Notes: The source is the OECD family database, SF3.1. See Fig. 1 for the classification of countries into groups.]

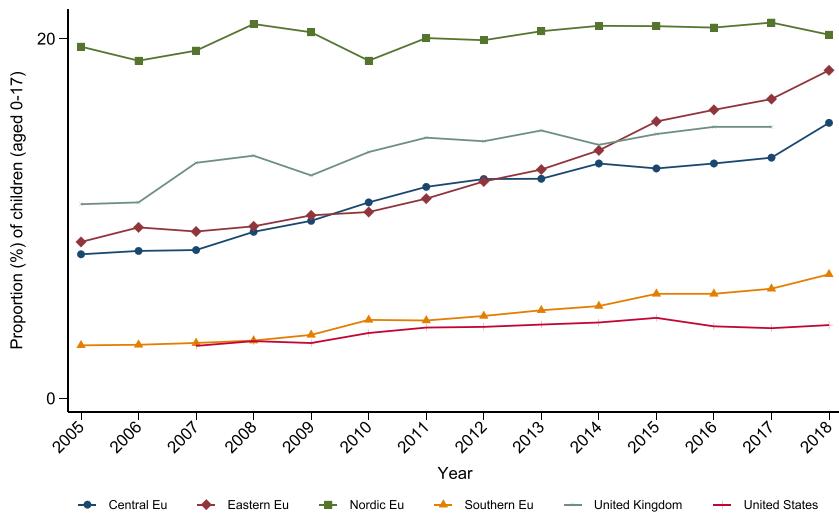


FIG. 9 Cohabitation trends, 2005–2018. [Notes: The source is the OECD family database, SF3.1. See Fig. 1 for the classification of countries into groups.]

Finally, in Fig. 10 we plot the fraction of children who reside in single-parent households over time. Rates are highest in the US and the UK and lowest in Southern Europe. While there has been an increase in single parenthood in Southern Europe over time, for the other country groups the rates have remained relatively stable.

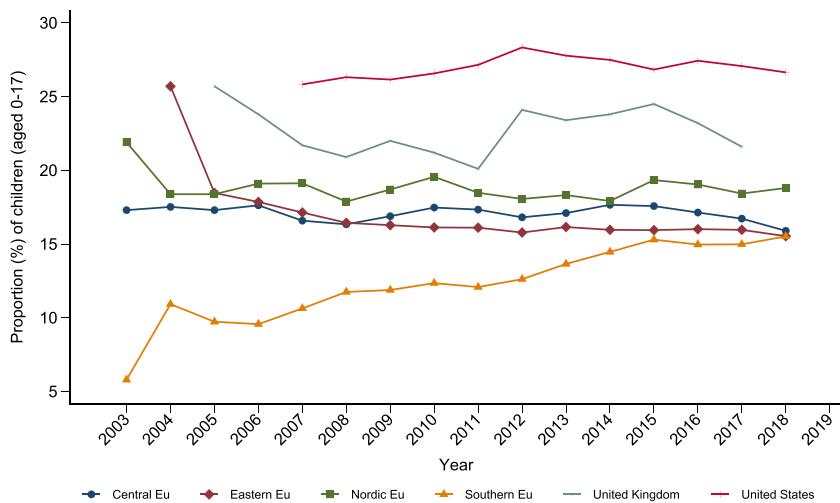


FIG. 10 Single parenthood trends, 2003–2018. [Notes: The source is the OECD family database, SF3.1. See Fig. 1 for the classification of countries into groups.]

5.2 Public policies affecting marriage, divorce, and cohabitation

In this section, we discuss how public policies affect marriage and divorce, limiting the discussion to policies discussed elsewhere in this chapter. We do not discuss changes in laws or the broader social support system.

In-kind subsidies. Most papers which study in-kind subsidies for families, i.e., parental leave or subsidized childcare, do not study marriage or divorce as primary outcomes. We therefore highlight just a handful of studies. [Cygani-Rehm et al. \(2018\)](#) finds that paid parental leave in Germany increases the probability that a newborn lives with cohabiting parents and reduces the risk of single parenthood. In a study examining the introduction of one month of parental leave earmarked for fathers, [Olafsson and Steingrimsdottir \(2020\)](#) find that this reduces divorce up to 15 years after the birth of a child. Effects are strongest for couples where the mother is more educated than the father.

In contrast, [Avdic and Karimi \(2018\)](#) finds that paid paternity leave in Sweden increases the probability of separation, with stronger effects for couples where the mother has equal or higher education than the father. Examining a series of policy reforms in Norway which expanded shared parental leave, [Dahl et al. \(2016\)](#) find no impact on either marriage or divorce. Looking at an extension of fathers' leave allotment in Norway, [Hart et al. \(2022\)](#) likewise find no effect on union stability.

While there are papers which examine how publicly funded childcare affects parental wellbeing ([Baker et al., 2008](#) and [Schmitz, 2020](#)), we found almost no research for the causal impact on marriage, divorce, and cohabitation. Looking at the extension of the school day in Mexico, which can be viewed as an implicit childcare subsidy, [Padilla-Romo et al. \(2022\)](#) find that this increases divorce rates.

Cash and near-cash transfers. There is a larger literature on how the social safety net affects marriage, divorce, and cohabitation. Studying one of the earliest safety net programs in the US, [Aizer et al. \(2024\)](#) find that the Mothers' Pension Program delayed marriage. Early studies of two welfare programs in the US, the Aid to Families with Dependent Children (AFDC) and Temporary Assistance to Needy Families (TANF) found little evidence that traditional welfare discouraged marriage (see [Moffitt, 1998](#) for a review). Subsequent work (see [Moffitt et al., 2020](#) and the cites therein) finds that most reforms to US welfare programs did not have large impacts on marriage, divorce, or cohabitation, but that some work-related reforms decreased marriage and increased single parenthood. A deeper discussion of welfare support systems and their effect on marriage is outside the scope of this chapter.

Another strand of the literature focuses on how tax systems influence the decision to marry, divorce, or cohabit. Looking at the Earned Income Tax Credit (EITC), [Bastian \(2017\)](#) finds that the program increased marriage and reduced cohabitation. This reflects a net effect, as there are marriage gains for some couples but a penalty for others.

In several countries, including the US, France, Germany, and Switzerland, taxation occurs at the household rather than individual level. Joint taxation can introduce either a marriage tax or a marriage penalty, depending on the relative earnings of each partner. Early work on this topic in the US is summarized by [Alm et al. \(1999\)](#); this early literature finds that marriage penalties in the tax system have the predicted effects on marriage, divorce, and separation, but the size of the overall effects are relatively modest.

Subsequent work largely confirms this conclusion both in the US ([Fisher, 2013; Frazier and McKeehan, 2018](#)) and other countries. Using data for Switzerland and a simulated instrumental variables strategy, [Myohl \(2024\)](#) finds that joint taxation reduces marriage rates for lower-income cohabiting households and those without children. Using German data, [Fink \(2020\)](#) finds that partners with unequal earnings marry earlier, presumably to take advantage of the benefits of joint taxation.

In related work, [Persson \(2020\)](#) finds that the elimination of spousal survivor's insurance in Sweden reduced entry into marriage and increased divorce. [Baker et al. \(2004\)](#) study surviving spouse pensions in Canada, and find that removing remarriage penalties increases the remarriage rate of widows. In new work studying same-sex couples, [Friedberg and Isaac \(2024\)](#), find that the recognition of same-sex marriages for tax purposes led to a small increase in marriages.

Do these in-kind and cash transfers explain the variation in partnerships and fertility across countries? This is a challenging question, as laws and the welfare state more generally vary widely across the globe. But on a broader level, [Halla et al. \(2016\)](#) argue that expansions in public social spending in OECD countries increases both marriage and divorce.

6 Family labor supply

6.1 Rates and trends

Fig. 11 compares the labor force participation rate for mothers across OECD countries in 2019. Higher rates are found in the Nordic countries, followed by Central Europe, then Eastern Europe and Anglo-Saxon countries, and finally Southern Europe (the miscellaneous countries vary widely). For example, over 80 % of mothers work in Sweden, while less than 60 % do in Italy.

Fig. 12 shows trends by geographical areas since 2003. There is generally a positive trend for most country groups, particularly in the later half of the period. However, for the Nordic countries, maternal employment in 2018 is at the same level as in 2003. Levels for the US likewise did not change much between 2003 and 2018, although there was a dip and a rebound after the Great Recession.

In **Fig. 13**, we plot the rate of part time employment for OECD countries. Here, stark differences emerge. The Nordic countries, Southern Europe and Eastern Europe all have relatively low rates, while the Anglo-Saxon countries and particularly Central Europe have a high rate of part-time work for mothers. As seen in **Fig. 14**, these are persistent level differences, with almost no change in part-time employment in any country group over time. This implies that the rising rates in maternal employment seen in **Fig. 12** are due to increases in full-time employment. From these graphs, we cannot tell whether the increase in full-time employment is due to mothers entering the labor force to take

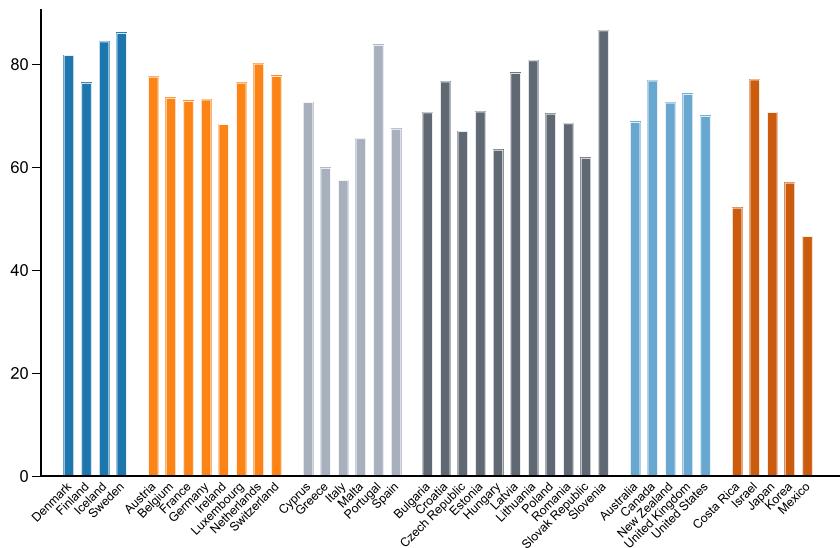


FIG. 11 Maternal employment rates across OECD countries, 2019. [Notes: The source is the OECD family database, LMF1.2. See **Fig. 1** for the classification of countries into groups.]

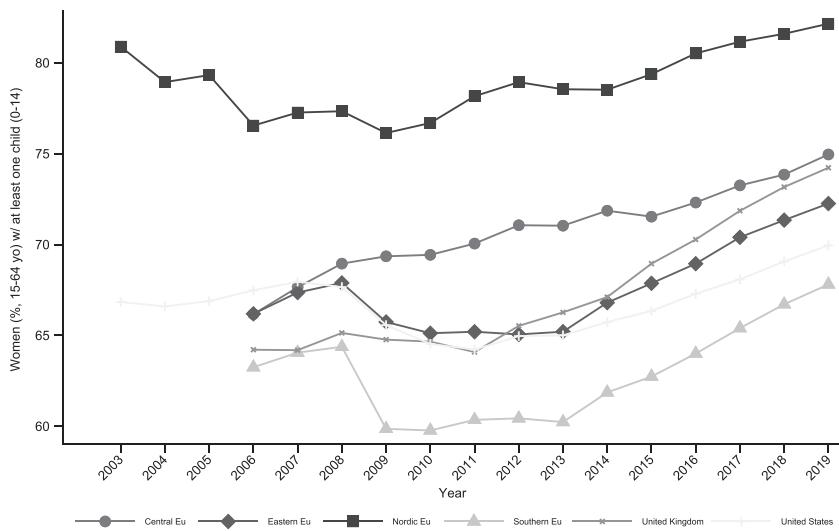


FIG. 12 Maternal employment trends, 2003–2019. [Notes: The source is the OECD family database, LMF1.2. See Fig. 1 for the classification of countries into groups.]

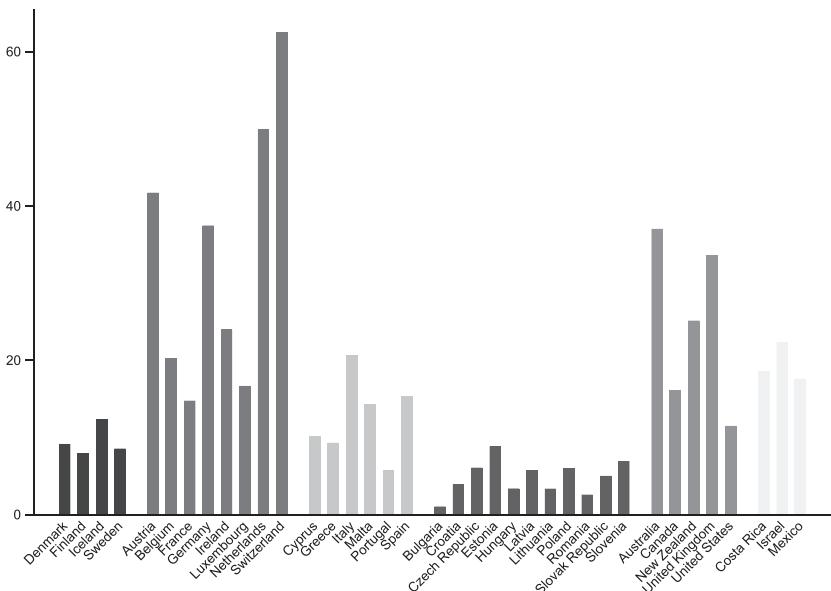


FIG. 13 Part-time maternal employment rates across OECD countries, 2019. [Notes: The source is the OECD family database, LMF1.2. See Fig. 1 for the classification of countries into groups.]

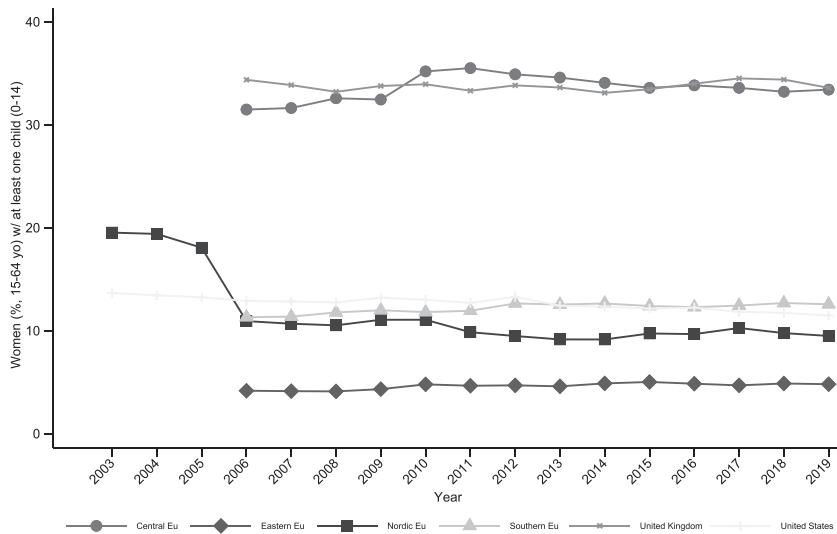


FIG. 14 Part-time maternal employment trends, 2003–2019. [Notes: The source is the OECD family database, LMF1.2. See Fig. 1 for the classification of countries into groups.]

full-time jobs or alternatively due to some mothers entering the labor force to take part-time jobs and other mothers shifting from part-time to full-time work.

While we could not obtain similar data for father's employment across countries, we can look at male employment more generally. In Fig. 15 we plot male employment over time by our six country groupings. There has been a small decline in most country groups, with the exception of Eastern Europe, which has seen a large increase. Interestingly, male employment rates have largely converged over time, with a full-time equivalent employment rate of roughly 75 % for most countries (the UK has a somewhat higher rate). This convergence contrasts with the larger divergence in maternal employment across countries.

A natural question is whether public policies can explain these differences in maternal employment across countries and over time, or whether preferences and norms are largely responsible.

6.2 Public policies and family labor supply

Harkening back to our conceptual framework, public policies could impact family labor supply through changes in the tax code which affect take-home wages, through social assistance programs, or through earmarked subsidies, such as for childcare. There is a sizeable and continually expanding literature on how various public policies affect family labor supply.

When the first volume of the Handbook of Labor Economics was written in 1986, there were chapters devoted to female labor supply, models of marital

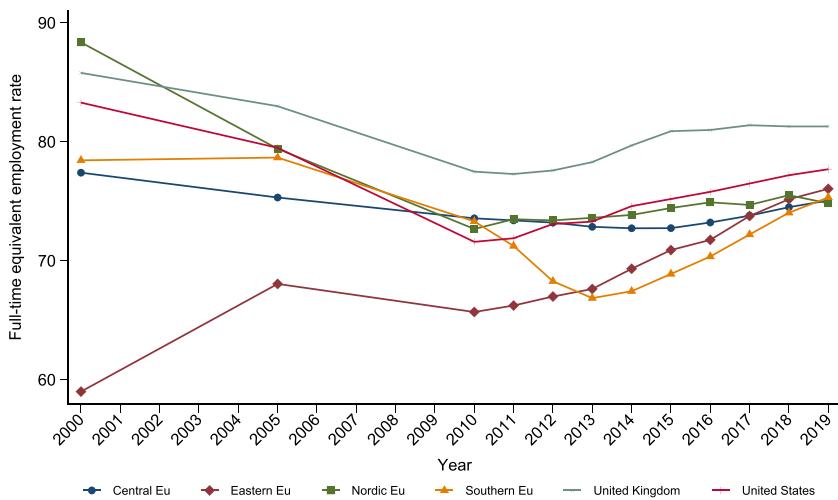


FIG. 15 Male employment trends, 2000–2019. [Notes: The source is the OECD.Stat. See Fig. 1 for the classification of countries into groups.]

status, and home production. Volume 3 in 1999 revisited the labor supply issue by discussing alternative estimation approaches. When Volume 4 was written in 2011, a few chapters touched only indirectly on family labor supply. Therefore, while the current chapter focuses most heavily on contributions since the last volume of the Handbook appeared, we briefly highlight several papers earlier than this on family labor supply to provide a more comprehensive overview.

Cash and near-cash transfers. A first class of public policies which could affect family labor supply are cash or near-cash transfer programs. In virtually all countries, the social safety net treats families different from individuals, offering more generous social assistance for families with children. Such redistributive policies face a potential tradeoff, as they could disincentive work. The first such cash transfer program in the US was the Mothers' Pension Program, which provided mothers a cash allowance to allow their children to remain at home instead of being placed in institutional care or forced to work. [Aizer et al. \(2024\)](#) show that this program had no effect on women's labor force participation. As follow-ups to this early program, both the Aid to Families with Dependent Children (AFDC) and the Temporary Assistance to Needy Families (TANF) programs provided cash transfers to low income families in the US. There is some evidence for work disincentives, but much of this work is based on older statistical methodologies (see [Moffitt \(1992\)](#) and [Ziliak \(2015\)](#) for reviews).

Studying the near-cash transfer program of food stamps in the US, [Hoynes and Schanzenbach \(2012\)](#) find reductions in both the intensive and extensive margins of labor supply, with effects concentrated among single-women

households. A recent paper using Italian data studies the impact of married couple survivor's benefits in Italy ([Giupponi, 2019](#)). Using a regression discontinuity design, the paper finds that benefit cuts sharply increased the surviving spouse's labor supply, consistent with both substantial work capacity and a high value of additional income in widowhood.

In many countries over the past three decades, including most prominently the US and the UK, social assistance programs supporting poor families have shifted away from unconditional cash transfers towards policies which simultaneously aimed to increase paid employment. In the US, there was a decline in the generosity of traditional welfare and an expansion of the Earned Income Tax Credit (EITC). The EITC provides a wage subsidy through the tax code for families with children. The subsidy increases in family earnings in the phase-in range, then plateaus, followed by a phase out range where benefits are taxed back. It was designed to "promote both the values of family and work". Classic studies of the EITC include [Eissa and Liebman \(1996\)](#), [Meyer and Rosenbaum \(2001\)](#), and [Eissa and Hoynes \(2004\)](#). The conclusion from these papers is that families respond to EITC incentives, but not in a uniform manner. EITC expansions promoted employment among eligible unmarried women with children. But since the policy is based on family income, it disincentivized work for secondary earners – who are typically women. [Nichols and Rothstein \(2015\)](#) survey these studies and come up with a consensus estimate that the EITC increases the labor supply of single mothers by 3–6 % points.

More recently, [Kuka and Shenhav \(2024\)](#), study the long-run effects of the large EITC expansion in 1993. They find that 10–19 years after a first birth, single mothers increase cumulative work by 0.6 years and have 4 % higher earnings due to more work experience. [Bastian and Lochner \(2022\)](#) study the introduction of the EITC in 1975 and finds that it increased maternal employment by 6 %.

Around the same time as the EITC expansions, the UK enacted a similar program called the Working Families' Tax Credit (WFTC). The reform likewise led to a an increase in unmarried mothers' labor supply, both by retaining and drawing in lone mothers in the labor market (see [Blundell et al., 2000](#); [Francesconi and Van der Klaauw, 2007](#); [Gregg et al., 2009](#)). The reform included a generous childcare credit, which likely played a role in increasing labor supply. For overviews of such in-work benefit programs across countries, see [Brewer et al. \(2009\)](#). As noted by [Aizer et al. \(2022\)](#), this shift linking welfare benefits to work may have lessened disincentive effects for adult labor supply, but it simultaneously removed much of the safety net for vulnerable children in the lowest income households.

A broader lesson to be learned from the EITC and WFTC studies is that the unit of taxation – the individual or the family – matters for labor supply responses. [Kleven et al. \(2009\)](#) highlight that secondary earners in a household arise when they either have low home productivity or a low cost of

participating in the labor market. Focusing on low participation costs, they conclude the optimal tax system should have negative joint taxation but positive tax rates on secondary earnings. Using a collective model of the household with intra-household bargaining, Alesina et al. (2011) argue that higher marginal tax rates on men, regardless of marital status, is optimal. More recently, Gayle and Shephard (2019) use a collective model and find that the optimal tax system should have negative jointness, but that empirically, the welfare gains from jointness are small.

As a recent paper building on this larger literature highlights, joint family taxation can result in nuanced marriage penalties which result in disparate treatment across racial and ethnic lines (Alm et al., 2023). In this vein, two recent papers using structural life-cycle models examine how marriage-related taxation and benefits for social security in the US affect the labor supply of married men and women (Borella et al., 2023; Groneck and Wallenius, 2021). Both papers conclude that marriage-related provisions decrease the labor supply of married women throughout their lifetimes, and that eliminating the marriage provisions would result in both equity and efficiency gains for much of the population.

Another recent paper (Isaac, 2023) estimates the labor supply effects of joint taxation for same-sex couples, using changes in the recognition of same-sex marriages for tax purposes. The paper finds that joint taxation reduces labor force participation of the lower earner.

Zooming out, research using calibration methods suggests that joint family versus individual taxation policies are partially responsible for the observed differences across countries in married women's labor force participation and the decision to work full versus part time (Bick and Fuchs-Schündeln, 2018).

In-kind subsidies. A second class of public policies which could affect family labor supply are in-kind subsidies which make it easier for mothers and fathers to raise children and work at the same time. One widely implemented policy is subsidized childcare. These subsidies can either be broad-based or targeted to low-income families. They have the potential to increase entry into the labor force, as well as lengthen the number of work hours by lowering the fixed and variable costs of paid work.

Both across countries and over time, more accessible and cheaper child care is correlated with maternal employment. For example, Del Boca et al. (2008) explore variation across European countries in a variety of family policies (including childcare subsidies, parental leave, and family allowances), and conclude the policies have the potential to account for a non-negligible portion of the differences in women's labor market participation across countries.

A key question recent work tries to address is whether these correlations are causal. While the subsidies could be responsible for increased labor force participation of mothers, it is also possible that in countries where more mothers work there is simply a higher demand for paid child care. Work on this topic in the past two decades largely uses quasi-experimental designs. Some studies use

geographic differences in costs and availability (e.g., [Tekin, 2007](#); [Brilli et al., 2016](#)), while other designs leverage variation in the implementation of subsidies across time or geography (see [Morrissey, 2017](#) for a review). We focus on some of the more influential papers in this quasi-experimental literature.

[Baker et al. \(2008\)](#) study the introduction of broad-based, subsidized child care in Quebec, Canada using panel data. They find this heavily subsidized program led to a large increase in the use of non-parental child care for pre-school aged children and a corresponding rise in the employment of women in two-parent families. The design did not allow an analysis of single women. While the increased labor supply of women in two-parent families generated additional tax revenue, this did not cover the subsidy costs, in part because the program resulted in some crowding out of informal child care arrangements. Research by [Lefebvre and Merrigan \(2008\)](#) on the Quebec reform reached similar conclusions, finding larger effects for mothers with young children. A more recent study by [Thomas \(2024\)](#) uses a staggered difference-in-difference design to study the introduction of a child care program in Nova Scotia, Canada. They find a labor supply response which is three times as large as in Quebec, which could be explained by the program being zero cost and with guaranteed availability.

[Havnes and Mogstad \(2011a\)](#) likewise study the introduction of heavily subsidized, universally accessible child care, but in Norway. Using a difference-in-differences design, they find no evidence of a labor supply response despite there being a large correlation between maternal employment and childcare use. Similar to findings for Canada, they find a crowd out of informal child care arrangements and conclude the net cost of such subsidies is high.

[Givord and Marbot \(2015\)](#) look at a temporary increase in childcare subsidies in France. Using a difference-in-differences design, they find a large increase in the use of paid childcare and a small increase in maternal labor supply.

[Cascio \(2009\)](#) studies a related universal program, namely, the introduction of kindergarten in the US. They find a sizeable 30 % increase in the labor force participation for single women whose five year olds enrolled in kindergarten. But there is no corresponding increase for married mothers, suggesting that targeted expansions towards single-parent families are likely to be more cost effective than universal programs. Work studying the introduction of universal preschool programs in Georgia and Oklahoma finds little evidence that maternal labor supply increased but a large crowd-out of private preschool ([Cascio and Schanzenbach, 2013; Bassok et al., 2014](#)). [Fitzpatrick \(2012\)](#) finds that public preschool does not generally increase the labor supply of mothers, except for single mothers without additional children.

Turning to other countries, [Bauernschuster and Schlotter \(2015\)](#) study the introduction of early kindergarten slots based on day-of birth cutoffs in Germany. Using both instrumental variable and difference-in-difference methods, they find positive impacts on maternal employment. Sharp increases in maternal labor supply are also found for Argentina ([Berlinski et al., 2011](#)),

Spain ([Nollenberger and Rodríguez-Planas, 2015](#)), and Arab mothers in Israel ([Schlosser, 2024](#)). However, using both a regression discontinuity design and a difference-in-difference design for Switzerland, [Gangl and Huber \(2022\)](#) find, if anything, only a moderate impact on mothers' labor market outcomes. In Mexico, the extension of the school day by 3.5 h lead to an increase in mothers' labor supply ([Padilla-Romo and Cabrera-Hernández, 2019](#)).

Family leave is another widely implemented policy which has the potential to increase family labor supply. [Lalive and Zweimüller \(2009\)](#) study a major Austrian leave reform which expanded the duration from one year to two years, and find that employment and earnings decreased in the short run, but not in the long run. They conclude that both job protection guarantees and cash transfers are key elements of family leave policies. [Lalive et al. \(2014\)](#) examine a series of further policy changes in Austria. They find that longer cash benefit periods, particularly during the period with job protection, cause mothers to significantly delay their return to work. Despite this, there is no significant impact of benefit duration or job protection duration on mothers' labor market outcomes in the medium run. Using a structural model and counterfactual policy simulations, the authors conclude that combining cash with job protection is critical for maintaining maternal labor market attachment.

Several studies examine increases in the generosity of paid leave in Germany. [Schönberg and Ludsteck \(2014\)](#) analyze several expansions in maternity leave using a difference-in-difference design around policy reforms governing maternity leave. They find the expansions lowered maternal employment immediately after birth, but had only a small impact on longer-run labor market outcomes. [Kluve and Tamm \(2013\)](#) study a 2007 reform and likewise find a significant decrease in short-run maternal employment post birth, but with a rebound in employment after the transfer ends. [Kluve and Schmitz \(2018\)](#) build on this work using a regression discontinuity design, finding increases in mothers' employment up to 5 years post-birth. The effects are driven by increases in part-time employment, with no effect for full-time employment. Mothers return to their prior employer at a higher rate and have better quality jobs.

In contrast, work for Norway ([Dahl et al., 2016](#)) finds that expansions in paid leave from 18 to 35 weeks via a series of reforms had virtually no impact on labor market participation in either the short or long run. The paid leave expansions had negative redistribution properties; since there was no crowd out of unpaid leave and income was replaced at 100 %, the extra leave amounted to a leisure transfer, primarily to middle and upper income families.

The US is an outlier in that there is no paid leave for new mothers at the national level. However, several states have implemented their own policies, with California being the first to do so in 2004. [Rossin-Slater et al. \(2013\)](#) use Current Population Survey data and find that California's 2004 paid leave program increased the average duration of maternity leave, with some evidence for an increase in work hours and wages when children are between the

ages of 1 to 3. [Byker \(2016\)](#) and [Baum and Ruhm \(2016\)](#) confirm this short-term response for California and [Byker \(2016\)](#) documents a similar short-run impact for New Jersey's program.

However, recent work using administrative Internal Revenue Service (IRS) and Social Security Administration (SSA) data and a regression discontinuity design by [Bailey et al. \(forthcoming\)](#) finds no overall long-run effect mother's employment or earnings from California's 2004 reform. For first-time mothers, they find the program reduces employment and earnings a decade after women give birth. More narrowly, related work using administrative data which focuses on high earners in California finds that wage replacement has no discernable impact on short-term employment ([Bana et al., 2020](#)).

Taken together, the literature on in-kind subsidies yields mixed findings with heterogeneous effects. [Rossin-Slater and Stearns \(2020\)](#) summarize the findings for Europe and North America as follows: leaves shorter than a year often improve job continuity and have no effect on wages, while longer paid leave often harms long-run career advancement. The literature highlights that universal policies can often be costly, in part because there can be substantial crowd out of private or informal care. Similarly, labor supply responses are often limited to certain subgroups, providing support for targeted policies. Of course, programs such as universal pre-school and parental leave have other goals besides increasing labor supply, such as improving child development, and hence a more holistic analysis of such policies is warranted. We briefly discuss the impact on children of such policies in Section 8.

7 Gender inequality

7.1 Rates and trends

While the prior section focused on maternal labor supply, a related question is whether public policies can narrow the gender wage gap after a child is born. As pointed out in a review article by [Goldin and Mitchell \(2017\)](#), women experience a substantial drop in employment and earnings after the birth of a child, but men do not. In an ongoing series of influential papers, Kleven, Landais and a rotating set of coauthors provide compelling empirical evidence on this "child penalty".

A first paper ([Kleven et al., 2019](#)) using Danish panel data reveals that the birth of a child leads to an immediate and long-lasting gender gap in earnings of 20 %. This is due to a combination of less participation in the labor market, fewer work hours, and lower wages. These child penalties have increased over the last 30 years or so. Building on this work, in a second paper they create a world atlas of child penalties using pseudo event studies based on cross-sectional data ([Kleven et al., forthcoming](#)). While the penalties are widespread across the globe, they vary in size, and explain a larger fraction of gender inequality in high wage countries. Fig. 16 shows estimates of the child penalty for various countries from [Kleven et al. \(forthcoming\)](#). The penalties are

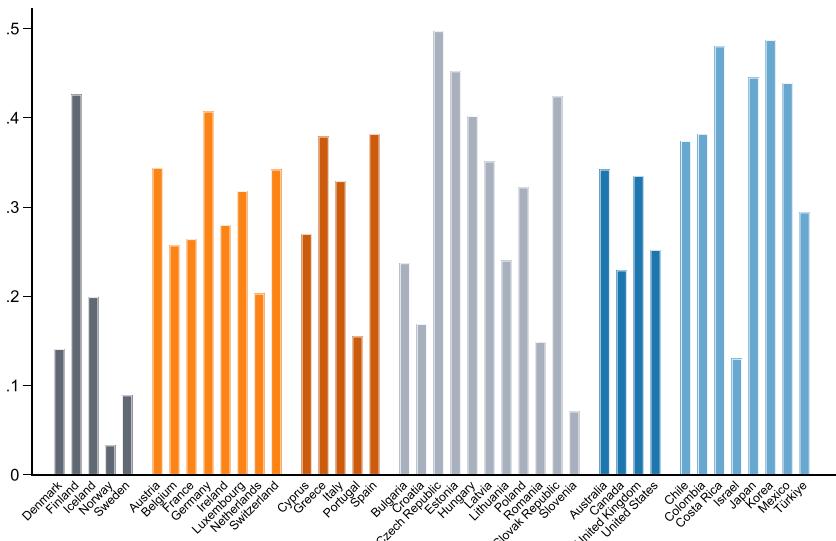


FIG. 16 Average child penalty over years 0–10 after birth of first child. [Notes: The source is Kleven, Landais, and Leite-Mariante, forthcoming. See Fig. 1 for the classification of countries into groups.]

smallest in Norway and highest in the Czech Republic. But the graph reveals substantial variation within country groups; for example, the penalty is relatively large in Finland compared to Norway, and relatively small in Croatia compared to the Czech Republic.

An important question is whether family policies are responsible for these gender inequalities observed in the labor market. Indeed, in many countries, a stated goal of childcare and parental leave programs is in part to reduce gender inequality. Relatedly, can such programs help to explain the general decline in men's labor supply in recent decades?

7.2 Public policies and gender inequality

One policy which explicitly aims to get fathers more involved in child rearing are “daddy quotas”. These paternity leave programs set aside a specific amount of parental leave which can only be taken by the father. The quotas were developed as a way to jump start the leave taking of fathers, since they seldom took any portion of shared leave within a couple. While such policies do increase paternity leave take up (e.g., Dahl et al., 2014; Ekberg et al., 2013), evidence on their effectiveness in reducing gender inequality in the labor market is mixed.

Looking at the introduction of a one month daddy quota in Sweden, Ekberg et al. (2013) find no impact on fathers' versus mothers' long-term wages and employment. Using Norwegian data, one study found that paternity leave

reduces fathers' earnings in Norway but does not impact mothers' earnings ([Rege and Solli, 2013](#)), while another found no effect on fathers' earnings and a reduction in mothers' earnings ([Cools et al., 2015](#)). Studying a Quebec reform which reserved 5 weeks for fathers, [Patnaik \(2019\)](#) finds the policy reduces sex inequality, with mothers spending more time in the paid market-place and men increasing their nonmarket activities at home.

Thinking about shared leave, generous policies could unintentionally increase gender inequality within a household if women become less attached to the labor market as they take longer leaves. There is some evidence for this type of labor supply response for mothers in some countries, such as Germany, as discussed in the prior section. However, other work finds no impact on gender inequality from extensions to shared parental leave. Looking at Norway's expansion of shared leave from 18 to 35 weeks, [Dahl et al. \(2016\)](#) find no effect on mothers' or fathers' earnings or labor force participation in either the short or long run, and hence no impact on gender equality on these dimensions.

Recent work by [Kleven et al. \(2024\)](#) uses high-quality register data from Austria to quasi-experimentally estimate the effects of all family policy reforms over 60 years on men's and women's earnings. They find that while there were large expansions in parental leave and child-care programs, they had almost no impact on gender inequality in the labor market.

In sum, the evidence on paternal labor supply and the gender wage gap does not support the notion that public policies are the main driver of observed gender differences in the labor market. Indeed, [Andresen and Nix \(2022\)](#) rule out both the arrival of a child and fathers' comparative advantage in paid work as mechanisms for the child penalty, arguing that preferences, gender norms, or discrimination must be in play. They reach this conclusion by comparing child penalties in heterosexual nonadopting, adopting, and same-sex couples.

8 Child outcomes

Public policies targeting the labor market activities of mothers and fathers have the potential to affect child outcomes, as this can change their time allocation and family income. This in turn will change the mix of child inputs (parental time investments and market or publicly-provided investments). There is a large and growing literature on how child outcomes are affected by resources, and reviewing it in its entirety is beyond the scope of this chapter. Instead, we focus on public policies which affect the longer-run labor market outcomes of children when they are adults (education and labor market outcomes), highlighting a few of the more prominent and recent studies.

8.1 Public policies and child outcomes

In-kind subsidies. There is a rich literature on how in-kind child care investments affect child development. We focus on universal and large-scale programs that have been evaluated using quasi-experimental methods.

Large-scale programs, especially Head Start in the US and subsidized universal programs in Europe and Canada, have been extensively evaluated. We will only review the more recent studies that are especially relevant for our context of longer run outcomes for children. We refer to [Duncan et al. \(2023\)](#) for an in-depth review of investments in early childhood development in preschool and at home. Their summary is that existing research reaches mixed conclusions, with more work remaining to be done to better understand “when and why the impacts of the home environment and preschool interventions fade out” (p. 1).

The most recent literature on the Head Start program includes papers by [Johnson and Jackson \(2019\)](#) and [Bailey et al. \(2021\)](#). Both use geographic differences in the rollout of the program between 1965 and 1980. Each paper finds positive long-term effects on a set of educational outcomes and increases in labor supply and earnings. [Johnson and Jackson \(2019\)](#) also finds reductions in the likelihood of poverty and incarceration in adulthood. [De Haan and Leuven \(2020\)](#) use a partial identification approach to estimate bounds on treatment effects and also find evidence that Head Start has a positive effect on education and wages. [Kline and Walters \(2016\)](#) underlines the importance of understanding counterfactual care when evaluating early child care interventions. They find that Head Start was only positive for children that transitioned from care at home, while for children previously attending other pre-school programs there is no effect.

[Havnes and Mogstad \(2011b\)](#) was an early paper evaluating a roll-out of universal child care to 3–5 year olds in Norway. They find positive effects on educational attainment at age 30 and in a follow-up paper they find a positive effect on adult income ([Havnes and Mogstad, 2015](#)). The effects are largest at the lower end of the income distribution and even turn negative at the top. Universal child care thus has the potential to level the playing field.

[Cornelissen et al. \(2018\)](#) investigate heterogeneity in the effects of a preschool program in Germany. They find a pattern of reverse selection on gains, with fewer children from disadvantaged backgrounds participating in child care even though the treatment effect for this group was the largest. [Felfe and Lalivé \(2018\)](#) finds similar results when using a reform for younger children in Germany.

For younger children, the evidence is more mixed. [Fort et al. \(2020\)](#) evaluates a center-based program in Italy for ages 1–2 and finds negative effects on IQ and personality for children from more affluent households. [Drange and Havnes \(2019\)](#) finds a positive effect on performance in math and language at age 6–7 of attending child care at ages 1–2, induced by an assignment lottery.

In sum, recent literature on the effects of child care on children’s outcomes show positive long run effects, especially for children from more disadvantaged backgrounds. An important caution is that many evaluations are done in countries with fairly high quality child care. More research needs to be done to understand which quality parameters are essential for a good program. In addition, the counterfactual mode of care is important and key to drawing broader lessons from policy evaluations ([García et al., 2020](#)). Finally, as

[Duncan et al. \(2023\)](#) emphasizes, few studies discuss the difference between average and marginal returns. While child care subsidies might raise participation for marginal groups, it also provides a cash transfer to families that would have used childcare regardless. Such deadweight losses should be taken into account when evaluating universal child care policies.

Turning to parental leave, there are a handful of studies which look at the long-run outcomes of children. [Dustmann and Schönberg \(2012\)](#) find no long-term effects on children of extending maternal leave in Germany. [Rasmussen \(2010\)](#) also finds no effect of extending parental leave in Denmark on children's long run outcomes. On the other hand, [Carneiro et al. \(2015\)](#) shows that the introduction of paid parental leave in Norway had some positive effects on children's future education and income. However, [Dahl et al. \(2016\)](#) shows that extending the duration beyond the initial few months in Norway provides no additional improvement in children's test scores.

Other papers likewise find mixed results. [Liu and Skans \(2010\)](#) finds positive effects on test scores in Sweden, but only for children of well-educated mothers. [Danzer and Lavy \(2018\)](#) finds no average effect on school outcomes at age 15 in Austria, but positive effects for certain subgroups. [Ginja et al. \(2020\)](#) finds positive effects for the older child but not the younger, due to the "speed premium" in Sweden which gives mothers higher benefits for a subsequent child if births are closely spaced.

[Bütkofer et al. \(2021\)](#) shows that the initial introduction of paid leave in Norway which [Carneiro et al. \(2015\)](#) found to increase long-term child outcomes also had positive effects on maternal health. This provides a possible mechanism for why studies looking at subsequent expansions in Norway do not find long-run effects on children (i.e., there were no further improvements to maternal health from subsequent leave expansions).

More generally, the pattern of mixed results across studies could be due to heterogeneous effects. For example, some studies find larger long-run effects for children with more educated mothers. Heterogeneity could arise through income effects, counterfactual care arrangements, differential labor supply, or varying fertility responses. We conclude there is more work to be done on how to best target in-kind transfers if long-run improvements in child outcomes are a policy goal.

Cash and near-cash transfers. While there is a sizable literature on the effects of cash transfers on children in the short term, there is less research for the longer run. Even so, it is beyond the scope of this chapter to review how money matters in general for children's long-term labor market outcomes. Here we review the findings for only a handful of studies as a way to illustrate the potential for programs which are targeted to increase parental labor supply to have long-term spillover effects on their children. Taking the example of the EITC, [Dahl and Lochner \(2012\)](#) examines how changes in family income from EITC expansions in the US affects children's academic performance. They find positive effects on math and reading test scores, with larger effects for younger

children, boys, and those from more disadvantaged backgrounds. [Milligan and Stabile \(2011\)](#) find similar benefits in academic achievement in their study of the Canadian Child Benefit Expansions. Building on this work, [Bastian and Michelmore \(2018\)](#) estimate the long-term impact of the EITC on children. They find that an extra \$1000 in EITC income increases the chances of graduating from high school (by 1.3 %) and college (4.2 %), as well as increasing young adult employment (1.0 %) and earnings (2.2 %).

As a second example, [Black et al. \(2014\)](#) studies a child care subsidy reform in Norway that increased disposable income of families, holding child care attendance fixed. This increased disposable income has positive effects on child test scores at age 16. These two examples highlight that cash transfers targeted to low income families have the potential to benefit children in the longer run.

9 Norms and spillovers

Public policies can also directly influence societal norms regarding family outcomes such as fertility, marriage, and labor supply. Although relatively limited, a growing body of literature examines how public policies impact program uptake and shape norms in these domains. We anticipate significant growth in this area, driven by the increased availability of data. A promising data advance is the merging of register data with survey data which provides more nuanced measures of societal norms.

9.1 Public policies and norms and spillovers

Several papers have examined whether participation in government programs targeting families can spread more broadly through society via peer effects. [Dahl et al. \(2014\)](#) estimate spillover effects of program participation in paternal leave. Coworkers and brothers are 11 % and 15 % points, respectively, more likely to take paternity leave if their peer was exogenously induced to take up leave. [Diallo and Lange \(2023\)](#) replicate the findings and find similar patterns using a Canadian reform in paternal leave. Another study by [Welteke and Wrohlich \(2019\)](#) for Germany also finds spillover effects of changes in maternal leave policies. Taking into account spillover effects of public policies leads to long-run participation rates which are substantially higher than would otherwise be expected from analyzing only the direct effect on the targeted population.

[Agostinelli et al. \(forthcoming\)](#) study spillover effects of a large-scale randomized control trial to help parents in Chicago and document large spillover effects on both treatment and control children who live near treated children. Their findings underscore the interaction of parenting with neighborhood and peer effects when evaluating the cost-benefit of programs.

[Ichino et al. \(2022\)](#) exploit variation from Swedish tax reforms regarding the use of parental childcare, contrasting effects for native and immigrant

couples from a variety of countries characterized by different gender norms. They find that couples originating from countries with relatively conservative norms are more likely to reallocate childcare to mothers following a reduction in the father's tax rate, and less likely to reallocate childcare to fathers following a reduction in the mother's tax rate. This reinforces the traditional allocation of childcare across these conservative-normed parents.

[Doepe and Kindermann \(2019\)](#) build a quantitative model of household bargaining in which the distribution of the burden of child care between mothers and fathers is a key determinant of fertility. They show that fertility is responsive to targeted policies that lower the child care burden specifically for mothers. This could be an explanation for the heterogenous fertility effects of public policies across countries.

In a recent working paper, [Fontenay and González \(2024\)](#) examine how paternity leave policies affect gender role attitudes of the next generation in six countries. Using an RD design, they find that male children whose parents were exposed to a paternity leave reform have less gender-stereotypical attitudes when they grow up and are more likely to choose female-stereotypical occupations such as healthcare and education.

Although research on norms is still in an early phase, papers on spillover effects of public policies consistently find sizeable effects, especially for close peers. Accurate evaluations of the benefits and costs of public programs need to take these spillovers into account, as they can often be nontrivial. Existing culture and norms in a country are also an important factors to consider when reforming public policies. While there is much to learn from country-specific studies, there are also large differences in institutions and norms that affect the generalizability of these early findings.

10 Lessons learned and avenues for future research

10.1 Short summary

In this chapter, we have documented significant changes in fertility, marriage, divorce, cohabitation, maternal employment, and gender inequality across OECD countries from 2001 to 2019. Analyzing these changes in relation to various public policies yields two main insights.

In the short term, while public policies can notably impact certain family outcomes, the overall pattern across the studies we have examined suggests that these policies are unlikely to reverse the trends of declining fertility and marriage rates or significantly increase maternal labor supply. Additionally, many of the public policies discussed entail considerable program costs and redistribution concerns, which limit their feasibility given relatively small benefits. However, a combination of policies, such as integrating tax incentives with targeted subsidies for childcare, shows more promise on the benefit side. This also suggests conducting broader cost-benefit analyses which consider both total benefits and costs for the whole system and not only single policies in isolation.

In the long term, there is a greater potential for impact. Changes in family norms around child-rearing and spillover effects to groups closely connected to targeted families, including the next generation, provide a stronger case for policy intervention. However, altering norms takes considerably longer and is more challenging to identify with commonly used research methods and data so the evidence base is still rather limited.

10.2 Avenues for future research

A particularly concerning trend is the significant decline in fertility rates around the globe, which has proven difficult to explain (e.g., [Kearney et al. \(2022\)](#)). Future research should explore family decision-making processes and how to design policies that better support the joint decisions of dual-working parents and fertility. [Doepke et al. \(2023\)](#) provides a valuable starting point for discussions on fertility and potential policies.

Second, the literature to date has predominantly focused on married families or single-parent households. However, contemporary families take various forms, including cohabiting couples, blended families, multi-generational households, and same-sex couples. An open avenue for research is to look into how general models and insights can be adapted to account for the dynamic nature of family structures.

Third, classic public policies need to be reconsidered in light of an evolving labor market landscape. Policies aimed at increasing female labor supply and fertility must account for changes in the modern labor market. For instance, new policies should address the implications of emerging technologies, which may differ for couples with and without children. [Goldin \(2014\)](#) suggests that structural changes to the labor market, rather than existing public policies, might be a more fruitful avenue for rethinking how interactions between families and labor markets will evolve. One such change is the increase in flexible work arrangements, especially the rise in remote work prompted by the COVID-19 pandemic. Further research is needed to understand the implications of these changes for families.

References

- [Adda, J., Dustmann, C., Stevens, K., 2017. The career costs of children. *J. Political Econ.* 125 \(2\), 293–337.](#)
- [Agostinelli, F., Doepke, M., Sorrenti, G., Zilibotti, F., forthcoming. It takes a village: the economics of parenting with neighborhood and peer effects. *J. Political Econ.*](#)
- [Aizer, A., Cho, S., Eli, S., Lleras-Muney, A., 2024. The impact of cash transfers to poor mothers on family structure and maternal well-being. *Am. Econ. J.: Appl. Econ.* 16 \(2\), 492–529.](#)
- [Aizer, A., Hoynes, H., Lleras-Muney, A., 2022. Children and the US social safety net: balancing disincentives for adults and benefits for children. *J. Econ. Perspect.* 36 \(2\), 149–174.](#)
- [Alesina, A., Ichino, A., Karabarbounis, L., 2011. Gender-based taxation and the division of family chores. *Am. Econ. J.: Econ. Policy* 3 \(2\), 1–40.](#)

- Alm, J., Dickert-Conlin, S., Whittington, L.A., 1999. Policy watch. *J. Econ. Perspect.* 13 (3), 193–204.
- Alm, J., Leguizamon, J.S., Leguizamon, S., 2023. Race, ethnicity, and taxation of the family: the many shades of the marriage penalty/bonus. *Natl Tax. J.* 76 (3), 525–560.
- Andresen, M.E., Nix, E., 2022. What causes the child penalty? Evidence from adopting and same-sex couples. *J. Labor. Econ.* 40 (4), 971–1004.
- Ang, X.L., 2015. The effects of cash transfer fertility incentives and parental leave benefits on fertility and labor supply: evidence from two natural experiments. *J. Family Econ. Issues* 36, 263–288.
- Ashenfelter, O., Heckman, J., 1974. The estimation of income and substitution effects in a model of family labor supply. *Econometrica* 42 (1), 73–85.
- Avdic, D., Karimi, A., 2018. Modern family? Paternity leave and marital stability. *Am. Econ. J.: Appl. Econ.* 10 (4), 283–307.
- Bailey, M.J., Byker, T., Patel, E., Ramnath, S., forthcoming. The long-run effects of California's paid family leave act on women's careers and childbearing: new evidence from a regression discontinuity design and US tax data. *Am. Econ. J.: Econ. Policy*.
- Bailey, M.J., Sun, S., Timpe, B., 2021. Prep school for poor kids: the long-run impacts of Head Start on human capital and economic self-sufficiency. *Am. Econ. Rev.* 111 (12), 3963–4001.
- Baker, M., Gruber, J., Milligan, K., 2008. Universal child care, maternal labor supply, and family well-being. *J. Political Econ.* 116 (4), 709–745.
- Baker, M., Hanna, E., Kantarevic, J., 2004. The married widow: marriage penalties matter!. *J. Eur. Econ. Assoc.* 2 (4), 634–664.
- Bana, S.H., Bedard, K., Rossin-Slater, M., 2020. The impacts of paid family leave benefits: regression kink evidence from California administrative data. *J. Policy Anal. Manag.* 39 (4), 888–929.
- Bartel, A.P., Rossin-Slater, M., Ruhm, C.J., Stearns, J., Waldfogel, J., 2018. Paid family leave, fathers' leave-taking, and leave-sharing in dual-earner households. *J. Policy Anal. Manag.* 37 (1), 10–37.
- Bassford, M., Fisher, H., 2020. The impact of paid parental leave on fertility intentions. *Econ. Rec.* 96 (315), 402–430.
- Bassok, D., Fitzpatrick, M., Loeb, S., 2014. Does state preschool crowd-out private provision? The impact of universal preschool on the childcare sector in Oklahoma and Georgia. *J. Urban. Econ.* 83, 18–33.
- Bastian, J., 2017. Unintended consequences? More marriage, more children, and the EITC. *Proc. Annu. Conf.* 110, 1–56.
- Bastian, J., Lochner, L., 2022. The earned income tax credit and maternal time use: more time working and less time with kids? *J. Labor. Econ.* 40 (3), 573–611.
- Bastian, J., Michelmore, K., 2018. The long-term impact of the earned income tax credit on children's education and employment outcomes. *J. Labor. Econ.* 36 (4), 1127–1163.
- Bauernschuster, S., Hener, T., Rainer, H., 2016. Children of a (policy) revolution: the introduction of universal child care and its effect on fertility. *J. Eur. Econ. Assoc.* 14 (4), 975–1005.
- Bauernschuster, S., Schlotter, M., 2015. Public child care and mothers' labor supply: evidence from two quasi-experiments. *J. Public. Econ.* 123, 1–16.
- Baum, C.L., Ruhm, C.J., 2016. The effects of paid family leave in California on labor market outcomes. *J. Policy Anal. Manag.* 35 (2), 333–356.
- Becker, G.S., 1973. A theory of marriage: Part I. *J. Political Econ.* 81 (4), 813–846.
- Berlinski, S., Galiani, S., Mc Ewan, P.J., 2011. Preschool and maternal labor market outcomes: evidence from a regression discontinuity design. *Econ. Dev. Cultural Change* 59 (2), 313–344.

- Bick, A., Fuchs-Schündeln, N., 2018. Taxation and labour supply of married couples across countries: a macroeconomic analysis. *Rev. Econ. Stud.* 85 (3), 1543–1576.
- Black, S.E., Devereux, P.J., Løken, K.V., Salvanes, K.G., 2014. Care or cash? The effect of child care subsidies on student performance. *Rev. Econ. Stat.* 96 (5), 824–837.
- Blundell, R., Duncan, A., McCrae, J., Meghir, C., 2000. The labour market impact of the working families' tax credit. *Fisc. Stud.* 21 (1), 75–104.
- Borella, M., De Nardi, M., Yang, F., 2023. Are marriage-related taxes and social security benefits holding back female labour supply? *Rev. Econ. Stud.* 90 (1), 102–131.
- Brewer, M., Francesconi, M., Gregg, P., Grogger, J., 2009. Feature: in-work benefit reform in a cross-national perspective-introduction. *Econ. J.* 119 (535), F1–F14.
- Brilli, Y., Del Boca, D., Pronzato, C.D., 2016. Does child care availability play a role in maternal employment and children's development? Evidence from Italy. *Rev. Econ. Househ.* 14, 27–51.
- Browning, M., Chiappori, P.-A., 1998. Efficient intra-household allocations: a general characterization and empirical tests. *Econometrica* 1241–1278.
- Bütikofer, A., Riise, J., Skira, M.M., 2021. The impact of paid maternity leave on maternal health. *Am. Econ. J.: Econ. Policy* 13 (1), 67–105.
- Byker, T.S., 2016. Paid parental leave laws in the United States: does short-duration leave affect women's labor-force attachment? *Am. Econ. Rev.* 106 (5), 242–246.
- Cannonier, C., 2014. Does the family and medical leave act (FMLA) increase fertility behavior? *J. Labor. Res.* 35, 105–132.
- Carneiro, P., Løken, K.V., Salvanes, K.G., 2015. A flying start? Maternity leave benefits and long-run outcomes of children. *J. Political Econ.* 123 (2), 365–412.
- Cascio, E.U., 2009. Maternal labor supply and the introduction of kindergartens into American public schools. *J. Hum. Resour.* 44 (1), 140–170.
- Cascio, E.U., Schanzenbach, D.W., 2013. The impacts of expanding access to high-quality pre-school. *Brook. Pap. Economic Act.* 2013 (Fall), 127–178.
- Chiappori, P.-A., 1992. Collective labor supply and welfare. *J. Political Econ.* 100 (3), 437–467.
- Cohen, A., Dehejia, R., Romanov, D., 2013. Financial incentives and fertility. *Rev. Econ. Stat.* 95 (1), 1–20.
- Cools, S., Fiva, J.H., Kirkebøen, L.J., 2015. Causal effects of paternity leave on children and parents. *Scand. J. Econ.* 117 (3), 801–828.
- Cornelissen, T., Dustmann, C., Raute, A., Schönberg, U., 2018. Who benefits from universal child care? Estimating marginal returns to early child care attendance. *J. Political Econ.* 126 (6), 2356–2409.
- Cygan-Rehm, K., 2016. Parental leave benefit and differential fertility responses: evidence from a German reform. *J. Popul. Econ.* 29, 73–103.
- Cygan-Rehm, K., Kuehnle, D., Riphahn, R.T., 2018. Paid parental leave and families' living arrangements. *Labour Econ.* 53, 182–197.
- Dachille, G.P., De Paola, M., Nisticò, R., Tridico, P., 2023. Guaranteed minimum income and fertility.
- Dahl, G.B., Lochner, L., 2012. The impact of family income on child achievement: evidence from the earned income tax credit. *Am. Econ. Rev.* 102 (5), 1927–1956.
- Dahl, G.B., Løken, K.V., Mogstad, M., 2014. Peer effects in program participation. *Am. Econ. Rev.* 104 (7), 2049–2074.
- Dahl, G.B., Løken, K.V., Mogstad, M., Salvanes, K.V., 2016. What is the case for paid maternity leave? *Rev. Econ. Stat.* 98 (4), 655–670.
- Danzer, N., Lavy, V., 2018. Paid parental leave and children's schooling outcomes. *Econ. J.* 128 (608), 81–117.

- De Haan, M., Leuven, E., 2020. Head start and the distribution of long-term education and labor market outcomes. *J. Labor. Econ.* 38 (3), 727–765.
- Del Boca, D., Pasqa, S., Pronzato, C., 2008. Motherhood and market work decisions in institutional context: a European perspective. *Oxf. Econ. Pap.* 61 (suppl_1), i147–i171.
- Diallo, Y., Lange, F., 2023. Peer effects at work on parental leave.
- Doepke, M., Hannusch, A., Kindermann, F., Tertilt, M., 2023. The economics of fertility: a new era. *Handbook of the Economics of the Family*, vol. 1. Elsevier, pp. 151–254.
- Doepke, M., Kindermann, F., 2019. Bargaining over babies: theory, evidence, and policy implications. *Am. Econ. Rev.* 109 (9), 3264–3306.
- Drange, N., Havnes, T., 2019. Early childcare and cognitive development: evidence from an assignment lottery. *J. Labor. Econ.* 37 (2), 581–620.
- Duncan, G., Kalil, A., Mogstad, M., Rege, M., 2023. Investing in early childhood development in preschool and at home. *Handb. Econ. Educ.* 6, 1–91.
- Dustmann, C., Schönberg, U., 2012. Expansions in maternity leave coverage and children's long-term outcomes. *Am. Econ. J.: Appl. Econ.* 4 (3), 190–224.
- Eissa, N., Hoynes, H.W., 2004. Taxes and the labor market participation of married couples: the earned income tax credit. *J. Public. Econ.* 88 (9–10), 1931–1958.
- Eissa, N., Liebman, J.B., 1996. Labor supply response to the earned income tax credit. *Q. J. Econ.* 111 (2), 605–637.
- Ekberg, J., Eriksson, R., Friebel, G., 2013. Parental leave—a policy evaluation of the Swedish "daddy-month" reform. *J. Public. Econ.* 97, 131–143.
- Elmallakh, N., 2023. Fertility and labor supply responses to child allowances: the introduction of means-tested benefits in france. *Demography* 60 (5), 1493–1522.
- Farré, L., González, L., 2019. Does paternity leave reduce fertility? *J. Public. Econ.* 172, 52–66.
- Felfe, C., Lalivé, R., 2018. Does early child care affect children's development? *J. Public. Econ.* 159, 33–53.
- Fink, A., 2020. German income taxation and the timing of marriage. *Appl. Econ.* 52 (5), 475–489.
- Fisher, H., 2013. The effect of marriage tax penalties and subsidies on marital status. *Fisc. Stud.* 34 (4), 437–465.
- Fitzpatrick, M.D., 2012. Revising our thinking about the relationship between maternal labor supply and preschool. *J. Hum. Resour.* 47 (3), 583–612.
- Fontenay, S., González, L., 2024. Can public policies break the gender mold? Evidence from paternity leave reforms in six countries. *Barcelona School of Economics Working Paper*, (1422).
- Fort, M., Ichino, A., Zanella, G., 2020. Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families. *J. Political Econ.* 128 (1), 158–205.
- Fortin, B., Lacroix, G., 1997. A test of the unitary and collective models of household labour supply. *Econ. J.* 107 (443), 933–955.
- Francesconi, M., Van Der Klaauw, W., 2007. The socioeconomic consequences of "in-work" benefit reform for British lone mothers. *J. Hum. Resour.* 42 (1), 1–31.
- Frazier, N., McKeehan, M., 2018. Hesitating at the altar: an update on taxes and the timing of marriage. *Public. Financ. Rev.* 46 (5), 743–763.
- Friedberg, L., Isaac, E., 2024. Same-sex marriage recognition and taxes: new evidence about the impact of household taxation. *Rev. Econ. Stat.* 106 (1), 85–101.
- Gangl, S., Huber, M., 2022. From homemakers to breadwinners? How mandatory kindergarten affects maternal labour market outcomes.
- García, J.L., Heckman, J.J., Leaf, D.E., Prados, M.J., 2020. Quantifying the life-cycle benefits of an influential early-childhood program. *J. Political Econ.* 128 (7), 2502–2541.

- Gayle, G.-L., Shephard, A., 2019. Optimal taxation, marriage, home production, and family labor supply. *Econometrica* 87 (1), 291–326.
- Ginja, R., Jans, J., Karimi, A., 2020. Parental leave benefits, household labor supply, and children's long-run outcomes. *J. Labor. Econ.* 38 (1), 261–320.
- Giupponi, G., 2019. When income effects are large: Labor supply responses and the value of welfare transfers.
- Givord, P., Marbot, C., 2015. Does the cost of child care affect female labor market participation? An evaluation of a French reform of childcare subsidies. *Labour Econ.* 36, 99–111.
- Goldin, C., 2014. A grand gender convergence: Its last chapter. *Am. Econ. Rev.* 104 (4), 1091–1119.
- Goldin, C., Mitchell, J., 2017. The new life cycle of women's employment: disappearing humps, sagging middles, expanding tops. *J. Econ. Perspect.* 31 (1), 161–182.
- González, L., Trommlerová, S.K., 2023. Cash transfers and fertility: how the introduction and cancellation of a child benefit affected births and abortions. *J. Hum. Resour.* 58 (3), 783–818.
- Gregg, P., Harkness, S., Smith, S., 2009. Welfare reform and lone parents in the UK. *Econ. J.* 119 (535), F38–F65.
- Groneck, M., Wallenius, J., 2021. It sucks to be single! Marital status and redistribution of social security. *Econ. J.* 131 (633), 327–371.
- Haan, P., Wrohlich, K., 2011. Can child care policy encourage employment and fertility? Evidence from a structural model. *Labour Econ.* 18 (4), 498–512.
- Halla, M., Lackner, M., Scharler, J., 2016. Does the welfare state destroy the family? Evidence from OECD member countries. *Scand. J. Econ.* 118 (2), 292–323.
- Hart, R.K., Andersen, S.N., Drange, N., 2022. Effects of extended paternity leave on family dynamics. *J. Marriage Family* 84 (3), 814–839.
- Havnes, T., Mogstad, M., 2011a. Money for nothing? Universal child care and maternal employment. *J. Public. Econ.* 95 (11–12), 1455–1465.
- Havnes, T., Mogstad, M., 2011b. No child left behind: subsidized child care and children's long-run outcomes. *Am. Econ. J.: Econ. Policy* 3 (2), 97–129.
- Havnes, T., Mogstad, M., 2015. Is universal child care leveling the playing field? *J. Public. Econ.* 127, 100–114.
- Hoynes, H.W., Schanzenbach, D.W., 2012. Work incentives and the food stamp program. *J. Public. Econ.* 96 (1-2), 151–162.
- Ichino, A., Olsson, M., Petrongolo, B., Thoursie, P.S., 2022. Taxes, childcare, and gender identity norms.
- Isaac, E., 2023. Suddenly married: joint taxation and the labor supply of same-sex married couples after us v. windsor. *J. Hum. Resour.*
- Johnson, R.C., Jackson, C.K., 2019. Reducing inequality through dynamic complementarity: evidence from Head Start and public school spending. *Am. Econ. J.: Econ. Policy* 11 (4), 310–349.
- Kearney, M.S., Levine, P.B., Pardue, L., 2022. The puzzle of falling us birth rates since the great recession. *J. Econ. Perspect.* 36 (1), 151–176.
- Kleven, H., Landais, C., Leite-Mariante, G., forthcoming. The child penalty atlas. *Rev. Econ. Studies*.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A., Zweimüller, J., 2024. Do family policies reduce gender inequality? Evidence from 60 years of policy experimentation. *Am. Econ. J.: Econ. Policy* 16 (2), 110–149.
- Kleven, H., Landais, C., Søgaard, J.E., 2019. Children and gender inequality: evidence from Denmark. *Am. Econ. J.: Appl. Econ.* 11 (4), 181–209.

- Kleven, H.J., Kreiner, C.T., Saez, E., 2009. The optimal income taxation of couples. *Econometrica* 77 (2), 537–560.
- Kline, P., Walters, C.R., 2016. Evaluating public programs with close substitutes: the case of Head Start. *Q. J. Econ.* 131 (4), 1795–1848.
- Kluve, J., Schmitz, S., 2018. Back to work: parental benefits and mothers' labor market outcomes in the medium run. *ILR Rev.* 71 (1), 143–173.
- Kluve, J., Tamm, M., 2013. Parental leave regulations, mothers' labor force attachment and fathers' childcare involvement: evidence from a natural experiment. *J. Popul. Econ.* 26, 983–1005.
- Kotsadam, A., Finseraas, H., 2011. The state intervenes in the battle of the sexes: causal effects of paternity leave. *Soc. Sci. Res.* 40 (6), 1611–1622.
- Kuka, E., Shenhav, N., 2024. Long-run effects of incentivizing work after childbirth. *Am. Econ. Rev.* 114 (6), 1692–1722.
- Lalive, R., Schlosser, A., Steinhauer, A., Zweimüller, J., 2014. Parental leave and mothers' careers: the relative importance of job protection and cash benefits. *Rev. Econ. Stud.* 81 (1), 219–265.
- Lalive, R., Zweimüller, J., 2009. How does parental leave affect fertility and return to work? Evidence from two natural experiments. *Q. J. Econ.* 124 (3), 1363–1402.
- Laroque, G., Salanié, B., 2014. Identifying the response of fertility to financial incentives. *J. Appl. Econ.* 29 (2), 314–332.
- Lefebvre, P., Merrigan, P., 2008. Child-care policy and the labor supply of mothers with young children: a natural experiment from canada. *J. Labor. Econ.* 26 (3), 519–548.
- Liu, Q., Skans, O.N., 2010. The duration of paid parental leave and children's scholastic performance. *BE J. Econ. Anal. Policy* 10 (1).
- Lundberg, S., Pollak, R.A., 1993. Separate spheres bargaining and the marriage market. *J. Political Econ.* 101 (6), 988–1010.
- Meyer, B.D., Rosenbaum, D.T., 2001. Welfare, the earned income tax credit, and the labor supply of single mothers. *Q. J. Econ.* 116 (3), 1063–1114.
- Milligan, K., 2005. Subsidizing the stork: new evidence on tax incentives and fertility. *Rev. Econ. Stat.* 87 (3), 539–555.
- Milligan, K., Stabile, M., 2011. Do child tax benefits affect the well-being of children? Evidence from Canadian child benefit expansions. *Am. Econ. J.: Econ. Policy* 3 (3), 175–205.
- Mincer, J., 1962. Aspects of labor markets (chap). *Labor Force Participation of Married Women*. Princeton University Press, pp. 63–97 (chap).
- Moffitt, R., 1992. Incentive effects of the US welfare system: a review. *J. Econ. Lit.* 30 (1), 1–61.
- Moffitt, R.A., 1998. *The Effect of Welfare on Marriage and Fertility*. National Academy Press, Washington, DC.
- Moffitt, R.A., Phelan, B.J., Winkler, A.E., 2020. Welfare rules, incentives, and family structure. *J. Hum. Resour.* 55 (1), 1–42.
- Mörk, E., Sjögren, A., Svaleryd, H., 2013. Childcare costs and the demand for children – evidence from a nationwide reform. *J. Popul. Econ.* 26, 33–65.
- Morrissey, T.W., 2017. Child care and parent labor force participation: a review of the research literature. *Rev. Econ. Househ.* 15 (1), 1–24.
- Myohl, N., 2024. Till taxes keep us apart? The impact of the marriage tax on the marriage rate. *Int. Tax. Public. Financ.* 31 (2), 552–592.
- Nichols, A., Rothstein, J., 2015. The earned income tax credit. *Economics of Means-Tested Transfer Programs in the United States* 1. University of Chicago Press, pp. 137–218.
- Nollenberger, N., Rodríguez-Planas, N., 2015. Full-time universal childcare in a context of low maternal employment: quasi-experimental evidence from Spain. *Labour Econ.* 36, 124–136.

- Olafsson, A., Steingrimsdottir, H., 2020. How does daddy at home affect marital stability? *Econ. J.* 130 (629), 1471–1500.
- Padilla-Romo, M., Cabrera-Hernández, F., 2019. Easing the constraints of motherhood: The effects of all-day schools on mothers' labor supply. *Econ. Inq.* 57 (2), 890–909.
- Padilla-Romo, M., Peluffo, C., Viollaz, M., 2022. Parents' effective time endowment and divorce: evidence from extended school days.
- Patnaik, A., 2019. Reserving time for daddy: the consequences of fathers' quotas. *J. Labor. Econ.* 37 (4), 1009–1059.
- Persson, P., 2020. Social insurance and the marriage market. *J. Political Econ.* 128 (1), 252–300.
- Rasmussen, A.W., 2010. Increasing the length of parents' birth-related leave: the effect on children's long-term educational outcomes. *Labour Econ.* 17 (1), 91–100.
- Raute, A., 2019. Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits. *J. Public. Econ.* 169, 203–222.
- Reader, M., Portes, J., Patrick, R., 2022. Does cutting child benefits reduce fertility in larger families? Evidence from the UK's two-child limit.
- Rege, M., Solli, I.F., 2013. The impact of paternity leave on fathers' future earnings. *Demography* 50 (6), 2255–2277.
- Riphahn, R.T., Wiynck, F., 2017. Fertility effects of child benefits. *J. Popul. Econ.* 30, 1135–1184.
- Rossin-Slater, M., Ruhm, C.J., Waldfogel, J., 2013. The effects of California's paid family leave program on mothers' leave-taking and subsequent labor market outcomes. *J. Policy Anal. Manag.* 32 (2), 224–245.
- Rossin-Slater, M., Stearns, J., 2020. Time on with baby and time off from work. *Future Child.* 30 (2), 35–52.
- Sandner, M., Wiynck, F., 2023. The fertility response to cutting child-related welfare benefits. *Popul. Res. Policy Rev.* 42 (2), 25.
- Schlosser, A., 2024. Public preschool and the labor supply of arab mothers: evidence from a natural experiment. CESifo Working Paper No. 10904.
- Schmitz, S., 2020. The impact of publicly funded childcare on parental well-being: evidence from cut-off rules. *Eur. J. Popul.* 36 (2), 171–196.
- Schönberg, U., Ludsteck, J., 2014. Expansions in maternity leave coverage and mothers' labor market outcomes after childbirth. *J. Labor. Econ.* 32 (3), 469–505.
- Tekin, E., 2007. Childcare subsidies, wages, and employment of single mothers. *J. Hum. Resour.* 42 (2), 453–487.
- Thomas, J., 2024. Free, full-day programming for four-year-old children in Nova Scotia and women's labour market outcomes. *Can. J. Econ./Revue Canadienne d'économique*.
- Wang, H., 2022. Fertility and family leave policies in Germany: optimal policy design in a dynamic framework. Tech. Rep., Working paper.
- Welteke, C., Wrohlich, K., 2019. Peer effects in parental leave decisions. *Labour Econ.* 57, 146–163.
- Wood, J., Neels, K., 2019. Local childcare availability and dual-earner fertility: variation in childcare coverage and birth hazards over place and time. *Eur. J. Popul.* 35 (5), 913–937.
- Yonzan, N., Timilsina, L., Kelly, I.R., 2024. Economic incentives surrounding fertility: evidence from alaska's permanent fund dividend. *Econ. Hum. Biol.* 52, 101334.
- Ziliak, J.P., 2015. Temporary assistance for needy families. *Economics of Means-Tested Transfer Programs in the United States* 1. University of Chicago Press, pp. 303–393.

This page intentionally left blank

Chapter 8

The evolution of gender in the labor market[☆]

Claudia Olivetti^{a,b,*}, Jessica Pan^{c,d}, and Barbara Petrongolo^{e,f,g}

^aDartmouth College, NH, United States, ^bNational Bureau of Economic Research, MA, United States, ^cNational University of Singapore, Singapore, ^dUniversity of Bonn, IZA, Bonn, Germany, ^eOxford University, Oxford, United Kingdom, ^fThe Centre for Economic Performance, London School of Economics and Political Science, London, United States, ^gCentre for Economic Policy Research, London, United Kingdom

*Corresponding author. e-mail address: Claudia.Olivetti@dartmouth.edu

Chapter Outline

1 Introduction	619	5.3 Monopsonistic labor markets	649
2 Real world and academic developments in gender inequality	624	6 The role of identity and norms in understanding gender inequalities	652
3 Women's labor supply and the gender gap	630	6.1 Relevance for labor supply and household specialization	653
3.1 The labor supply of the secondary earner	631	6.2 Stereotypes, beliefs, and discrimination	655
4 Evolving perspectives on gender inequality	635	6.3 What drives gender norms and how malleable are they?	658
4.1 Preferences, traits, and constraints	636	7 Micro-macro linkages	664
4.2 Career-family trade-offs	639	8 Conclusion	665
5 The anatomy of the career costs of motherhood	644	Appendix A. Gender in economic journals	667
5.1 Gender biology and productivity	644	References	669
5.2 Differential job sorting and the organization of work	646		

1 Introduction

The remarkable progress of women in the labor market marks one of the most significant economic and social changes of the past half a century. Accompanying these developments has been a large increase in interest in

☆ We thank the editors Christian Dustmann and Thomas Lemieux and seminar participants at the 2023 Handbook Conference in Berlin (generously funded by Rockwool Foundation Berlin (RFBerlin)) for very helpful comments and suggestions. We are grateful to Dorothy Ting and Teh Renjie for excellent research assistance.

gender topics in the economics profession since the 1990s, culminating in the award of the 2023 Nobel Prize in Economics to Claudia Goldin for her pioneering work on understanding women's labor market outcomes through the centuries.

While understanding women's outcomes in the labor market is an important topic of inquiry in its own right, the study of gender in itself has significantly contributed to modern labor economics more generally. As Claudia Goldin remarks in her 2006 American Economic Association (AEA) Presidential Address:

It would not be much of an exaggeration to claim that women gave "birth" to modern labor economics, especially labor supply. Economists need variance to analyze changes in behavioral responses, and women provided an abundance of that. Men, by and large, were not as interesting, since their participation and hours varied far less in cross section and over time. (p. 3)

[Fig. 1](#) illustrates how interest in gender topics within the economics profession has changed over time by comparing the share of papers in the top 30 economics journals that are on gender-related topics versus race-related topics.¹ We classify paper topics based on keywords pertaining to gender (i.e., female, women, gender) or race (black, ethnic, Hispanic, race) in the title. In addition, we include a more expansive set of keywords (wife, maternity, mother, girl) to identify gender-related topics. While the share of race-related papers in the top 30 economics journals has remained relatively constant at about 1 % from the 1970s to the 2020s, the share of gender-related papers has increased steadily over the period, from around 0.8 % in the 1970s to about 1.8 % in the 2020s. The patterns are even more striking when we use the more inclusive set of keywords to identify gender-related topics.²

The study of gender has also expanded from fairly niche topics in labor and family economics to other fields in economics. [Fig. 2](#) further classifies gender-related papers in the top 30 economics journals into various sub-topics based on keywords in the title. Not surprisingly, most gender papers are about the labor market or family related topics. Nevertheless, between the 1970s/1980s and the 2010/2020s, the share of gender papers relating to traditional labor-

¹ The list of top 30 economics journals used for this exercise can be found in Appendix [Table A.1](#).

² We acknowledge that classifying papers based only on their titles is likely to lead to an undercount of the number of race and gender papers. We explored the issue using a similar keyword search procedure applied to paper abstracts; however, this tends to lead to an overcount since many papers that are not primarily about race or gender report dimensions of heterogeneity on the basis of these characteristics. We use data on assigned ``subjects'' to papers available from EBSCO, a leading provider of research databases and e-journals, to show that the patterns described above hold when papers are classified using subject keywords instead of titles. The results from this exercise are reported in Appendix [Fig. A.1](#) for 21 of the 30 journals for which subject information is available from EBSCO.

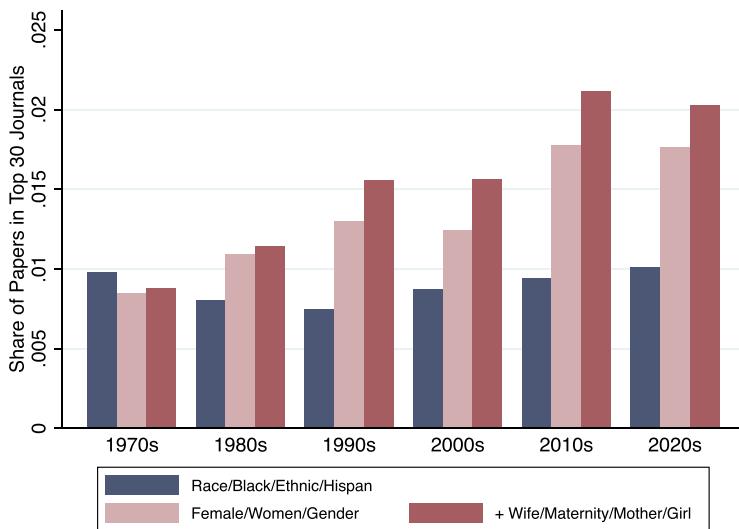


FIG. 1 Gender vs. race papers in top 30 economics journals. Note: Titles of papers published in the top 30 economics journals are extracted from Econlit. Each time period covers all papers published in the top 30 economics journals during that period. The 2020s time period is limited to the years up to 2023. To identify gender and race papers, we first perform text normalization on the titles by stemming the prefixes and suffixes followed by a keyword search using the lexical items listed in the legend above. The blue bars show the share of race-related papers; the red bars show the share of gender-related papers using different sets of gender-related keywords. Table A.1 lists the top 30 economics journals used for this exercise.

related topics halved, and were replaced by papers relating to development, health, political economy, finance, and behavioral economics. The share of gender papers relating to family economics and education remained relatively constant over this period. As such, gender-related papers have become much more evenly distributed across subfields in economics today.

This chapter traces the evolution of the study of gender, focusing on how academic thinking on this topic has evolved, and how past insights inform current perspectives on addressing the remaining gender disparities in the labor market.

In Section 2, we begin by describing the main developments in gender inequality in the labor market since the 1980s and how academic research has evolved alongside. Most of the evidence discussed refers to the United States, but we argue that the key takeaways provide a representative picture of gender inequalities in most high-income countries. While women have made significant progress in closing gender gaps in earnings, the allocation of work is to date heavily gendered, both in the labor market and in the home. Women continue to be less likely to participate in the labor market, and those who are employed work fewer hours than men. Even among those fully attached to the labor market, women continue to earn less per hour worked. Moreover, women's under-representation in market work is more than

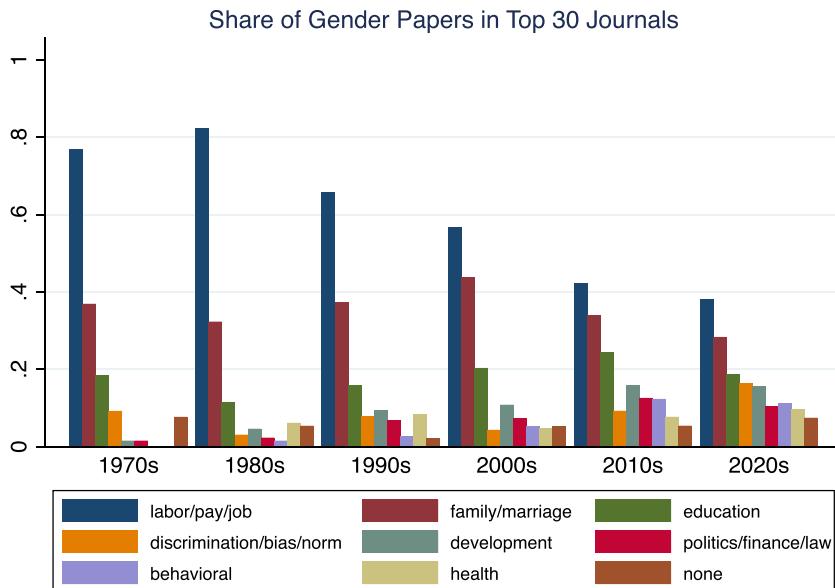


FIG. 2 Evolution of topics among gender papers. Note: Titles of papers published in the top 30 economics journals are extracted from Econlit. Each time period covers all papers published in the top 30 economics journals during that period. The 2020s time period is limited to the years up to 2023. The subset of gender papers is identified using a keyword search of the paper titles by first performing text normalizations on the titles by stemming the prefixes and suffixes followed by a procedure that flags out gender-related keywords (female/women/gender/wife/matern/mother/girl). Papers are classified into the various pre-defined topic groups based on the presence of the topic-related words in the title. The list of keywords used to identify each topic can be found in Appendix Table A.2.

offset by their disproportionate share of unpaid work in the home. Thus, women tend to enjoy less leisure time than men and their work is overall less conducive to social prestige or bargaining power within the household. As women have overtaken men in terms of completed years of schooling, and narrowed their gap in work experience relative to men, slightly more than half of the gender wage gap is now accounted for by the differential sorting of women and men into occupations and industries, with the remainder “unexplained” by observable characteristics.

The role of these forces is represented in Section 3 within a model of female labor supply, in which gender gaps in earnings reflect gender differences in the allocation of working time between the home and the market, productivity, and wage markdowns below productivity levels. Despite its simplicity, the model provides a versatile tool to illustrate how unequal gender roles in the household and departures from competitive wage setting can shape earnings gaps even once productivity differentials vanish.

Why do men and women work different hours in the market and the home, sort into different jobs, and face different wage returns? Current academic thinking emphasizes two fundamentally different explanations for the existence

of such gaps, which we elaborate in [Section 4](#). One view is that men and women have *inherently* different preferences, skills, or psychological traits that drive their choices in education and careers. In this case, gender inequality is simply a manifestation of essential differences between men and women. The other view posits that men and women are similar in the relevant dimensions, but face different opportunities and constraints. In this case, gender inequality can be a symptom of misallocation, and policies that promote gender equality can improve allocative efficiency. Naturally, a key challenge is that observed gender differences in skills, traits, or preferences could themselves be endogenous to constraints in the form of norms, stereotypes, and discrimination.

In [Section 4.1](#), we summarize findings from a body of research investigating gender differences in psychological traits and preferences since the 2000s. The emerging consensus is that those differences play, at best, a modest role in accounting for the observed gaps in pay ([Blau and Kahn, 2017](#)). Moreover, research in social psychology that has studied gender differences in a wide variety of domains including cognitive traits, communication styles, personality and social traits, establishes that, with a small number of exceptions, the data suggests that women and men are more alike than they are different ([Hyde, 2014](#)).

The relevance of gendered constraints for understanding the remaining gender gaps has shifted the academic discourse to be more upfront about the allocative efficiency consequences of persistent inequality, recognizing that enabling both women and men to reach their full potential in the labor market can confer significant economic gains through improved talent allocation, and need not come at the expense of the other group. Supporting this view, seminal work by [Hsieh et al. \(2019\)](#) documents the recent economic growth gains resulting from improved access to labor market opportunities for women and black men in the US.

Women's primary role of childbearers and carers is emphasized as one key hurdle to their continued participation and especially to their entry and retention into highly-paid but time-demanding careers. In [Section 4.2](#), we provide an in-depth review of how the literature has approached the study of the differential trade-off between family and career for mothers and fathers. This literature, which has gained momentum over the past decade, has renewed interest in, and created links with, early work in family economics, bringing richer data and a varied set of methodologies to the identification of the career costs of parenthood. The clear consensus from this research indicates that parenthood drives widening gender gaps in earnings and, following the decline in productivity gaps and outright pay discrimination, the remaining gender gaps in developed countries "are mostly about children."

[Section 5](#) describes the anatomy and dynamics of motherhood penalties, highlighting how differential constraints result in equally able women and men sorting into different types of jobs that reward workers differently to accommodate career-family considerations. Recent work has emphasized the role of

preferences for job amenities such as shorter hours and commutes, work flexibility and working from home. These translate into earnings gaps whenever women have a higher willingness to pay for family-friendly amenities than men. Such constraints have demand-side implications as well, whereby women's smaller choice set over jobs could result in wage-setting power for employers in monopsonistic labor markets.

[Section 6](#) turns to the discussion of gender identity norms. The observation that work-family issues continues to remain a "woman's problem" despite women's economic progress has brought to the fore the relevance of cultural and identity-related factors in understanding the remaining disparities in the labor market. Indeed, since the last Handbook chapter, an influential body of work has firmly established the importance of gender norms for family formation, household specialization, and labor supply. We then discuss how stereotypes and beliefs about the women's (and men's) abilities and the appropriate set of activities that they should engage in could lead to pre-market discrimination in the form of constraints to skill investment and educational choices, as well as differential treatment by employers. The net effect is a self-fulfilling cycle where individuals' preferences, traits, and skills are endogenous to gendered norms and societal expectations.

The relevance of norms for understanding gender inequality has sparked an active literature that seeks to understand what drives gender norms and how to change them. In [Section 6.3](#) we discuss relevant work on the historical origins of norms, the drivers of cultural change, transmission channels, and an emerging strand of work that suggests that information gaps could be an important contributor to the stickiness of norms. Finally, [Section 8](#) concludes with some suggestions for future research.

2 Real world and academic developments in gender inequality

The convergence in gender trends in all high-income countries, alongside persistent inequalities to date in most indicators of labor market success, have spawned decades of research on gender. To understand the development of academic perspectives on this topic, we start by describing the evolution of gender differences in labor income in the US, using data from the Panel Survey of Income Dynamics from 1980 onwards.

Between 1980 and 2018, women's employment to population ratio in the US has risen from 58 % to 74 %, average weekly hours for those in work have increased from 38 to 41 per week (while men's average weekly hours were stable at 46 h per week), and their hourly wages have risen from 62 % to 76 % of male wages. We capture these trends by showing the evolution of the gender gap in earnings, defined as the difference between male and female average earnings, relative to men's earnings. This is a summary measure that captures gender differences in all dimensions of working life, reflecting whether and

how much men and women work, the types of jobs they do, their experiences and skills, the returns to these, and frictions in wage setting, if any, including discrimination.

Fig. 3 shows trends in the gender gap in labor earnings for each decade from 1980 to 2018 for men and women aged 25 to 64. Individuals who are not working are assigned zero earnings. In 1980, the gender gap in earnings, as a percentage of men's earnings was 69 % (i.e., women earned less than a third of men's earnings). The gap fell considerably over the next two decades, and in 1998, women's earnings were about 50 % that of men's. Convergence continued, albeit at a slower pace in the last two decades. In 2018, the gender gap in earnings stood at 40 %.

Following the procedure outlined in Kleven and Landais (2017), we decompose the observed earnings gaps in each time period across these three

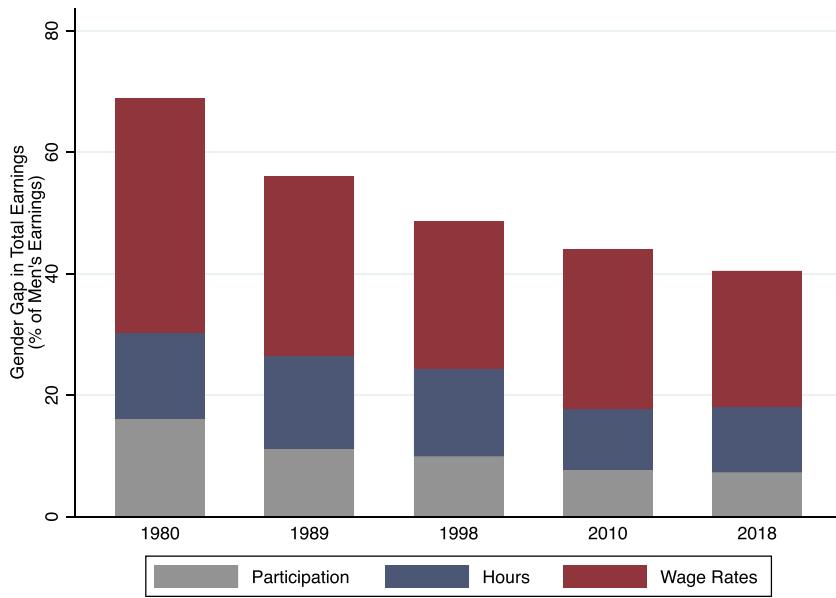


FIG. 3 Gender gap in earnings: role of participation, hours, and wage rates. Note: The data is from the 1981, 1990, 1999, 2011, and 2019 waves of the Panel Study of Income Dynamics (PSID) and includes household heads and spouses/cohabitators aged 25–64. The choice of years is based on an attempt to create 10-year blocks and the availability of PSID data which are only available biennially post-1997. The gender gap in total earnings is given by $\Delta^Y = (\bar{Y}^m - \bar{Y}^f)/\bar{Y}^m$. Following Kleven and Landais (2017), this can be decomposed into the sum of components ($\Delta^Y = G^W + G^H + G^P$) driven by gaps in wages ($G^W \equiv (\bar{w}^m - \bar{w}^f)/\bar{w}^m$), gap in hours conditional on wages being equal ($G^H \equiv (\bar{h}^m - \bar{h}^f)/\bar{h}^m \times \bar{w}^f/\bar{w}^m$) and gaps in participation conditional on wages and hours being equal ($G^P \equiv (\bar{p}^m - \bar{p}^f)/\bar{p}^m \times \bar{w}^f/\bar{w}^m \times \bar{h}^f/\bar{h}^m$), where \bar{h}^g and \bar{p}^g denote average hours of gender g conditional on working, and the employment rate, respectively. \bar{w}^g is the hours-weighted average wage rate for each gender g (i.e., $\bar{w}^g = \frac{\sum w_i^g \cdot h_i^g d_i^g}{\sum h_i^g d_i^g}$, where d_i^g indicates whether individual i participates, and w_i^g and h_i^g are the wages and hours of individual i , respectively).

margins: women being employed at lower rates (in gray), employed women working fewer hours than employed men (in blue), and women earning less per hour than men (in red).³ Across all time periods, while all three margins play an important role in accounting for gender differences in earnings, differences in wage rates typically account for more than half of the overall gap in earnings. The decline in the earnings gap over the past five decades has been driven by improvements in women's relative outcomes across the three margins. As a proportion of the overall gap, the contribution of wage gaps has been relatively stable over time, while the contribution of gender differences in participation has declined from about 23 % of the overall earnings gap to about 18 % in 2018. Correspondingly, the portion of the overall earnings gap due to women working shorter hours has increased over this period.

All three margins also play an important role in most high-income countries (see [Andrew et al., 2024, Fig. 1](#)), with some interesting patterns. First, the gap in hours tends to be larger where the gap in participation is smaller. It is likely that in countries where most women work outside the home, jobs have adjusted to facilitate the combination of home and market work and part-time work becomes widespread (e.g. in the Netherlands, the UK, and Ireland). Where fewer women work, most jobs are full-time, and gaps in hours are smaller (e.g. in southern Europe). Second, the gender wage gap is also negatively correlated to the employment gap. [Olivetti and Petrongolo \(2008\)](#) highlight that this correlation is consistent with positive selection on labor market returns, implying that in countries with lower female participation, high-wage women tend to be over-represented in the employed population. The US is among the countries in which gaps in hourly wages explain the largest share of earnings gaps.

We further analyze the sources of the gender gap in wage rates in the US by using a traditional Oaxaca-Blinder decomposition of male-female differences in log wages into a component accounted for by differences in characteristics and an unexplained component. Of particular interest in such a decomposition is the role played by human capital characteristics (e.g., education and experience), job characteristics (e.g., occupational, industry), background characteristics (e.g., race, region, and union status), as well as well as the residual (unexplained) gap.

We build on the decomposition reported by [Blau and Kahn \(2017\)](#), extending the analysis to include a more recent time period and a couple of intermediate years. The sample is similar to that for the previous figure, except that – because we are focusing on wage rates – we further restrict the sample to

³ The “participation” component reflects gaps in participation conditional on wages and hours being equal, the “hours” component reflects gaps in average hours conditional on wages being equal, and the “wage rates” component reflects gaps in average wages per hour worked. Note also that for this decomposition to work, average wages for each gender is an hours-weighted average wage rate. More details can be found in the note to [Fig. 3](#).

non-farm wage and salary workers who worked full-time, for at least 26 weeks during the preceding year.

As shown in [Fig. 4](#), in 1980 women's wages in the US were, on average, 62 % of men's wages. Controlling for gender differences in human capital closes the gap by about 11 log points, and additionally controlling for job characteristics – occupation, industry – closes the gap further by about 9 log points. Over time, gaps in human capital (education and experience) explained increasingly less of the remaining gap such that by 2018, as women outpaced men in terms of educational attainment, controlling for these variables served to *raise* women's relative wages by 9 %. By 2018, women's wages had risen to 80 % of men's wages, with more than half of the gender wage gap accounted for by industry and occupation.

While the evidence shown is restricted to labor market outcomes, gender gaps in earnings are associated to reverse gaps in unpaid work in the home. Indeed, in the vast majority of countries, including the US, the allocation of work inside the home is more heavily gendered than in the market, with women performing twice as much unpaid work as men on average across OECD countries ([Andrew et al., 2024, Fig. 2](#)). The implication is that men can enjoy more leisure time than women: in the same data, women's leisure time is, on average, 86 % that of men's. Moreover, paid and unpaid work do not convey the same economic power and prestige. In fact, work inside the home is not counted as "employment", while the same activities – e.g. educating children, keeping accounts or cleaning – would be filed under employment if performed outside the home.

Along with these broader trends in the labor market, previous volumes of the Handbook offer insights as to how perspectives on gender have evolved over time among labor economists. [Table 1](#) provides a summary of the different phases of research in gender, highlighting the real-world developments in gender inequality alongside the developments in academic research in each decade from the 1980s to the present day. Over time, there has been a clear shift away from viewing women and men as single, representative agents, toward a household-centric view where men and women take on dual roles in the labor market and the home, shaped by work-family trade-offs and cultural influences.

In the first Handbook volume published in 1986, [Killingsworth and Heckman \(1986\)](#) and [Montgomery and Trussell \(1986\)](#) document and model pre-1980 trends in female labor supply and fertility, respectively, against the backdrop of an "exogenous" rise in female relative wages. A decade later, the 1999 Handbook volume suggests a more integrated approach toward the study of gender, with [Altonji and Blank \(1999\)](#) providing a comprehensive overview of race and gender in the labor market, focusing on the role of human capital accumulation, work experience, and discrimination as key determinants of observed differentials in earnings and participation. A separate chapter is devoted to approaches to modeling labor supply and discusses family labor-supply models in detail ([Blundell and MacCurdy, 1999](#)).

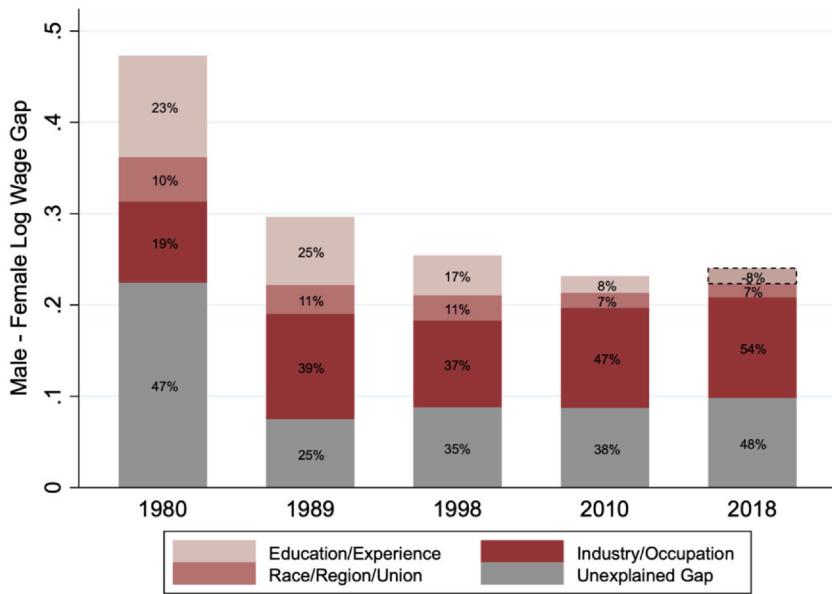


FIG. 4 Decomposition of gender log wage gap over time. Note: The data is from the 1981, 1990, 1999, 2011, and 2019 waves of the Panel Study of Income Dynamics (PSID) and includes household heads and spouses/cohabitators aged 25–64 who were full time, non-farm, wage and salary workers who worked for at least 26 weeks during the preceding year. The dependent variable is the log of average hourly earnings, which is computed by dividing annual labor earnings by annual hours worked for the year prior to the interview. The decomposition procedure and sample restrictions follow that of [Blau and Kahn \(2017\)](#) except that region controls are used instead of Metropolitan Statistical Area (MSA) as that variable is not available in the public-use version of the PSID. In addition, the analysis includes the intermediate years of 1989 and 1998, and extend the analysis to 2018.

[Altonji and Blank \(1999\)](#) summarized the 1990s consensus on the role of women's human capital gains in the ongoing process of gender convergence in earnings. Despite the closing of gaps in schooling and actual labor market experience, sizeable gaps in earnings remained, to a large extent associated with systematic differences in the jobs done by men and women. Important open questions remained about the drivers of substantial occupational segregation, which potentially reflected women's lack of specific human capital for entry into high-earnings occupations, stronger discrimination in certain occupations, differential preferences, or a combination of the three.

With the narrowing of pre-market differences, a major development starting in the early 2000s was the focus on “new classes of explanations” for gender differences in earnings and occupational choice. With application of the experimental approach to studying gender and the availability of rich, administrative databases for several countries – as well as the influence of social psychology in economics –

TABLE 1 Summary of different phases in research on gender.

Real-world developments	Academic research	
1980s	<ul style="list-style-type: none"> ● Substantial increase in female labor supply from the 1940s to 1980. ● Stylized facts on female labor supply over the lifecycle, by marital status, and presence of children. ● Large gender pay gaps in 1980: 48 log points (20 % explained by education/experience) 	<ul style="list-style-type: none"> ● Modeling and estimating labor supply elasticities of men and women. ● Understanding selection of women into participating in the labor force. ● Models of family labor supply ● Models of household formation, childbearing, and home production, with implications for female labor supply.
		HLE Vol. 1 (1986): Labor Supply of Men (Pencavel), Female Labor Supply (Killingsworth/ Heckman), Models of Marital Status and Childbearing (Montgomery/Trussel), Home Production (Gronau)
1990s	<ul style="list-style-type: none"> ● Continued increase in female LFP and decline in gender pay gap. ● Gender pay gap between 25 to 30 log points in the 1990s, partly explained by education and experience, and to a larger extent by occupation and industry differences. 	<ul style="list-style-type: none"> ● Understanding determinants of gender differentials in pay and participation through decomposition methods. ● Role of occupational segregation and discrimination as key determinants of gender disparities. ● Further development of family labor supply models.
		HLE Vol. 3 (1999): Race and Gender in the Labor Market (Altonji/Black) Labor Supply: A Review of Alternative Approaches (Blundell/McCurdy)
2000s	<ul style="list-style-type: none"> ● Gender wage gaps in the U.S. plateau at around 20 log points in the early 2000s, but continue to narrow in other rich countries. ● Increase in female LFP began to slow and plateau in the 1990s. ● Reversal of the gender gap in education. Human capital differences explain little to none of the gender pay gap. 	<ul style="list-style-type: none"> ● Focus on “new classes of explanations” for gender differences in earnings and occupational choice such as gender differences in preferences and psychological attributes and the role of gender identity norms.
		HLE Vol. 4 (2011): New Perspectives on Gender (Bertrand)

Continued

Table 1 Summary of different phases in research on gender.—Cont'd

Real-world developments	Academic research
2010s - Present	<ul style="list-style-type: none"> ● Substantial gender gaps in earnings and participation continue to persist. ● Large earnings/labor supply declines associated with parenthood for women, but not for men. <ul style="list-style-type: none"> ● Work-family trade-offs faced by women result in differential sorting across and within jobs and firms. ● Clearer distinction between the role of inherent differences between men and women and differential opportunities and constraints. ● Emphasis on gender norms and stereotypes as a fundamental source of differential constraints.

research has taken on board novel questions such as the study of gender differences in preferences and psychological traits and the role of identity norms in prescribing appropriate behavior for men and women in the family, the labor market, and society at large. The most recent Handbook chapter by [Bertrand \(2011\)](#) developed these novel perspectives on gender and laid the path to an especially active strand of research on the role of identity norms in shaping preferences, peer influences, family formation and career choices. This work has built on stark gender differences in unpaid work, coupled with the differential value attached to home vs. market work, as a significant barrier for women seeking a career and, conversely, for men seeking to spend more time on unpaid work.

3 Women's labor supply and the gender gap

We next illustrate in a simple framework how various forces (technological, institutional, or cultural) operate on the convergence or – conversely – the persistence of gender gaps in labor market outcomes. The purpose of the model is illustrative, not exhaustive, making a number of simplifying assumptions to ensure a parsimonious representation of the economic mechanisms underlying women's labor choices and the gender gap in earnings.

We model female labor supply taking wages as given. An individual's wage w can be decomposed into a latent "competitive" wage, equal to the marginal product of labor p , and a markdown below the competitive compensation level. Non-competitive forces may drive markdowns for both genders but, given the focus of this chapter, we assume for convenience that men are paid the competitive wage ($w_m = p$), while equally productive women are paid a fraction $w_f = \varphi p$, where φ represents the mark-down. For example, women might face

statistical, taste based, or monopsonistic discrimination on a given job (Lundberg and Startz, 1983; Flabbi, 2010; Manning, 2003). In addition, women may face entry barriers in certain occupations, or constraints to the range of acceptable jobs (Goldin, 2014; Goldin, 2014). As our model does not explicitly model occupational choices, we can subsume occupational “downgrading” into the parameter φ .

Men and women may also differ in their productivity (p_g , $g = f, m$), reflecting human capital differences (years of education, college major, work experience, etc.) and technological features. Whenever women are on average less productive than men, a gender pay gap would emerge even in competitive labor markets ($p_f < p_m$, $\varphi = 1$). As women’s human capital becomes more similar to, or surpasses, men’s (Blau and Kahn, 2017), and brawn-saving technologies compensate women’s comparative disadvantage in physical tasks (Heathcote et al., 2010; Ager et al., 2023), other factors become more relevant determinants of the gender earnings gap, subsumed in the wedge φ . Our framework will illustrate the importance of household specialization as a determinant of the gender earnings gap and discuss comparative statics results related to factors that affect household allocation decisions, human capital, discrimination or other frictions.

3.1 The labor supply of the secondary earner

We model the labor supply of the secondary earner within a unitary (opposite-sex) household, deriving utility from consumption of commodities (e.g. meals, vacations, childcare) produced with combination of market goods (m) and home production (H). The specific approach taken builds on the informal conceptual framework of Blau and Winkler (2021, ch. 6). The assumption here is that all household consumption is a public good, and market goods and home time are intermediate inputs in the production of the final good.

Each partner in the household has a unit time endowment. We assume that the husband works full-time in the market, supplying a fixed amount of time \bar{h}_m at the wage rate w_m , and spending the remaining time $1 - \bar{h}_m$ in home production activities.⁴ The couple jointly chooses the wife’s labor supply to the market (h_f) and her contribution to the production of the household public good $1 - h_f$.

Home production H combines the time inputs of the two spouses according to $H = (1 - h_f)^\eta(1 - \bar{h}_m)^{1-\eta}$, where η is the relative importance of the wife’s time for home making. This parameter may reflect gender absolute advantages in home making, intrinsic preferences, or gendered norms about the division of home production. Based on the assumptions made, the household’s budget constraint is given by $m = w_f h_f + w_m \bar{h}_m$. The relationship between income

⁴ The assumption that there is no private consumption or that men are the primary earners can be easily relaxed.

and the wife's home time is shown graphically in Fig. 5, with a kink corresponding to the case when the wife fully-specializes in home production ($h_f = 0$).

We consider a logarithmic household's utility function in market goods and home production, $U = (1 - \theta)\ln(m) + \theta\ln(H)$, where θ represents time intensity in the production of the commodity. The couple's maximization problem can be written as

$$\max_{h_f \geq 0} (1 - \theta)\ln(w_f h_f + w_m \bar{h}_m) + \theta[\eta\ln(1 - h_f) + (1 - \eta)\ln(1 - \bar{h}_m)]. \quad (1)$$

The parameters θ and η can vary across households, giving rise to interesting comparative statics. For example, households with higher θ have a preference for time-intensive commodities (e.g. cooking meals from scratch or parental childcare), or more limited access to time-saving technologies. Households with higher η more strongly value women's involvement in home production, for example because they believe that having a working mother is especially detrimental to the well being of young children.

The household maximization problem has an interior solution $U'(h_f) = 0$ whenever the market wage is larger than the wife's reservation wage, w_r , representing the value of the wife's home time when she fully specializes in home production, i.e. $h_f = 0$. If $w_f < w_r$, the market wage is not sufficient to let household deviate from full specialization. Analytically, the reservation wage is given by $w_r = \frac{\theta}{1-\theta}\eta w_m \bar{h}_m$ and it depends solely on preferences for commodity production and husband's income. This is higher in households with more traditional gender roles (higher η), a stronger preference for time-intensive consumption (higher θ), or higher income (with $w_m \bar{h}_m$ capturing income effects).

When $w_f \geq \frac{\theta}{1-\theta}\eta w_m \bar{h}_m$, the household's choice is described by the first-order condition:

$$(1 - \theta)\frac{w_f}{w_f h_f + w_m \bar{h}_m} = \theta\eta \frac{1}{1 - h_f}. \quad (2)$$

Recall that households are choosing the optimal combination of market goods and home hours in the production of the public good consumed. The left hand side of Eq. (2) is the marginal benefit of buying an additional unit of market goods via longer wife's hours in the market. The right hand side represents the marginal (opportunity) cost of doing so, in terms of lost utility from the home-produced services. Re-arranging, the optimal home time for a working wife is given by:

$$h_f^* = \frac{1 - \tilde{\theta}\eta \frac{w_m \bar{h}_m}{w_f}}{1 + \tilde{\theta}\eta}, \quad (3)$$

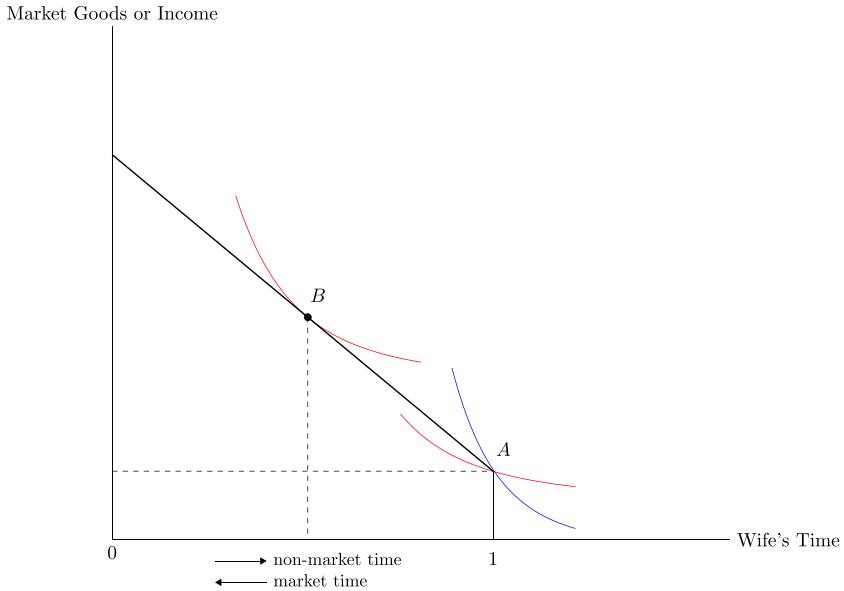


FIG. 5 The labor supply decisions of married women.

where $\tilde{\theta} = \frac{\theta}{1-\theta}$ is increasing in θ .

Fig. 5 illustrates the optimal time allocation for two households with identical husband's earnings and wife's wage but heterogeneous preferences. Household A, characterized by the steeper indifference curve, has more conservative gender roles (higher η) and/or a stronger taste for time-intensive commodities (higher θ). Household B, with flatter indifference curves, has more gender-equal norms (lower η) and favors market-intensive consumption (lower θ). In household A the wife's reservation wage (i.e. the slope of the indifference curve computed at full home specialization, $h_f = 0$) is larger than the wage rate (the slope of the budget constraint), hence she fully specializes in home production. In household B, the lower reservation wage implies that the wife is working in the market a strictly positive share of time $1 - h_f$.

In this stylized model, the comparative statics properties of the gender earnings ratio $\frac{w_f h_f^*}{w_m \bar{h}_m}$ reflect the role of wage components – productivity and the markdown – and optimal labor supply $h_f^*(w_f)$. Recalling $w_f = \phi p_f$ and $w_m = p_m$:

$$\frac{w_f h_f^*}{w_m \bar{h}_m} = \begin{cases} \frac{\phi^{p_f} \bar{h}_m^{-1} - \tilde{\theta}\eta}{1 + \tilde{\theta}\eta}, & \text{for } \phi p_f \geq \tilde{\theta}\eta p_m \bar{h}_m \\ 0, & \text{for } \phi p_f < \tilde{\theta}\eta p_m \bar{h}_m \end{cases} \quad (4)$$

Expression (4) reflects wage and labor supply contributions to the gender earnings' gap, highlighting intensive and extensive margins. Higher female productivity p_f and/or more competitive wage setting (lower p_f) increase both the probability that the wife works in the market and the earnings ratio for those employed. Preferences for time-intensive consumption and conservative gender norms $\tilde{\theta}\eta$ and income effects $p_m\bar{h}_m$ push in the opposite direction. An increase in overall productivity that leaves the gender ratio p_f/p_m unchanged does not impact the gender earnings gap. This follows from the assumption of Cobb-Douglas preference with income and substitution effect canceling each other out – an assumption that can easily be relaxed.

This simple framework is versatile and can be used to think about the role of children, the evolution of gender norms, or technical change. We can use Fig. 5 to illustrate this. As children are an especially time-intensive component of the households' public consumption, childbirth can be interpreted as an increase in θ , causing a decline in the earnings ratio. In addition, the career cost of children for mothers (discussed extensively in Sections 4.2 and 5) rises with gendered norms (higher η) and father's earnings potential, including the long-hour culture in male-dominated jobs (\bar{h}_m). Although the framework is static, the loss in labor market experience due to work interruptions can lead to a decline in latent productivity p_f , as well as additional constraints on acceptable job opportunities, leading to more monopsonistic labor markets (lower $P = F(\eta^*)$).

New technologies can affect relative earnings via both time-saving appliances in the household (Greenwood et al., 2005), lowering θ , and female-friendly technological progress (Heathcote et al., 2010), raising p_f/p_m .⁵ The resulting increase in the wage ratio w_f/w_m has a larger impact on relative earnings for household with flatter indifference curves, who are more responsive to economic incentives due to larger substitution effects relative to income effects.

3.1.1 Modeling gender norms and beliefs

Gender norms may be themselves endogenous to female labor supply outcomes via a variety of transmission processes within and across generations, as will be documented extensively in Section 6. Gender biases in social norms may be eroded by exposure to working women via role models, peer influences and learning (see Fernández (2013), and references therein). Here, we model the bilateral interplay between norms and female outcomes in a simplified version of our model that focuses on the extensive margin of labor supply, in which women either work fulltime ($h_f = \bar{h}_f$), or fully specialize in the home ($h_f = 0$).

⁵ An important dimension of the rise in relative female productivity relates to the expansion of the service economy, which is more female-intensive than manufacturing, see Ngai and Petrongolo (2017) and references therein.

This discrete choice model can easily be extended to allow for part-time work, $h_f = \{0, \bar{h}_L, \bar{h}_H\}$, where $\bar{h}_L < \bar{h}_H$, and can provide a framework for occupational choice, where occupations differ in time demands. As above, men have a fixed workweek \bar{h}_m .

We let norms vary across households. In particular, norms of household i are summarized in η_i , which is distributed according to $F(\eta)$. The maximization problem of household i becomes:

$$\max_{h_f \in \{0, \bar{h}_f\}} \theta \ln(w_f h_f + w_m \bar{h}_m) + (1 - \theta)[\eta_i \ln(1 - h_f) + (1 - \eta_i) \ln(1 - \bar{h}_m)], \quad (5)$$

with women working the fixed amount \bar{h}_f if

$$\theta \ln(w_f \bar{h}_f + w_m \bar{h}_m) + (1 - \theta) \eta_i \ln(1 - \bar{h}_f) \geq \theta \ln(w_m \bar{h}_m),$$

and non participating otherwise. Let's define η^* as the norms of the marginal household, who is indifferent between the case in which the wife participates to the labor market and the case in which she fully specializes in home production. Based on condition (6),

$$\eta^* = \tilde{\theta} \ln \left(\frac{w_m \bar{h}_m}{w_f \bar{h}_f + w_m \bar{h}_m} \right) \frac{1}{\ln(1 - \bar{h}_f)} \quad (7)$$

In households with $\eta < \eta^*$ the wife works in the market, whereas in households with η_{0i} she only works in the home. Given the distribution of norms, female labor force participation is given by $P = F(\eta^*)$. The comparative statics on female participation encompass the same factors as in [Section 3.1](#), which are now captured by the summary statistics η^* .

We next endogenize norms by letting η_i vary with female labor force participation, $\eta_i = \eta_{0i} - \eta_l P$, where the parameter η is constant across households, while η_{0i} is distributed according to F . The marginal household is now characterized by

$$\eta_0^* = \tilde{\theta} \ln \left(\frac{w_m \bar{h}_m}{w_f \bar{h}_f + w_m \bar{h}_m} \right) \frac{1}{\ln(1 - \bar{h}_f)} + \eta_l P \quad (8)$$

The resulting participation rate is then found as the fixed point solution to $P = F(\eta_0^*(P))$, which potential leads to multiple equilibria and coordination mechanisms, as in models by [Fernández \(2011\)](#) and [Hazan and Maoz \(2002\)](#).

4 Evolving perspectives on gender inequality

The framework above is useful to highlight several channels leading to gender gaps in earnings. With gender convergence in human capital and productivity, the literature on gender inequality has coalesced around the study of gender

differences in preferences and constraints, imperfectly competitive wage setting, as well as the understanding of gendered norms.

4.1 Preferences, traits, and constraints

Over the last decade, research has made significant inroads in understanding the separate roles of preferences and constraints in driving gender gaps. As a large portion of the remaining gaps in earnings is explained by differences in the pay and attributes of the jobs that men and women perform on average, it is important to establish whether gender differences in the job allocation reflect differential preferences for job attributes and/or differential skills in those jobs, versus differential barriers to entry into certain careers or the labor market as a whole. Naturally, a key challenge to making this distinction is that observed gender differences in skills, traits, or preferences could themselves be endogenous to constraints in the form of norms, stereotypes, and discrimination.

Building on seminal experimental work discussed in [Croson and Gneezy \(2009\)](#) and [Bertrand \(2011\)](#), research on gender differences in preferences and psychological attributes has continued to expand around traits like risk-aversion, self-confidence, competitiveness, willingness to negotiate and ask, as well as other-regarding preferences. These traits are relevant for the choice to enter prestigious and financially rewarding careers, which often develop in competitive and risky environments. There are clear hurdles to estimating the causal impact of psychological traits on labor market outcomes. One challenge is about measurement, as one would ideally need to measure those individual traits before they may be contaminated by the outcomes of interest. Another challenge is about the distinction between direct effects, which affect the performance of an individual on a given job, and indirect effects, which shape human capital investments and labor supply choices. [Blau and Kahn \(2017\)](#) draw important lessons on these questions from a small number of studies that relate psychological traits to gender pay gaps. Overall, available results suggest that psychological factors account for a modest portion of both the raw or the adjusted wage gap for individual and job characteristics. However, a few studies suggest that even similar traits may feed into gaps in outcomes when they are regarded and rewarded differently depending on whether they are displayed by a man or a woman.

Importantly, recent studies in psychology have shown that mean gender differences in many individual traits are small relative to their variation within each gender. [Hyde \(2005\)](#) pools results from meta analyzes on gender differences in several cognitive and non-cognitive skills, communication, personality traits, well-being, and moral reasoning. For 78 % of the 124 characteristics assessed, there is a sufficiently large overlap between the male and female distributions to conclude that men and women are more alike than they are different in many relevant traits (see also the discussions by [Hyde, 2014](#) and [Bertrand, 2020](#)). Despite this evidence, gender differences are often

exacerbated in the perceptions of economists. For example, in the study of [Bandiera et al. \(2022\)](#), over three quarters of economists surveyed believe that women are on average under-confident, while men are on average over-confident. By contrast, a meta-analysis of experimental results in economics reveals that both men and women are on average overconfident, and a Bayesian hierarchical model that aggregates available estimates cannot reject the hypothesis that they are equally over-confident. In addition, the estimated pooling factor across studies is low, implying that each study contains limited information over a common phenomenon. The discordance between perceptions and empirical results may be reconciled if economists overestimate the pooling factor across available estimates, or have priors that are both biased and precise.

The gender similarity hypothesis of [Hyde \(2005\)](#) acknowledges clear exceptions in a few domains. Men perform better on average on measures of motor skills, behave differently in some measures of sexuality, and are more likely to be physically aggressive. While it may be tempting to hypothesize that these dimensions do not directly shape labor market success, recent work has documented important consequences of sexual harassment and violence against women in general on the economic outcomes of victims, perpetrators, and their workplaces. Understandably, these themes have attracted increasing attention among economists since the #MeToo movement has made salient the pervasiveness of sexual harassment at work and its personal and professional costs for victims.

[Folke and Rickne \(2022\)](#) provide a novel, comprehensive study of sexual harassment in workplaces in Sweden and its effects on gender inequality in the labor market at large. In nationally representative survey data, women are three times more likely than men to report sexual harassment over the past year (12.6 % versus 4.2 %). For both men and women, harassment risk rises markedly with the share of opposite-gender co-workers. As harassment victims are more likely to move employer, victimization leads women to quit male-dominated firms, which also happen to be high-wage firms, and viceversa for men. Harassment-related mobility hence exacerbates sex segregation and pay inequality, explaining about 10 % of the gender wage gap.

Using data on police reports in Finland, [Adams-Prassl et al. \(2024\)](#) link cases of violence between co-workers to the economic outcomes of suspect perpetrators (84 % of whom are male) and victims (evenly split between male and female). Violence at work drives sudden and persistent employment losses for both parties, with starkly asymmetric effects between cases of male-on-male and male-on-female violence. While male victims experience smaller negative repercussions than their male assailants, female victims bear a *larger* economic penalty than their male assailants. This asymmetry is often related to systematic differences in seniority between victims and perpetrators. Importantly, following violence against women at work, incumbent women are more likely to leave the firm, and fewer women are hired in the medium term. These

patterns are concentrated in male-managed firms, while in female-led firms perpetrators experience similar employment losses as their victims. The economic costs of assaults for women are not limited to workplace violence. [Bindler and Ketel \(2022\)](#) find large and persistent earning losses among women who experience intimate partner violence and [Adams et al. \(forthcoming\)](#) document detrimental employment and earning effects of cohabiting with men who are ever reported for intimate-partner violence. Their results suggest that economic coercion is a central component of abusive relationships, even before a partner is reported for physical violence. One important lesson from this literature is that gender differences in traits such as sexual abuse or violence may translates into a barrier to women's economic success.

Recent perspectives on barriers to women's labor market involvement have produced major advances in economists' understanding of gender gaps. The first key advance consisted in acknowledging that differential gender opportunities and barriers naturally lead to questions of allocative efficiency. Starting from the premise that the distribution of innate talent does not systematically vary by gender, the under-representation of women in certain professions implies that female talent is not efficiently allocated in the economy. Indeed, [Hsieh et al. \(2019\)](#) estimate that between 20 % and 40 % of growth in GDP per capita in the US over the past half a century can be explained by the improved allocation of talent, thanks to improved access to education and declining occupational segregation for women and black men.⁶

Consistent with this narrative, one should expect productivity gains from hiring more women in male dominated contexts, in which women are likely to be positively selected. Using personnel records from a multinational firm, [Ashraf et al. \(2024\)](#) show that the performance of female employees within the organization is higher in countries where women are underrepresented in the candidate pool. These tend to be countries in which predominant gender norms discourage women's participation to the labor market as a whole. Similarly, [Chiplunkar and Goldberg \(2024\)](#) and [Mertz et al. \(2024\)](#) find evidence of improved firm performance when women face lower entry barriers in entrepreneurship in India and Denmark, respectively.

An important, symmetric question is whether men's under-representation in certain female-dominated professions implies that their talent is not efficiently allocated. [Delfino \(2024\)](#) shows that attracting male applicants into the UK social care sector – in which the share of female employees hovers around 80 % – improves the selection of male talent. [Schaede and Mankki \(2024\)](#) find and that lifting a 40 % male hiring quota from the Finnish education sector in the late 1980s led to higher female concentration in the sector and lower

⁶ Using a similar framework in an international context, [Chiplunkar and Kleineberg \(2024\)](#) estimate that the reductions in gender-specific wedges in employment and remuneration between 1970 and 2015 account for around a third of labor reallocation towards manufacturing and services in six large countries.

attainment among pupils. Evidence suggests that the diversity quota achieved a more efficient allocation of talent than the unconstrained selection process that followed, as this seemed to penalize valuable traits and skills among the under-represented group.

While these studies provide evidence of productivity gains from the entry of men and women into contexts where they are under-represented, it is hard to draw conclusions on aggregate welfare from analyzes of specific sectors, without a symmetric analysis of gains and losses in the respective feeder sectors. By considering job allocation across the whole economy, work by [Hsieh et al. \(2019\)](#) is an important exception. However, even in their analysis the household sector is not explicitly modeled. A fuller understanding of the welfare consequences of a more equitable allocation of talent would require a general equilibrium perspective that considers both genders' comparative advantage in the labor market and the home.

The second advance consisted in pushing the research frontier on the understanding of the nature and sources of the main entry barriers to the labor market or to specific professions. Women's primary role of child-bearers and carers is typically emphasized as one key hurdle to their continued participation and especially to their entry and retention into highly-paid but time-demanding careers. We will describe below how research on career-family trade offs, which has gained momentum over the past decade, has renewed interest in and created links with early work in family economics, bringing a new perspective on gender identity norms and richer data to measure them.

4.2 Career-family trade-offs

While female labor supply and fertility are deeply intertwined, much of the early work has dealt with each issue in isolation. Since at least [Becker \(1960\)](#), several authors have studied fertility in static or life-cycle settings, treating labor supply decisions as given. Conversely, the early labor literature typically focused on female participation, treating fertility as exogenous (see, for example, [Heckman and MacCurdy, 1980](#)). Seminal work on the joint labor supply and fertility decisions of women by [Moffitt \(1984\)](#) has modeled, in a dynamic setting, the simultaneous evolution of wages, labor force participation, and fertility, allowing for correlated shocks to fertility and participation. In his model, the detrimental effect of children on mothers' participation and wages reflects both the time involved in raising children and the loss of actual labor market experience.

[Francesconi \(2002\)](#) introduces the intensive margin of labor supply in joint dynamic, structural models of participation and fertility. He considers part-time employment as an alternative to labor market breaks when the disutility of work increases during childbearing years. However, the model estimates imply that part-time employment hardly cushions maternal earnings in the long-run relative to career breaks, as returns to part-time work experience appear to be

substantially lower than the returns to full-time experience. In particular, the convex relationship between returns to experience and working time suggests that part-time and full-time jobs differ systematically beyond the length of the typical workweek, and most notably in the types of occupations typically available on a full-time or part-time basis.

[Adda et al. \(2017\)](#) propose a natural modeling of the career costs of children, subsumed in occupational choice. Occupations capture the bulk of gender inequality in earnings, as only a small portion of the remaining gender gap in contemporary data is explained by unequal pay for equal work. If the choice of occupation is a key factor underlying diverging career paths for men and women, especially after parenthood, it is important to understand the drivers of such choices and the main features of occupations sought out by women with children, and those who intend to have them in the future. [Adda et al. \(2017\)](#) incorporate occupational choices in a dynamic life-cycle model of participation, fertility, and asset accumulation where occupations differ in their wage profiles, the speed of skill depreciation associated to career breaks, as well as their family friendliness. Estimating their model on men and women completing apprenticeships in Germany, the authors find that abstract occupations have relatively high returns to experience, high penalties for career breaks, and poor amenity value once women have children. This implies that the interplay between fertility choices and career concerns are therefore far more relevant in abstract than routine occupations (with manual occupations faring somewhere in the middle). Based on their model estimates, the authors conclude that about three quarters of the career costs of children stem from reduced or intermittent participation, with the rest being explained by occupational choices, skill depreciation, and a reduction in working hours.

By interacting career and fertility decisions over the lifecycle, an important feature of adda's model is that fertility plans are allowed to shape women's human capital and occupational choices ahead of childbirth. However, the authors estimate that the anticipation effects of motherhood are relatively small. For example, the choice of apprenticeship track during teenage years by women who intend to have children explains about 5 % of the lifetime cost of fertility, and the earnings gap with respect to men only starts to build up for women who intend to have children around age 26, just before the average age at first birth. These results are consistent with survey evidence that women systematically underestimate the impacts of prospective fertility on their labor market involvement ([Kuziemko et al., 2018](#)).

Another strand of work has addressed the potential endogeneity of fertility and its timing by using instruments for the number of children in female labor supply equations, such as twin births ([Rosenzweig and Wolpin, 1980](#); [Bronars and Grogger, 1994](#)) and sibling sex composition ([Angrist and William, 1998](#)). These instrumental variable (IV) estimates typically deliver negative impacts of fertility on maternal labor supply, although these tend to be relatively short-lived and smaller than those obtained from OLS.

A drawback of these early papers is that the fertility impacts are limited to the arrival of a second or third child, and therefore miss the role of the extensive margin of fertility. Later work leverages (in)fertility shocks to investigate differential labor supply outcomes between mothers and childless women (see, among others, Hotz et al. 2005; Lundborg et al. 2017; Gallen et al. 2023; Bögl et al. 2024). In particular, by comparing women who conceive through in-vitro fertilization (IVF) in Denmark to those who attempt to conceive through IVF but fail, Lundborg et al. (2017) detect large impacts of fertility on maternal earnings in the short run, with much smaller impacts beyond a child's second birthday. Bögl et al. (2024) obtain similar results for Norway, by comparing women who have a live birth after their first medically-assisted conception to those who miscarry. It is important to note in these comparisons that the childless (control) group is made of women who experience failure in their struggle against infertility, as Bögl et al. (2024) find that women who miscarry after their first medically-assisted conception are significantly more likely to take mental health medication than those who have a live birth. The interpretation is that both the arrival of children and the mental health toll of a miscarriage have detrimental impacts on earnings, thus the comparison between treatment and control groups may underestimate the overall impact of fertility on earnings. Unlike Bögl et al. (2024), Lundborg et al. (2017) find negligible impacts of infertility on the incidence of depression in Denmark.

The internal validity of the IV approach requires the probability of success of fertility treatments to be orthogonal to earning trajectories. This point has been questioned by Groes et al. (2024), who find that college-educated women in Denmark have a 9 % higher live birth chance upon IVF than high school-educated women, and 25 % higher chances than high-school dropouts. The external validity of local average treatment effects obtained on women who opt for IVF treatment crucially relies on the representativeness of this selected sample.

To capture the overall treatment effect of fertility, the past decade has seen a proliferation of event-study evidence on the career costs of childbirth. This approach leverages sharp changes in outcomes around first childbirth for mothers relative to fathers. The fundamental assumption that the timing of fertility is independent of counterfactual outcomes is typically motivated by the occurrence of sharp breaks in career trajectories upon birth, without major anticipatory effects. The wide consensus from this body of work is that, while childbirth is roughly neutral for men's labor market trajectories, it drives a sudden and largely persistent setback in women's earnings. Angelov et al. (2016) estimate that Swedish couples experience a widening of about 30 % points in the spousal gap in earnings during the first 15 years of parenthood. Kleven et al. (2019) estimate a long-run "child penalty" in Denmark of about 20 %: this measures the extent to which female earnings fall relative to male earnings due to childbirth, encompassing reduced maternal participation,

reduced hours for mothers who participate, and lower hourly wages.⁷ Importantly, the child penalty in Denmark has hovered around 20 % since the 1980s, against a backdrop of rapidly declining gender gaps in human capital as well as unexplained gaps. Therefore, while in the 1980s child-related inequality was explaining about 40 % of the gender gaps in earnings, its role had doubled by 2013. Research on additional countries has revealed similar patterns, with some variation in magnitudes: between five and ten years into parenthood, women's earnings typically fall behind men's earnings by 20 %–25 % in Denmark and Sweden, 30 %–40 % in the US and the UK, and up to 50 %–60 % in Germany, Austria and Italy (Kleven et al., 2019; Casarico and Lattanzio, 2023).

While the event-study approach of Kleven et al. (2019) requires high-quality panel data, Kleven et al. (2024) show that results from this approach can be closely replicated on cross-sectional data organized as a pseudo-panel, effectively extending the feasibility of child penalty estimates to most countries around the world. This requires building pre-childbirth employment histories for parents, based on employment outcomes of childless individuals with matching characteristics. An interesting finding from the child penalty "atlas" of Kleven et al. (2024) is that female employment losses associated with marriage and childbirth are negligible in countries with very low levels of GDP per head, then rise at intermediate levels of development, before starting to fall again towards the top of the country ranking. Using twin birth and same gender instruments for incremental fertility, Aaronson et al. (2020) document similar patterns on a large cross-country panel spanning over two centuries.

These trends clearly correlate with various dimensions of structural transformation. In predominantly agricultural societies, most women work flexibly on or near the household premises, and their work is compatible with marriage and childcare. The transition towards industrialization and the service economy, in tandem with urbanization and the de-localization of work, drives progressively larger child-related gaps in employment, as childcare requires some degree of specialization. At highest income levels, economies can create family-friendly jobs that ease the combination of families and careers. The hump-shape pattern in the family penalty mirrors the U-shape pattern in female employment emphasized by Goldin (1990) and Ngai et al. (2024), among others.⁸

The IVF-based IV approach and the event-study approach recover conceptually different treatment effects of fertility based on different identifying assumptions. Event studies are centered around the time of first birth and identify

⁷ Comparable life-cycle evidence has been shown in Bertrand et al. (2010) and Goldin and Mitchell (2017), among others.

⁸ The relationship between marriage and (the timing of) fertility has evolved over time and varies across countries. Papers that separately identify the two channels find that, conditional on fertility, marriage penalties rapidly decline with levels of development, and that the whole "family" penalty is accounted for by the presence of children among recent cohorts of parents in high-income countries (Kleven et al., 2024; Juhn and McCue, 2017; Albanesi et al., 2023).

dynamic treatment effects of fertility by comparing the earning trajectories of women who give birth to those of women of the same age who give birth at different ages. In doing this, they assume exogenous birth events with respect to counterfactual career outcomes (conditional on included controls) and smooth counterfactual outcomes around childbirth.⁹ The IV-IVF approach is centered around the time of the first IVF attempt and identifies treatment effect on compliers, i.e. those who conceive at the first attempt. This requires assuming that IVF success is orthogonal to career outcomes and it only affects outcomes via fertility. However, many women who initially fail to conceive through IVF become mothers later, via IVF or otherwise. Estimated treatment effects of fertility would thus be downward biased by delayed fertility behavior.

[Besnes et al. \(2023\)](#) combine both approaches using data on Norwegian women undergoing IVF treatment and their partners. To address biases related to delayed fertility, they estimate an event-study model centered around first birth and, to address concerns of endogenous timing of birth, they capture a woman's intention to conceive by controlling flexibly for time since the first IVF treatment. Their results show a 23 % widening of the parental gap in earnings after birth, shrinking to 13 % in the long-run (mostly driven by a fertility premium for partners). This long-run penalty is smaller than the 18 % estimate obtained with the conventional event-study approach and larger than the 4.8 % IVF-based IV estimate. This is consistent with both the role of endogenous birth timing, if women tend to time fertility when their earning profiles flatten, and the a downward bias induced by delayed fertility in IV estimates.

Most recent contributions in this literature highlight additional interesting patterns in child penalties. [Adams et al. \(2024\)](#) find that a sizeable portion of the child penalty in Denmark is explained by spells of parental leave, when mothers are not working or earning a salary but are entitled to return to their pre-birth job and pay within their leave entitlement. This finding stresses the importance of distinguishing between “incapacitation” effects of parental leave, which are typically incurred soon after each birth, and longer-term impacts that may happen via the loss of actual work experience and adjustments in labor supply at intensive or extensive margins. [Kuka and Shenay \(2024\)](#) document the important role played by loss of actual experience during career breaks among single mothers in the US. In particular, those who were exposed to work incentives immediately after birth rather than 3–6 years later, accumulate 0.62 additional years of experience and have 4.2 % higher earnings conditional on working. These results suggest that work experience soon after birth may be rewarded with steeper returns. We will expand on detected patterns of labor supply adjustment in Section.

⁹The smoothness assumption is not sufficient for identification of long-run child penalties, as it would be necessary to assume that women who give birth at very different ages provide valid control groups for each other.

Finally, most of the literature on child penalties emphasizes changes in maternal labor market outcomes, but what happens to men who have children, compared to those who do not? A long standing literature has detected marriage and fatherhood premia for men in the US, whether in cross-sectional or within-group estimates (see for example the discussion in Juhn and McCue, 2017). Fatherhood may shape wages through employer perceptions and possibly (positive) discrimination. Indeed, Korenman and Neumark (1991) find that the earnings profile of men steepen after marriage, to some extent thanks to more favorably rating by their managers. Similarly, Correll et al. (2007) find that fathers tend to be evaluated more positively than non-fathers, while the opposite happens to mothers, despite equivalent qualifications. In addition, men may respond to societal pressures linked to the breadwinner model by working longer and harder. For instance, Killewald (2013) finds that the fatherhood premium is largest among residential, married, biological fathers, who might feel greater incentives to improve their children's well-being, compared to stepfathers or non-residential fathers. Lundberg and Rose (2000) find a substantial reallocation of time and effort for married couples associated with the arrival of children. Following the birth of the first child, the father's wage increase by 7 % in households where the mother is continuously employed, and by 11 % in households where the mother has a career break. More recently, Goldin et al. (2024) find that fathers earn a wage premium that cannot be fully explained by selection into fatherhood (i.e., higher-ability or harder-working men being more likely to become fathers.) Similar results obtained by Besnes et al. (2023) using the IVF-based IV tend to exclude that fatherhood premia simply reflect endogenous fertility around men's earning growth. Goldin et al. (2024) also note that fatherhood premia in the US are larger among college graduates, and especially for men working in occupations that require long and/or inflexible hours. This evidence is consistent with progressive specialization of men and women in paid and unpaid work, respectively, once they become parents.

While available approaches differ on assumptions and strengths and weaknesses, consensus is that parenthood drives widening gaps in parental earnings, and the study of the anatomy and drivers of the child penalty is currently one of the most actively researched areas of gender inequality.

5 The anatomy of the career costs of motherhood

5.1 Gender biology and productivity

First-order questions on the drivers of child penalties are whether they reflect biological components of women's caring responsibilities, as opposed to acquired patterns of specialization, and to what extent they result from productivity differentials, as opposed to larger markdowns of wages below productivity for women after they become mothers.

As pregnancy, birth and breastfeeding may set limits on women's labor market involvement, [Kleven et al. \(2021\)](#) investigate the role of these factors by comparing earnings penalties for biological and adoptive mothers in Denmark. Similarly as for biological parents, earnings trajectories for adoptive parents evolve in parallel before adoption, and diverge persistently afterwards. Short-run penalties are smaller in adoptive than biological families, but long-run penalties are very similar. [Andresen and Nix \(2022\)](#) leverage additional evidence from same-sex female couples in Norway, as well as heterosexual adoptive couples. While heterosexual couples – whether biological or adoptive – experience similar setbacks in earnings for mothers and virtually no drop for fathers, same sex couples share the cost of children much more evenly, with a somewhat larger drop for the biological mother in the short-run, but virtually no difference between the biological mother and her partner in the long-run. These pieces of evidence establish that maternal biology is unlikely to drive persistent drops in earnings, although it plays a modest role within a couple of years from birth.

Estimates of gender differences in productivity are scant. Seminal estimates are from contexts in which productivity is easily measurable. [Azmak and Ferrer \(2017\)](#) find that male lawyers bill 10 % more hours and bring in more than twice as much new client revenue as female lawyers. Much of this gap is explained by the presence of your children and differential aspirations to become a partner in the firm. [Cook et al. \(2020\)](#) estimate that male drivers earn on average 7 % more than female drivers on the Uber rideshare platform. This differential reflects men's higher willingness to drive in more lucrative locations (with higher crime and more drinking establishments), their sector-specific human capital (as they typically accumulate more job experience), and their higher driving speed.

[Gallen \(2023\)](#) investigates gender differences in productivity in six large private sector industries in Denmark. Her estimates leverage variation in value added across firms employing different proportions of female employees, conditional on human capital, hours worked, and detailed occupation. She finds that wage gaps between mothers and men approximately reflect underlying productivity differences, although part of the productivity gap is driven by some gradual reallocation of women into lower-TFP firms once they have children. The fact that mother's pay is on average aligned with their relative productivity excludes (observable) discriminatory pay differences for equal work, although it may not explain sorting across occupations or employers, something that we will discuss in the next Section. Interestingly, [Gallen \(2023\)](#) documents evidence of uncompensated productivity premia for childless women, especially during their prime child-bearing years, possibly consistent with a statistical discrimination channel: if employers cannot reduce wages when women have children, they may offer lower wages to childless women in anticipation of motherhood.

For those working from home (WFH), the presence of children may directly impact productivity via work patterns. [Adams-Prassl et al. \(2023\)](#) find that that mothers working for the online MTurk platform are more likely to interrupt

their time on the platform, with consequences for the speed of completing tasks. [Ho et al. \(2024\)](#) shows similar findings on women in India who are offered the opportunity to WFH and multitask work with childcare. In both cases, piece-rate compensation implies that efficiency costs are borne by workers. However, efficiency losses may discourage firms from offering WFH under typical time-rate compensation.

5.2 Differential job sorting and the organization of work

A large body of work has complemented evidence on within-job gender differences in productivity with analyzes of differential job sorting of mothers and fathers. For example, [Kleven et al. \(2019\)](#) estimate that, soon after childbirth, working mothers tend to fall behind in the occupational ladder with respect to fathers, and are less likely to hold managerial roles. They are also more likely to move to the public sector and to firms led by female managers with children suggesting the pursuit of family-friendly working conditions at the expense of higher pay. In the presence of gender differences in preferences for working conditions, models of compensating differentials (e.g., [Rosen, 1986](#)) imply that women are willing to accept lower earnings in exchange for desirable job amenities that are costly for employers to provide. This view places special emphasis on the role of the organization of work in shaping gender gaps.

[Goldin \(2014\)](#) argues that a major source of the remaining pay disparities, especially among highly-educated (and equally qualified) men and women is the fact that many of the highest paying occupations are also those that disproportionately reward individuals who are willing to work long (and particular) hours. As women tend to work fewer hours, they tend to suffer greater earnings penalties relative to men in such occupations. Indeed, she documents that occupations that exhibit the greatest convexity of pay with respect to time worked also have the largest gender earnings gaps. The remuneration of family-unfriendly work schedules is particularly prevalent in the corporate, financial, and legal sectors, suggesting that such organizational practices are likely to be a key factor behind the substantial gender pay gaps that emerge over the lifecycle in these professions, especially with the arrival of children ([Bertrand et al., 2010; Noonan et al., 2005; Azmat and Ferrer, 2017](#)).

Building on these observations, [Cortés and Pan \(2019\)](#) show that relaxing the work hours constraint faced by highly-educated women, through the increase in the availability of low-cost and flexible household services in the form of low-skilled immigrant labor, increases the relative earnings of women in occupations that reward overwork. Moreover, in cities with greater inflows of low-skilled immigration, women are more likely to be found in higher quantiles of the male wage distribution, and young women are more likely to enter occupations with higher returns to overwork. Focusing on the medical profession, [Wasserman \(2022\)](#) shows that a policy that directly reduced a job's time requirements affected women's propensity to enter the job and the gender

wage gap. Using data on the universe of US medical school graduates, and exploiting a 2003 policy that capped the average workweek for medical residents at 80 h, she finds that medical specialties that experienced larger declines in weekly hours attracted more women, against roughly unchanged numbers of men. A back-of-the-envelope calculation reveals that the entry of women into high-compensation specialties due to the reform potentially closed the physician gender wage gap by 11 %.

The large differences in “flexibility penalties” across occupations suggest that organizational changes to reduce such penalties offer a promising solution to addressing gender pay gaps. Yet, our understanding of the precise sources of the returns to working long/inflexible hours remains limited. Goldin (2014) offers several examples of occupations and sectors such as obstetricians, pharmacists, and veterinarians that have moved toward greater hours flexibility by increasing substitutability among employees. Nevertheless, the causes of the changes appear to be quite varied and range from the re-organization of work to take advantage of economies of scale, lower labor costs, or employee pressure. Others have suggested that such practices could arise due to signaling considerations in situations where actual productivity is hard to observe, and firms rely on work hours as a proxy for productivity (see Landers et al., 1996; Tō, 2024). Additional case studies and research that can further elucidate the sources of occupational differences in the returns to overtime/inflexible hours and shed light on how workplace practices can be changed (ideally with little or no productivity costs) would be especially promising.

The surge in working from home (WFH) – especially in hybrid format – after the COVID-19 pandemic has provided economists with an unprecedented testing ground for investigating its benefits to employers and employees, but evidence on gendered impacts is thin, at least in the short-run. Evidence from the WFH experiment of Bloom et al. (2022) has shown that the introduction of hybrid work in a global travel-agent headquartered in Shanghai had no direct impacts on measured performance, but led to 33 % lower quits and higher employee satisfaction. Interestingly, women were disproportionately less likely to quit their jobs, relative to men, if given the opportunity to WFH. However, they were less likely than men to volunteer for the WFH experiment and ex-post take-up rates of WFH were very similar across genders. This apparent paradox could be possibly explained by gender differences in concerns over the career costs of signaling a preference for remote work. In addition, women opting for WFH may deepen gender roles within their households by increasing their availability for home-based duties.

While results from this body of evidence are consistent with compensating differentials associated to family-friendly working conditions, in practice it is not easy to infer workers’ valuation of job amenities from observational data on job choices since the observed relationships between earnings and specific job attributes tend to be confounded with unobserved worker characteristics and job attributes. To sidestep these issues, researchers have turned to the use

of hypothetical job choice experiments to estimate individual preferences for workplace attributes.

In these experiments, respondents are asked to choose which job they prefer (out of two or three job offers) from a series of hypothetical scenarios that are constructed to reflect a realistic menu of potential job offers that vary in earnings, and other job characteristics (e.g., hours of work, work hours flexibility, the option to work part-time, etc.). Job characteristics, including earnings, are randomly varied across job offers within each scenario. Individual preferences for each job attributes (or willingness to pay) is then measured in terms of the amount of earnings that respondents are willing to forego for a particular job attribute.

Using such an approach, [Wiswall and Zafar \(2018\)](#) elicit preferences over hypothetical job attributes among New York University undergraduates. They find that women express on average a much stronger preference for flexibility in working hours than men, with an implied willingness to pay (WTP) of 7.3 % compared to 1.1 % for men, while men have a higher WTP than women for higher earning growth. They also show that self-reported preferences for job attributes have a sizeable impact on major choice. Overall, the authors find that gender differences in preferences for these job attributes can explain as much as a quarter of the early-career gender gap in earnings. Relatedly, using a similar stated-preference approach, [Maestas et al. \(2023\)](#) document that women, in general, place a higher value than men on avoiding physically demanding work, paid time off, and the option to telecommute.

[Mas and Pallais \(2017\)](#) provide evidence on preferences for actual work arrangements, by introducing a discrete choice experiment in the application process for call center positions across the US. Applicants can express their preferences between a conventional 9-to-5, 5-day a week, office job and alternative arrangements featuring flexible scheduling, working from home (WFH), or employer discretion over work schedules. Wages are randomized across these options. While the large majority of applicants do not value flexible scheduling, on average they value the opportunity of WFH and dislike employer discretion in scheduling. Women, especially those with young children, express a higher WTP for these job attributes than men. However, in their setting, as the incidence of these attributes is fairly similar for men and women, gender differences in the WTP for them cannot lead to sizeable gender gaps in pay, even under large compensating differentials.¹⁰

¹⁰ Other studies have also sought to infer WTP for job attributes using other approaches besides discrete choice experiments. Using a revealed preference approach, [Felfe \(2012\)](#) infers mother's WTP for job amenities from the response of maternity leave take-up to the characteristics of jobs they are returning to, and [Hotz et al. \(2018\)](#) estimate that mothers value workplaces with higher shares of female co-workers more favorably than fathers do, where co-worker composition is interpreted as a correlate to unobservable amenities.

Recent work has also directly documented gender differences in job search strategies. Using administrative data on unemployed jobseekers in France, [Le Barbanchon et al. \(2021\)](#) document women's higher willingness to pay for shorter commutes. In particular, they estimate that gender gaps in reservation wages, post-unemployment wages, acceptable commutes and realized commutes all widen with age, and an important portion of the these gaps is related to the presence of children. By comparing acceptable job characteristics with realized outcomes, they estimate that women have a higher distaste for commute, leading them to trade-off a higher portion of potential earnings for being able to work closer to their homes. Model calibration for men and women with different household compositions predicts that gender gaps in the distaste for commuting explain around 10 % of wage gaps.

[Cortés et al. \(2023\)](#) offer a novel perspective on how gender differences in risk preferences and beliefs affect the types of jobs that men and women choose to accept. Focusing on recent graduates from Boston University, they find that women have lower reservation wages on average, and as a consequence, they tend to accept their first job upon graduation sooner than men, albeit with lower entry wages than that of a comparable male. An important portion of the gender gap in reservation wages (and accepted wages) is accounted for by higher risk tolerance and over-optimism about job search prospects among men. While the focus of the [Cortés et al. \(2023\)](#) study is on young, mostly childless, individuals, gender differences in job search may be amplified by the presence of children and care responsibilities, whether current or in expectation. [D'Angelis \(2023\)](#) shows that due to their higher willingness to pay for the amenity, college-educated millennial women's search for employers that offer parental leave can contribute to the early-career growth of the gender wage gap, well before having children. Relatedly, [Skandalis and Philippe \(2024\)](#) estimate that jobless mothers make fewer job applications than women without children because they are both more selective on acceptable job attributes and bear a higher opportunity cost of time spent on search.

5.3 Monopsonistic labor markets

Several of the mechanisms discussed can be interpreted through the lens of standard models of compensating differentials in perfectly competitive labor markets. In these models, wage differentials exactly compensate for the value of non-wage job attributes, such that different jobs provide the same level of utility to (equally productive) men and women.

At the same time, similar mechanisms can be exacerbated in monopsonistic labor markets, in which employers have significant market power in setting wages and working conditions, and gender differences in family-related constraints may provide employers with higher market power on female employees. Interestingly, one of the first explanations of gender gaps can be found in [Robinson \(1933\)](#)'s treaty on monopsony, where she notes that wage

discrimination between equally productive men and women can arise whenever their “conditions of [labor] supply are different” (p. 302–304). [Robinson \(1933\)](#)’s builds her argument about gender differences in labor supply on a model in which men are organized in a trade union and women are not, but this argument can be easily generalized to the case in which men and women differ in their evaluation of non-wage job attributes.

The distinction between competitive models of compensating differentials and models of monopsonistic labor markets rests on the behavior of labor supply. In perfectly competitive models with heterogeneous working conditions, labor supply is infinitely elastic to utility differentials, hence utility is equalized across employers, and wages are unrelated to labor supply to the individual employer because they are fully compensated by non-wage attributes. In a monopsonistic labor market, labor supply is only imperfectly elastic to utility and utility differentials across jobs persist in equilibrium. In this case, variation in wages does predict labor supply to the individual employer, and the wage elasticity of labor supply is inversely related to employer market power and to the markdown of wages below the marginal product of labor. There is extensive evidence that all margins to labor supply significantly respond to wages ([Manning, 2021](#); [Sokolova and Sorensen, 2021](#)), in support of the idea that labor markets are imperfectly competitive because employers have considerable monopsony power over workers.

In the household model of [Section 3](#), we have posited that women possibly face monopsonistic labor markets (with $1 - \varphi > 0$ denoting the wage markdown), while men are paid their marginal product. In a more general scenario, both genders may be paid below their marginal product, but women (especially mothers) face larger markdowns than men because their labor supply is less elastic to a firm’s wage. [Manning \(2003, ch. 7\)](#) contains early evidence on gender differentials in labor market transitions. In particular, women with children in the UK are more likely than any other group to report that family commitments hinder their job search and prevent them from moving jobs. Conditional on moving jobs, wage returns tend to be lower for women than for men, but gains in terms of non-pecuniary factors are higher.¹¹ While [Manning \(2003\)](#) does not detect clear-cut evidence of gender differentials in the elasticity of job separations to the wage in the UK, [Barth and Dale-Olsen \(2009\)](#) finds that women’s job separation in Norway are less responsive than men’s separations to firm-level wage premia.¹²

As highlighted by [Sokolova and Sorensen \(2021\)](#) and [Caldwell et al. \(2024\)](#), one of the main challenges in estimating labor supply elasticities is identifying credible variation in wages, i.e. cases of exogenous wage changes

¹¹ See also [Petrongolo and Ronchi \(2020\)](#) for evidence of differential gender gains in terms of geographic proximity to work.

¹² Qualitatively similar results are shown by [Hirsch et al. \(2010\)](#); [Ransom and Sims \(2010\)](#); [Webber \(2016\)](#).

that would not involve an endogenous adjustment of job amenities, recruitment effort or selectivity. The growing availability of field experiments and matched worker-firm data has much improved the reliability of elasticity estimates, although evidence on gender differences is to date scant.

In the experimental approach, [Caldwell and Oehlsen \(2023\)](#) offer a random sample of Uber drivers an earning premium of 10–50 % per trip for a week. Some of these drivers also have access to the competitor drive-share platform Lyft, providing variation in outside options. Their result suggest that women's labor supply is not less elastic to the firm than men's, and their labor supply to the market as a whole is more elastic. While these results imply that employers in the gig-economy do not have incentives to pay women below men, other factors may play a role in less-flexible set-ups.

[Sharma \(2024\)](#) investigates gender differences in labor supply elasticity in Brazilian manufacturing sector. She leverages firm-specific demand shocks, represented by the end of the Multi-Fiber Agreement in 2005, which lifted export quotas on very specific textile products from China to several high-income countries, and concurrently caused a 20 % fall of Brazilian exports of these products. The MFA expiry caused an equivalent decline of male and female wages in China-competing firms. However, men were substantially more likely to leave those firms than women, and their wages eventually recovered, while women's wages remained persistently lower. Differential separation elasticities would drive a 18 % gender wage gap among equally productive workers, explained in roughly equal proportions by women's stronger idiosyncratic preferences for their current employer and the higher concentration of their outside options.

The renewed interest in the consequences of monopsonistic labor markets for (gender) inequality has naturally called for direct evidence on the role played by firms. Firm-specific pay premia contribute to the gender wage gap whenever women sort into lower-paying firms and/or appropriate a smaller share of the firm-specific surplus than men. [Card et al. \(2015\)](#) quantify these channels by introducing gender-specific wage premia in the two-way fixed-effects framework of [Abowd et al. \(1999\)](#). Using matched employer-employee data for Portugal, they find that differential sorting and rent sharing mechanisms jointly explain about one fifth of the gender wage gap. [Morchio and Moser \(2023\)](#) propose microfoundations for worker sorting and wage setting in the [Card et al. \(2015\)](#) framework, based on a combination of compensating differentials, taste-based discrimination, and monopsony power. The key to identify the role of these components on matched employer-employee data for Brazil is a revealed-preference interpretation of worker flows. By comparing gender-specific utility at each firm to firm-level pay, they recover the gender-specific amenity values at each firm. Their results indicate that compensating differentials explain the bulk of gender wage gaps, implying that higher-ranked employers for men mostly offer higher wages, while for women they mostly

offer better amenities. By contrast, the utility differentials associated to job sorting appear to be small.

Some papers in this stream aim to directly identify the role of family-friendly working conditions as “productive” amenities for employers. [Goldin et al. \(2020\)](#) posit that firms have an incentive to offer paid parental leave to their employees whenever they invest sufficiently in firm-specific human capital, whose rewards attract them back at work at the end of their leave. Indeed, firms that provide paid leave in the US tend to be larger, with relatively younger workforces, operating in industries with higher incidence of on-the-job training. [Liu et al. \(2022\)](#) document that voluntary provision of paid parental leave – though not gender-neutral benefits – is also negatively correlated with the share of college-educated women in an industry, suggesting that employers offer female-friendly benefits to attract women in contexts where female talent is relatively scarce.

Using matched employer-employee data for Germany, [Costas-Fernandez et al. \(2024\)](#) complement existing evidence on firm incentives with an analysis of labor supply responses and show that firms offering childcare to employees have a higher share of returning mothers after maternity leave, especially so for high-wage mothers, who are presumably more difficult to replace. [Corradini et al. \(2024\)](#) consider changes in job amenities induced by a collective bargaining reform in Brazil that prioritized women’s needs, with an emphasis on paid maternity leave, childcare, and flexible work schedules. They find that firms treated by the reform saw a marked improvement in female-centric amenities, together with increased female hires and improved retention. Importantly, gains for women were realized without a trade-off in their wages, or in male employment and earnings, or even firm profitability. The interpretation is that the reform refocused unions’ priorities on pareto improvements that would not have gained enough support in the aggregation of workers’ interests at baseline.

6 The role of identity and norms in understanding gender inequalities

The disparate impact that parenthood has on the careers of women relative to men suggests that gender inequality in the labor market likely has its roots in gender roles within the household which are shaped by wider societal norms. In the presence of gender norms that dictate the appropriate role of women in society relative to men, deviating from the prescribed behavior of one’s social category is costly, thus imposing constraints on individuals’ behavior ([Akerlof and Kranton, 2000](#)).

In our model, as described in [Section 3.1.1](#), gender norms affect household utility through the parameter η that determines the utility value that the household places on the wives’ time at home. This simple representation of gender norms can serve to illustrate why work-family issues remain largely a

“woman’s problem” despite the converging economic roles of men and women in society. The model also provides some intuition as how η affects aggregate women’s labor force participation (and vice versa) in an economy, and how it can evolve dynamically over time and space as a result of social transmission mechanisms.

This section reviews the empirical evidence on the quantitative relevance of gender norms on economic behavior, followed by a discussion of what drives the formation, evolution, and transmission of gender norms.

6.1 Relevance for labor supply and household specialization

Among the “new classes of explanations” that [Bertrand \(2011\)](#) highlighted in the previous Handbook chapter, the role played by gender norms in explaining persisting gender gaps has attracted, by far, the most attention among economists in the past decade. Building on the theoretical foundations laid out by [Akerlof and Kranton \(2000\)](#) – where identity considerations are modeled to directly enter an individual’s utility function – earlier papers in this stream have sought to provide direct tests of the relevance of the gender identity model for understanding women’s relative outcomes. [Bertrand et al. \(2015\)](#) focus on the behavioral prescription that “a man should earn more than his wife” and show that adherence to this norm has wide-ranging economic and social consequences. The authors show, using administrative earnings data from the U.S., that the distribution of the share of household income earned by the wife exhibits a sharp drop-off at 0.5 – i.e., when the wife starts to out-earn her husband, consistent with the existence of gender identity norms that induce an aversion to a situation where the wife earns more than her husband.

The authors explore other potential manifestations of the norm in marriage formation, wives’ labor market outcomes, marital satisfaction, and the division of home production. They find that within marriage markets over time, when potential wives are more likely to out-earn potential husbands, the marriage rate declines. Looking within couples, when the wife’s potential income is more likely to exceed the husband’s, her labor supply is reduced and, even if she does work, her realized earnings falls further from her full earnings potential. The authors argue that this is consistent with the wife distorting her labor supply to avoid a gender-role reversal and appear “less threatening.” Finally, couples where the wife earns more than the husband are less happy, stable, and ultimately more likely to divorce. Moreover, in such couples, wives take on a greater share of the household, possibly to assuage their partner’s unease with the situation.

The findings on relative earnings and marriage durability are consistent with [Folke and Rickne \(2020\)](#) who exploit close elections as a source of plausibly exogenous variation in job promotions for politicians and show that promotions to top jobs substantially increases the likelihood of divorce for women relative to men in Sweden. They provide descriptive evidence that

similar results hold in the corporate sector for job promotions to CEO. Consistent with an identity-based channel, these effects are largely concentrated among gender-traditional couples where the promotion represents a larger deviation from initial gender role expectations at the time of marriage. Such trade-offs between career and marriage might explain why women continue to remain underrepresented in top jobs and leadership positions.

[Bursztyn et al. \(2017\)](#) provide further evidence supporting the idea that women might avoid actions that advance their careers due to perceived or actual trade-offs between marriage and career. Focusing on MBA students at UCLA, the authors first show that while married and unmarried women have similar grades on course components that are unobservable to other students such as exams and assignments, unmarried women have systematically lower participation grades. These descriptive patterns are consistent with unmarried female students downplaying their ability and ambition in the classroom setting to avoid signaling traits that might reduce their desirability as potential marriage partners. The authors provide direct evidence on this apparent trade-off with a field experiment using a real-stakes questionnaire on job preferences and personality traits that newly admitted students are required to complete for internship placement. Students were randomized into a “public” condition where they were told their answers would be discussed in the career class, and a “private” condition where they were told instead that their anonymized answers would be discussed. The authors find that single female students report less ambitious career goals and leadership attributes in the public condition, whereas neither non-single women nor men’s (regardless of relationship status) differed across the two conditions.

More recently, a related line of work by [Ichino et al. \(forthcoming\)](#) infers the strength of gender norms by studying how the spousal division of childcare responds to changes in the marginal tax rate faced by each spouse. Building on a household model in which spouses jointly choose to invest their time in market work and childcare, the authors point out that, under some assumptions, the degree of substitutability of spousal inputs in childcare is a key parameter that captures the strength of norms. Lower substitutability implies that couples have stronger preferences regarding specific combinations of spousal inputs in childcare and are less willing to reallocate their time when relative wage rates change, thereby sacrificing total household income. Exploiting variation in wage rates from Swedish tax reforms, the authors estimate elasticity parameters for different groups of couples that likely differ in their attitudes toward gender roles. They find that the allocation of home production among immigrant groups from countries with more traditional gender norms tend to react more strongly to a reduction in the husband’s than the wife’s tax rate. Native couples, on the other hand, have more symmetric responses. Taken to a larger scale, these findings imply that public intervention would face an uphill struggle in tackling gender inequalities whenever individual responses are mediated by conservative norms. Relatedly, [Giommoni and Rubolino \(2022\)](#)

examine bunching responses to an Italian tax policy that grants a credit to the main earner when the second earner reports income below a cutoff. They find that second earner women maximize family income by bunching at the cutoff, while second earner men do not, with more pronounced gender difference in bunching among immigrants/natives from more gender-traditional countries/municipalities. Overall, these findings highlight that, in the presence of binding gender norms, couples appear to be willing to incur considerable monetary costs to comply with these norms.

Other papers have taken a different approach to assess the role of norms by asking whether standard theories of comparative advantage and household specialization can explain observed gender inequality in the household and the labor market. Using rich time-use data collected from all members of a household as part of the Household, Income, and Labor Dynamics in Australia (HILDA) survey, [Siminski and Yetsenga \(2022\)](#) develop new measures of within-household specialization and test the predictions of a formal Beckerian domestic production model. They show that women do more domestic work than their male spouse at every point in the support of the relative wage distribution, and that the allocation of domestic work within the household is only weakly related to relative wages. Overall, they find that comparative advantage plays little to no role in the sexual division of labor within couple households. Relatedly, [Andresen and Nix \(2022\)](#) show that controlling for measures of predetermined relative labor market productivity differences between spouses as a proxy for comparative advantage does not eliminate motherhood penalties among heterosexual couples.

6.2 Stereotypes, beliefs, and discrimination

The relevance of norms for explaining persistent gender gaps in the household and labor market has gained increasing traction among economists. Much of the work focuses on a supply-side interpretation, where prevailing norms and stereotypes act as constraints to women's (and men's) decisions within the household and the labor market. Nevertheless, our understanding of the wider implications of stereotypes and norms on preferences and skills, and ultimately, its overall quantitative importance, remains lacking. [Bertrand \(2020\)](#) discusses in her 2020 AEA presidential address, the very nature of norms implies that individual decisions are shaped by powerful stereotypes about gender-specific roles and attributes. These stereotypes are not only descriptive, but prescriptive, and directly affects one's self-image, shaping preferences over what is appropriate given prescribed behaviors associated with one's gender group. The broader implication is that such stereotypes tend to be self-fulfilling, with men and women adapting their behavior to what is expected from their gender group, either consciously or unconsciously.

Viewed from this perspective, one has to be careful when attributing differences in choices and outcomes between men and women to observed

differences in skills, traits, or “preferences,” as these could themselves be shaped by prevailing stereotypes and norms (Lundberg, 2022). Assessing the extent to which underlying differences across genders along these dimensions are intrinsic or socially conditioned matters crucially for our interpretation of gender inequality and the design of policies to tackle the remaining gaps. Future research along these lines, particularly drawing on insights from related disciplines such as social psychology, would be highly valuable.

In the presence of powerful norms and stereotypes, a related conceptual challenge that labor economists have to grapple with is the distinction between “choice” and differential treatment, which economists have traditionally labeled as “discrimination.” Lundberg (2022) argues that to the extent that observed choices are the outcome of differential treatment or socialization by parents, schools, and society even before boys and girls enter the labor market, the discrimination versus choice dichotomy does not make sense either conceptually or empirically. Moreover, to the extent that prevailing stereotypes about gender-specific roles and attributes serve as a basis for employer discrimination either statistically or because of taste (e.g., violation of identity norms as suggested by Akerlof and Kranton (2000)), discrimination and norms may in fact be more appropriately viewed as two sides of the same coin.

Along these lines, a recent strand of work has started to emerge to explore how different types of discrimination and gender stereotyping affect women’s economic progress by hindering not only the allocation of talent across the occupational distribution but also women’s career advancement.

Differential gender access to the labor market may persist due to discriminatory beliefs about women diluting the prestige of male-dominated occupations (e.g., Goldin (2014)’s pollution theory). Greenberg et al. (2024) test this theory by studying women’s integration into combat and leadership roles in the U.S. Army, following the 2016 end of the Ground Combat Exclusion Policy for women. Using detailed personnel and survey data, they show that integrating women into previously all-male units does not negatively affect men’s or the unit’s personnel outcomes (e.g., retention, promotions, separations for misconduct). However, it does lead to a negative shift in male soldiers’ perceptions of workplace quality.

Overt or unintentional discrimination by employers, managers, and supervisors, or their beliefs regarding gender differences in the “treatment” effect of having children, may contribute to the divergent earnings trajectories of mothers and fathers as described in Section 4.2. Mothers could be deliberately passed over for promotions, or supervisors could engage in “sexist paternalism,” which, while intended to protect, actually harms them (Buchmann et al., 2024). By contrast, fathers (and married men in general) may be rewarded in the labor market based on perceptions of fairness or personal preference. For example, using data from a U.S. manufacturing firm in 1976, Korenman and Neumark (1991) found that men’s earnings profiles steepen

after marriage, which is linked to married men receiving higher performance evaluations from supervisors.

Attribution bias regarding women's ability to perform in historically male dominated leadership positions has been identified in fields like business, finance, and medicine. [Landsman \(2019\)](#) finds evidence of a gender punishment gap among S&P 1500 executives: following poor firm performance, female executives are more likely to lose their positions compared to male executives. [Sarsons \(2024\)](#) uses detailed Medicare data to test whether referring physicians assess patients' surgical outcomes differently depending on the surgeon's gender. She documents an asymmetric treatment of negative outcomes among U.S. surgeons, with female surgeons experiencing a larger drop in patient referrals relative to their male counterparts after a patient death. Conversely, male surgeons receive a larger increase in patient referrals following positive surgical outcomes. Sarsons' study is one of the most convincing in this area, as her data allow for the control of factors like patient and procedure risk and surgeon experience, isolating the gender-driven portion of these biases. [Egan et al. \(2022\)](#) find a similar asymmetric punishment gap in the financial advisory industry. Following an incident of misconduct, female advisers are 20 % more likely to lose their jobs and 30 % less likely to find new employment compared to male advisers. The study finds that this gap is not driven by gender differences in occupation, productivity, the nature of the misconduct, or recidivism. As with Sarsons' work, the study shows limited evidence that the punishment gap is driven by rational or Bayesian profit maximization. For example, the gap in hiring and firing diminishes in firms with a greater percentage of female managers and executives.

Differential treatment by employers could also extend to men when they deviate from commonly accepted behavior, making it equally challenging for them to engage in gender atypical behavior. Using a survey experiment and a large-scale audit study, [Weisshaar \(2018\)](#) find that fathers face a higher penalty when they take time off work to care for family relative to mothers, especially in tight labor markets, which she argues is due to a violation of "ideal worker" norms which are more rigidly applied to men than women. This might explain why in many countries, men are reluctant to take paternity leave even when such benefits are extended to them. Competition within the workplace, either real or perceived, could further exacerbate such mechanisms especially when workplace performance is tied to in-person presence or visibility. This is nicely illustrated by [Johnson et al. \(2024\)](#) who study the career impacts of a worker when his competitors (i.e., male workers in the same firm) take paternity leave because of their eligibility under a policy reform in Norway. The authors show that focal workers' earnings increase in the years post-birth when the share of leave-eligible competitors increase; however, the focal worker's own leave has no direct effect on his earnings trajectory as long as his competitors also take leave.

Another strand of work examines the role of stereotypes as a form of pre-market discrimination largely in the context of the gender gap in educational choices. Several studies have documented that adults shape gender-appropriate behavior in children, affecting their choices, preferences, and beliefs about their ability. For example, [Carlana \(2019\)](#) shows that assignment to math teachers with stronger implicit gender stereotypes widens the gender gap in math performance and leads girls to select into less demanding high schools. These effects are driven, at least in part, by girls' lower self-confidence in their math ability when exposed to gender-biased teachers. [Nosek et al. \(2009\)](#) and [Nollenberger et al. \(2016\)](#) document that gender attitudes matter for the gender gap in math and science performance among children across countries and across immigrant groups within the U.S., respectively. Other studies find that female role models are effective in encouraging women to major in economics ([Porter and Serra, 2020](#)), participate in STEM-related activities ([Del Carpio and Guadalupe, 2021](#)), and enroll in selective and male-dominated STEM programs in college ([Breda et al., 2023](#)), suggesting that women's lower preferences for STEM education and careers are likely to be socially constructed.

Such pre-market discrimination could in itself be a reaction to anticipated discrimination in the labor market and marriage market. [Manning and Swaffield \(2008\)](#) and [Kaestner and Malamud \(2023\)](#) provide some evidence in line with women and men experiencing differential treatment in the labor market when they deviate from commonly accepted behavior. Specifically, using data from the NLSY, [Kaestner and Malamud \(2023\)](#) show that women characterized as "headstrong" and boys who were considered as "dependent" when they were children experienced earnings penalties as adults, all else equal. In terms of the marriage market, [Wiswall and Zafar \(2021\)](#) find that women perceive a marriage market penalty to completing a degree in science or business, relative to a humanities or social science degree, and that such family expectations are particularly important for women's major choices. This is similar in spirit to [Bursztyn et al. \(2017\)](#)'s finding that MBA women choose to avoid public expressions of career ambition due to concerns that this would depress their marriage market prospects.

The emerging work suggests that the presence of gendered expectations and incentives invariably sets up a self-perpetuating cycle where demand-side and supply-side considerations – buttressed by stereotypes and norms – reinforce each other to impact preferences, skill investment decisions, and the labor market choices of women and men.

6.3 What drives gender norms and how malleable are they?

The growing recognition of the empirical relevance of gender norms for understanding gender inequality has brought to the fore the question of what drives the formation and evolution of gender norms.

6.3.1 *Historical origins and persistence*

A relatively large literature has established the historical origins of gender norms, and show how cultural persistence can lead to the stickiness of norms over long periods of time. This literature provides an indication of historical conditions that shape gender-role attitudes, including agricultural practices that promotes specialization along gender lines, changes in the relative demand for female labor, and bargaining in the marriage market.

One of the earliest papers in this stream, [Alesina et al. \(2013\)](#) demonstrate how traditional agricultural practices influenced historical gender roles and led to long-term persistence in female labor participation. Exploiting variation in historical geo-climatic conditions for growing crops using the plough versus shifting cultivation, the authors find that among ethnicities and countries whose ancestors practiced physical strength-intensive plough cultivation, which tended to favor male labor, women were historically less likely to participate in farm work, and, today, have lower rates of female labor force participation and hold less progressive gender-role attitudes. Follow-up work by [Hansen et al. \(2015\)](#) studies the role of agricultural history more generally and finds that societies with longer histories of agriculture (i.e., earlier Neolithic revolutions) have lower female labor force participation and less equal gender roles today. The authors argue that these patterns are likely driven by a combination of higher fertility in societies with longer agricultural histories due to greater technological advancement as well as the transition to cereal agriculture resulting in greater household specialization where women predominantly engaged in child-rearing and cereal processing.

Other papers in this stream have explored the historical role of uneven sex ratios and changes in the value of women's work. [Grosjean and Khattar \(2019\)](#) study the long-run impacts of historical male-biased sex ratios induced by the resettlement of convicts to Australia, and show that areas with more male-biased sex ratios historically are characterized by more traditional gender-role attitudes and greater gender inequality in the labor market in the present day, well after sex ratios are back to the natural rate. [Xue \(2023\)](#) explores how the cotton revolution in imperial China, which led to a sharp increase in high-value work opportunities for women, affected cultural beliefs about women's worth. Exploiting variation across counties in premodern cotton textile production, generated by weather-suitability for cotton weaving and distance from the national market, she finds that areas with higher premodern cotton textile production had lower sex ratio at birth in 2000, stronger position of women in the household, and more progressive gender-role attitudes. A common thread across these studies is the emphasis on the role of vertical cultural transmission in sustaining long-term persistence across generations ([Bisin and Verdier, 2001](#)).

6.3.2 *Cultural change and learning*

That historical forces continue to shape patterns of gender norms today can help to explain the stickiness of gender norms even as economic conditions change. Yet,

throughout history, there have been numerous instances where gender norms have changed relatively quickly in response to technological innovations, economic development, and changes in the social and political landscape. For example, in the case of the U.S., innovations in contraception, widespread adoption of home production technologies, improvements in maternal health, and the availability of substitutes to maternal inputs such as infant formula, provided women with greater ability to plan childbearing, reconcile work and domestic responsibilities, and invest in education and their careers (Goldin and Katz, 2002; Greenwood et al., 2005; Albanesi and Olivetti, 2016). As Goldin (2006) argues in her 2006 Ely Lecture, these changes, coupled with legislative changes that removed explicit barriers to women's work (e.g., marriage bars) and antidiscrimination legislation, are likely to have contributed to the altering of women's identity and changing gender roles beginning in the 1960s and accelerating from the 1970s onwards. Political institutions have also been shown to be an important driver of the adoption of new norms. Several studies show how exposure to state socialism – which promoted women's economic inclusion – has led to the adoption of more progressive gender-role attitudes, increased women's preferences for work, and altered gender roles within the household (Beblo and Görges, 2018; Campa and Serafinelli, 2019; Senik et al., 2020).

A few recent papers focus on the role of public policies in shifting gender norms. Bastian (2020) shows that, by boosting maternal employment, the introduction of the Earned Income Tax Credit (EITC) led to higher approval of working women. Examining the intergenerational effects of the introduction of paternity leave in Spain, Farre et al. (2023) find that children born after the policy change exhibit more gender egalitarian attitudes, engage more in counter-stereotypical household tasks, and are more likely to report future expectations regarding their own work and family choices that deviate from the traditional male-breadwinner model. The authors attribute these effects to children's exposure to greater involvement in childcare by fathers and greater willingness of mothers to return to work after childbirth that resulted from fathers' take-up of paternity leave due to the reform (Farre and Gonzalez, 2019). Other papers focus on school-based interventions or curriculum and show that these have the potential to shift norms. For example, Dhar et al. (2022) evaluate a randomized intervention in India that engaged adolescent boys and girls in classroom discussions about gender equality. The authors find that the program led to a persistent increase in progressive gender attitudes and self-reports of more gender-equal behavior. Hara and Rodriguez-Planas (2023) show that a Japanese educational reform that eliminated gender-typed and gender-segregated classes in industrial arts and home economics in junior high schools led to changes in beliefs regarding men's and women's gender roles and a shift toward less gendered specialization of tasks within the household.

While these papers identify particular forces that shape gender norms at a particular point in time in a given society, the underlying mechanisms that generate widespread cultural change or why cultural change happens more rapidly in some societies but not others is less well-understood. Motivated by

the S-shape patterns for female labor force participation and gender-role attitudes from 1940 to 2000, [Fernández \(2013\)](#) proposes a model of cultural change to explain the evolution of social beliefs in the U.S. In her model, cultural change results from a rational, intergenerational learning process where individuals with heterogeneous beliefs about married women's long-run payoff from working update their beliefs by observing the labor supply behavior of women in the preceding generation. Calibrating the model to key statistics for 1980 to 2000, the author finds that the model is able to replicate the dynamic path of married women's labor force participation from 1880 to 2000 and that the paths of both beliefs and earnings had an important role to play in the dramatic evolution of women's work over the past century.

[Fogli and Veldkamp \(2011\)](#) propose a related model where women learn about the effects of maternal employment on children by observing employed women nearby. When few women work, there is little information and participation rises slowly. As information accumulates in some regions, this reduces the uncertainty of the effects of maternal employment, leading to more women in those regions participating. Within these regions, learning accelerates, labor force participation rises faster, and regional participation diverges. Eventually, as information diffuses throughout the economy, beliefs converge to the truth, participation plateaus, and regions become more similar. Similar to [Fernández \(2013\)](#), the calibrated model delivers an S-shaped evolution of aggregate female labor force participation and gender-role attitudes. In addition, the local nature of the learning process generates geographically heterogeneous, but locally correlated reactions, similar to that observed in the data.

[Giuliano and Nunn \(2021\)](#) offer some insights into the question of why culture persists in some cases but not others by testing the empirical relevance of a key determinant that has emerged from the theoretical evolutionary anthropology literature – the stability of a society's environment across generations. The idea is that if the environment is very similar, cultural values and beliefs that have evolved and survived are likely to contain information that is relevant for the current generation; by contrast, if the environment changes a lot from one generation to another, the previous generation's values and beliefs are less likely to be relevant for the current generation. Using cross-generational variability in climate conditions and a variety of samples and empirical strategies, the authors show that populations with ancestors who lived in environments with greater climate instability place less emphasis on maintaining tradition and exhibit less persistence in cultural norms, including gender norms.

While these papers make some inroads in tackling the question of how and why gender norms change, there remains much scope for future work. For example, we still have limited understanding of how gender norms evolve in the face of market forces that are making these norms increasingly costly, the types of gender norms that are likely to change or become relevant as the economic and social environment changes, and what it takes to precipitate and sustain widespread cultural change.

6.3.3 *Transmission channels*

How beliefs are formed and transmitted are key to understanding why gender norms are persistent as well as how cultural change can be achieved. The literature typically emphasizes three forms of cultural transmission: vertical transmission from parents to children, oblique transmission based on non-parental and non-peer elders (e.g., role models or teachers), and horizontal transmission from peers.

There is growing evidence on the relevance of each of these channels for the transmission of gender norms and preference formation. The large literature on the relevance of origin country norms on the work and fertility preferences of second-generation immigrants implicitly assumes that the family is the primary channel through which norms are transmitted, reinforced perhaps by ethnic social networks (e.g., Fernández et al., 2009; Blau et al., 2013; Di Miceli, 2019). Several studies examine the vertical transmission mechanism more directly. For example, exploiting variation in the mobilization rates of men across U.S. states during WWII as a shock to mothers' labor force participation, Fernández et al. (2004) show that men whose mothers worked are more likely to have working wives. Interestingly, they find little evidence that married women's work behavior is affected by whether her mother works. Using linked parent and child surveys, Farré and Vella (2013) and Bertrand (2019) show that mothers' gender role attitudes and exposure to non-traditional family types are associated with their children's gender-role attitudes as young adults, and daughters' work behavior. Within the family, sibling sex composition has also been shown to affect gender-role attitudes and behavior. Women with a brother rather than a sister tend to hold more traditional family attitudes and exhibit a greater degree of gender conformity in their occupation and partner choice (Rao and Chatterjee, 2018; Cools and Patacchini, 2019; Healy and Malhotra, 2013) likely due to sex-typing in mixed-gender sibships or differential parental investments. Using time-use data from Denmark, Brenøe (forthcoming) provide some suggestive evidence of such "gendered-parenting" among children with an opposite-sex sibling.

Apart from vertical transmission from parents to children, several studies have also documented the oblique transmission of norms and counter-stereotypical behavior from non-parental elders within social groups. Exploiting variation in the employment status of mothers across different cohorts of students within a high school, Olivetti et al. (2020) show that adolescent women who were exposed to more peers with working mothers are less likely to feel that work interferes with family responsibilities, and are more likely to work for pay when they have children. This gender-role socialization effects through peers' mothers is above and beyond the effects of their own mothers. In addition, the papers discussed in Section 6.2 on the influence of female role models and teachers' stereotypical attitudes on the gender gap in STEM are further examples of the oblique transmission of norms and beliefs.

Finally, the horizontal transmission of norms through peers can amplify initial changes in the behavior of a given social group, facilitating sustained changes in norms over time. Maurin and Moschion (2009) and Mota et al. (2016) provide evidence from France that married mothers' decision to participate in the labor market is influenced by the labor supply behavior of other mothers living in the same neighborhood. Relatedly, Nicoletti et al. (2018) use Norwegian administrative data to examine the causal influence of the family network on mothers' labor supply decisions and find that cousins and sisters significantly affect the number of hours worked by mothers of preschool children. Moreover, they find, perhaps not surprisingly, that family peers have a stronger effect than neighborhood peers. The authors perform a back-of-the-envelope calculation to quantify the family peer effect and estimate a social multiplier factor of 1.5 for a given direct increase in labor supply. As emphasized by Fogli and Veldkamp (2011) and Fernández (2013), such horizontal transmission of norms provides a natural explanation for the large variation in labor supply behavior across subgroups of workers, geography, and over time.

6.3.4 *Information gaps*

It remains quite puzzling that gender norms remain persistent even in the face of evolving market forces that are making the adherence to these norms increasingly costly. An emerging strand of work suggests that systematic misperceptions of others' views toward counter-stereotypical behaviors can help explain why norms remain sticky even when the economic or social environment changes. Such a situation where most people personally reject a norm, but they incorrectly believe that most others accept the norm, and end up adhering to the norm because of the fear of social sanctions is referred to as "pluralistic ignorance" by social psychologists.

One of the first papers to examine the empirical relevance of misperceptions for understanding economic behavior is Bursztyn et al. (2020)'s study of female labor force participation in Saudi Arabia. The authors draw on several surveys to document that the vast majority of young married men in Saudi Arabia privately support women working outside their home, yet substantially underestimate the level of support by their peers. They provide evidence that experimentally correcting these beliefs increases men's willingness to let their wives to search for jobs and, consequently, increases the likelihood that their wives applied and interviewed for a job outside the home. A closely related study by Cameron et al. (2024) in Indonesia documents similar misperceptions in women's (but not men's) support for working women and the level of support among men for sharing childcare and that providing information about the true level of support in the community increased both genders' support for working women, especially among men whose wives were not working for respondents with school-aged children. Cortés et al. (forthcoming) show that misperceptions regarding the support for mothers' participation in the labor market contribute to the stickiness of gender norms even in the U.S., where the prevailing norm is much less extreme and women and men have similar access

to education and labor market opportunities.¹³ Such misperceptions can also potentially explain the slow take-up of policies aimed at directly trying to counter stereotypes such as encouraging paternity leave-taking among fathers (Miyajima and Yamaguchi, 2017).

The finding that misperceptions regarding gender norms are widespread and likely constrain the adjustment of prevailing norms and behavior in a way that reflects the true sentiment and beliefs of the population suggests a seemingly straightforward and concrete solution to facilitating the shift in gender norms – providing information. While the findings from the above-mentioned studies indeed suggest that such an approach is promising, research in this area is still in its infancy. Existing studies have largely focused on short-run changes in attitude and beliefs and, at most, medium term changes in intermediate outcomes related to the job search process; more evidence on the effectiveness of information provision in shifting actual behavior is still needed. Moreover, while economists have typically favored information provision in the form of simple messaging directed toward an individual, in the context of shifting norms, it is unclear that such a light-touch approach is sufficient to credibly and meaningfully address widely-held misperceptions. Social psychologists have long emphasized that norms exist within group processes and individuals are most likely to update their beliefs about the group norm if they can directly observe that stated views and actions of their peers through the dynamic processes of social proof and reality testing (e.g., Cialdini et al., 1999; Van Kleef et al., 2019). Providing support for this view, Dahl et al. (2014) document considerable peer effects in paid paternity leave-taking in Norway – coworkers and brothers are more likely to take paternity leave if their peer is exogenously induced to take leave. Importantly, the peer effects increase over time as more individuals within the peer group observe and take-up leave. Future work that considers how to embed social norm interventions within the context of group processes could be especially promising (Prentice and Paluck, 2020).

7 Micro-macro linkages

there is nice discussion about micro-macro linkages in slides, but should we have a section on this or it could fit in one of existing sections? the macro stuff is scattered in the paper.

¹³ Follow-up work by Bursztyn et al. (2023) takes on a cross-country perspective and studies actual and perceived gender norms regarding (1) the basic rights of women to work outside the home, and (2) gender affirmative action across 60 countries. The authors establish widespread misperceptions of gender norms in these two domains around the world with the patterns in the extent and direction of the misperception depending on the type of norm and how gender-equal the country is. The authors argue that the patterns can be best explained by overweighting of the minority view and gender stereotyping.

- Structural change -DONE- footnote briefly discussing paper by Gaurav and Tatyana, [Ngai and Petrongolo, \(2017\)](#) and Ngai et al. when discussing misallocation of talent (Hsiao).
- literature on tech or med change in macro is cited in discussion of reservation wage and the relative strength of income and substitution effects
- To do (maybe) - cite political economy stuff by Michele and Mathias (women's right, link to economic growth and development) when discussing norms & policies
- To do: cite macro labor supply elasticities, larger either because of dynamic consideration (keane) or because women are more intermittent participants (rogerson). Probably we can simply cite the lit review by keane and rogerson.
- A few micro-macro strands on gender economics:
- Labor supply elasticities (keane and rogerson - dynamic of labor supply), extensive vs intensive margin (ref??) – CITE IN DISCUSSION OF MONOPSONISTIC DISCRIMINATION>.

8 Conclusion

Gender is now a mainstream topic in economics. Over the last decade or so since the last Handbook Chapter, labor economists have zeroed in on several leading explanations for the remaining gender disparities in the labor market. There is now a clearer distinction between the role of preferences and constraints in driving gender gaps, and increasing recognition that essential differences between men and women in terms of preferences, skills, and psychological attributes play, at best, a modest role in explaining the remaining gaps.

The more prominent explanations today center around the differential constraints that women, especially mothers, face relative to men in the labor market due to the trade-offs involved in seeking to balance work and family responsibilities. There is now a wealth of evidence suggesting that parenthood drives a large wedge in the career trajectories of mothers and fathers, and an emerging body of work that analyzes the precise mechanisms at play. These include studies that document gender differences in the willingness to pay for family-friendly job attributes such as workplace flexibility, shorter hours, remote work, and shorter commutes, and the implications for gender-based sorting both across and within occupations. Such preference heterogeneity has also led to a renewed interest in the extent to which women may be further disadvantaged by firms' ability to exercise monopsonistic power over female employees. Future work could explore how firms and industries can be incentivized to reorganize work in a way that diminishes the penalties associated with family-friendly work arrangements.

Much of the extensive literature on career-family trade-offs focuses on mothers, but the impact on fathers' earnings is less explored, and existing findings suggest that investigating labor market and normative determinants of the fatherhood premium, in addition to those driving motherhood penalties, may be a promising area for further research.

While recent work has crystallized the key sources of the tension between work and family for women, there appears to be forces, both on the work and the home front, that may make it even more challenging for modern cohorts of women to close the gaps. There is some indication that work is becoming “greedier,” with high-powered occupations increasingly rewarding employees who are willing to put in long and inflexible hours. The exact reasons behind these trends are unclear, but possible drivers include the use of performance pay and promotion systems, and rising inequality. Such developments tend to increase the returns to specialization within the household, putting women at a further disadvantage.

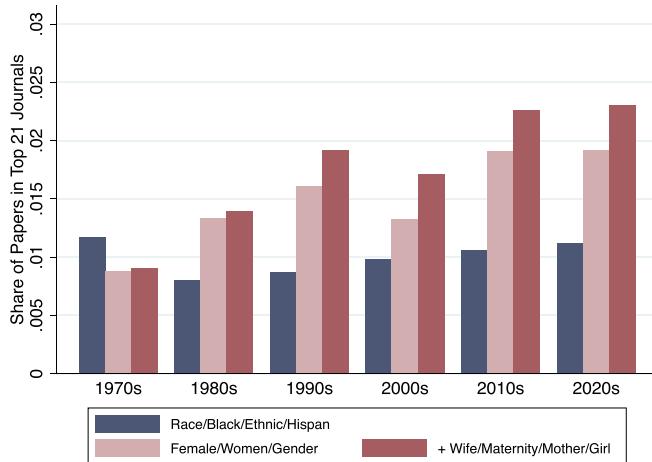
Another important labor market development has been the rise in gig work and remote working opportunities as more businesses embrace technology in the workplace. On the one hand, by increasing the substitutability of workers, penalties associated with flexibility should be minimized with gig work. Similarly, remote work opportunities should provide more flexibility to workers. Nevertheless, there are concerns that so long as women continue to be the dominant provider of childcare, such developments could serve to further entrench household specialization. The lack of structure and freedom of choice in gig work, paradoxically, also makes women even more available to take on household responsibilities. Further, productivity losses that may arise when multitasking WFH and childcare simultaneously could provide an additional channel for flexibility penalties. A clearer understanding of the sources of the returns to long (and continuous) hours in various settings – whether it is productivity-related or cultural – as well as a broader consideration of how the structure of the work environment interacts with existing gender roles would help with efforts to leverage technology in the design of more equitable workplaces.

On the family front, parenting has also become more time-intensive: over the past two decades, time spent with children has risen steadily for both parents, but especially so for college-educated mothers in the US and other high-income countries ([Kuziemko et al., 2018](#); [Borra and Sevilla, 2019](#); [Sani et al., 2016](#)). Once again, it is not entirely clear what the precise drivers of these trends are; however, some have suggested that rising parental time demands could be driven by competition for top colleges ([Ramey and Ramey, 2010](#)) and increasing returns to education and skills which in turn raise the returns to investment in children’s human capital ([Doepeke and Zilibotti, 2017](#)). Such developments are likely to lead to an increase in the real or perceived costs of motherhood, which could, in turn, affect fertility decisions.

We expect that efforts that seek to address gender disparities in the labor market will have to contend with some of these emerging challenges. Ultimately, the process for change may require a combination of efforts and policies to hasten the weakening of traditional gender norms and the promotion of more equitable workplaces and practices.

Appendix A. Gender in economic journals

(A) Classification Using Paper Title



(B) Classification Using Listed Subject(s)

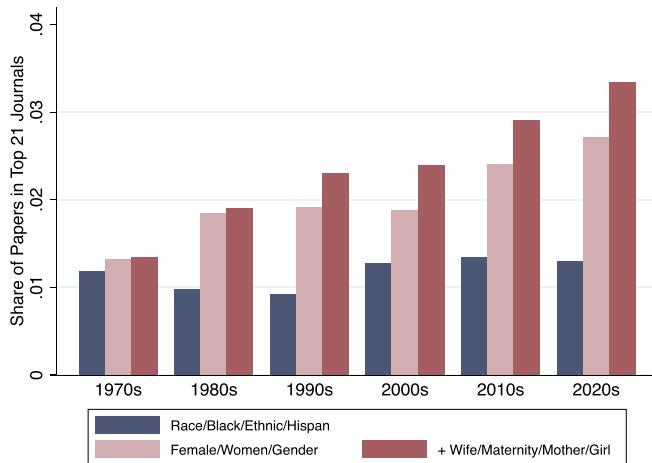


FIG. A.1 Gender vs. race papers in top 21 economics journals available in EBSCO. Note: Titles and subjects of papers are extracted from EBSCO's research database for 21 of the top 30 economics journals for which information on paper subject is available. See Table A.2 for the 21 journals that are included in this exercise. Each time period covers all papers published in the top 21 economics journals during that period. The 2020s time period is limited to the years up to 2023. The classification using paper title in Panel (A) follows the same procedure as described in Fig. 1 applied to the 21 journals. Panel (B) uses the same keyword search procedure applied to the EBSCO listed subject(s) of the paper instead of the paper title. The blue bars show the share of race-related papers; the red bars show the share of gender-related papers using different sets of gender-related keywords.

TABLE A.1 List of top 30 economics journals.

Included in Top 21 List	
Quarterly Journal of Economics	✓
American Economic Review	✓
Econometrica	✓
Review of Economic Studies	✓
Journal of Political Economy	✓
American Economic Journal: Macroeconomics	
American Economic Journal: Applied Economics	
Journal of the European Economic Association	✓
American Economic Journal: Economic Policy	
Journal of Labor Economics	✓
Theoretical Economics	
Review of Economics and Statistics	✓
Journal of Monetary Economics	✓
American Economic Journal: Microeconomics	
Journal of Human Resources	✓
Quantitative Economics	
Journal of Economic Growth	
Economic Journal	✓
RAND Journal of Economics	✓
Review of Economic Dynamics	✓
Journal of Business and Economic Statistics	✓
Journal of International Economics	✓
International Economic Review	✓
Journal of Economic Theory	✓
Journal of Public Economics	✓
Journal of Econometrics	✓
Experimental Economics	
Econometric Theory	
Journal of Development Economics	✓
Journal of Applied Econometrics	✓

TABLE A.2 Keywords used to identify topics among gender papers.

Topics	Keywords
Labor/Pay/Job	<i>Labor, labour, wage, earn, pay, work, firm, occup, wealth, unemploy, employ, particip, job, displac, busi, career, occup, salari, hire, incom, economi</i>
Family/Marriage	<i>Famili, marit, marriag, sibl, child, son, fertil, fecund, sibship, marri, household, mother, wife, matern, mate, birth, divorc, parent</i>
Education	<i>School, educ, colleg, academ, skill, stem, student, classroom, achiev, teacher, teach, math, human capit, major, scienc, engin, faculti, professor</i>
Discrimination/ Bias/Norm	<i>Discrimin, bias, stereotyp, norm, cultur, attitud, ident, gender role</i>
Development	<i>Develop, microfin, growth, tanzania, pakistan, bangladesh, bengal, china, india, nepal, africa, philippin, empower, corrupt, cast</i>
Politics/Finance/ Law	<i>Polit, suffrag, govern, vote, elect, institut, juri, ceo, financ, credit, entrepeneur, corpor, quota, lawyer, glass ceil</i>
Behavioral	<i>Risk, pressur, uncertainti, compet, stake, overconfid, selfish, altruism, generos, cooper, leadership, trust, negoti, generous, Cooperative, Co operation, willing, bargain</i>
Health	<i>Health, morbid, mortal, depress, diseas, bulim, medic, abort, hiv, pandem</i>

References

- Aaronson, D., Dehejia, R., Jordan, A., Pop-Eleches, C., Samii, C., Schulze, K., 2020. The effect of fertility on mothers' labor supply over the last two centuries. *Econ. J.* 131 (633), 1–32.
- Abowd, J.M., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. *Econometrica* 67 (2), 251–333.
- Adams, A., Huttunen, K., Nix, E., Zhang, N., The dynamics of abusive relationships. *The Quarterly Journal of Economics*, forthcoming.
- Adams, A., Jensen, M., Petrongolo, B., 2024. Birth timing and spacing: implications for parental leave dynamics and child penalties. CEPR Discussion Paper 19324.
- Adams-Prassl, A., Hara, K., Milland, K., Callison-Burch, C., 2023. The gender wage gap in an online labor market: The cost of interruptions. *Rev. Econ. Stat.* 1–23.
- Adams-Prassl, A., Huttunen, K., Nix, E., Zhang, N., 2024. Violence against women at work. *Q. J. Econ.* 139 (2), 937–991.
- Adda, J., Dustmann, C., Stevens, K., 2017. The career costs of children. *J. Political Econ.* 125 (2), 293–337.
- Ager, P., Goni, M., Salvanes, K.G., 2023. Gender-biased technological change: milking machines and the exodus of women from farming. CEPR Discussion Paper 18290.

- Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. *Q. J. Econ.* 115 (3), 715–753.
- Albanesi, S., Olivetti, C., 2016. Gender roles and medical progress. *J. Political Econ.* 124 (3), 650–695.
- Albanesi, S., Olivetti, C., Petrongolo, B., 2023. Families, labor markets, and policy. In: Lundberg, S., Voena, A. (Eds.), *Handbook of the Economics of the Family*, Volume 1. North-Holland, pp. 255–326.
- Alesina, A., Giuliano, P., Nunn, N., 2013. On the origins of gender roles: women and the plough. *Q. J. Econ.* 128 (2), 469–527.
- Altonji, J., Blank, R., 1999. Race and gender in the labor market. *Handbook of Labor Economics* 3C. Elsevier, pp. 3143–3259 chapter 48.
- Andresen, M.E., Nix, E., 2022. What causes the child penalty? Evidence from adopting and same-sex couples. *J. Labor. Econ.* 40 (4), 971–1004.
- Andrew, A., Bandiera, O., Costa Dias, M., Landais, C., 2024. Women and men at work. *Oxf. Open. Econ.* 3 (ement_1), i294–i322.
- Angelov, N., Johansson, P., Lindahl, E., 2016. Parenthood and the gender gap in pay. *J. Labor. Econ.* 34 (3), 545–579.
- Angrist, J.D., William, N.E., 1998. Children and their parents' labor supply: evidence from exogenous variation in family size. *Am. Econ. Rev.* 88 (3), 450–477.
- Ashraf, N., Bandiera, O., Minni, V., Quintas-Martínez, V., 2024. Gender gaps across the spectrum of development: local talent and firm productivity. *Tech. Rep.*
- Azmat, G., Ferrer, R., 2017. Gender gaps in performance: evidence from young lawyers. *J. Political Econ.* 125 (5), 1306–1355.
- Bandiera, O., Parekh, N., Petrongolo, B., Rao, M., 2022. Men are from mars, and women too: a Bayesian meta-analysis of overconfidence experiments. *Economica* 89 (S1), S38–S70.
- Barbanchon, T.L., Rathelot, R., Roulet, A., 2021. Gender differences in job search: trading off commute against wage. *Q. J. Econ.* 136 (1), 381–426.
- Barth, E., Dale-Olsen, H., 2009. Monopsonistic discrimination, worker turnover, and the gender wage gap. *Labour Econ.* 16 (5), 589–597.
- Bastian, J., 2020. The rise of working mothers and the 1975 earned income tax credit. *Am. Econ. J. Econ. Policy* 12 (3), 44–75.
- Beblo, M., Görge, L., 2018. On the nature of nurture. The malleability of gender differences in work preferences. *J. Econ. Behav. Organ.* 151, 19–41.
- Becker, G.S., 1960. An economic analysis of fertility. In: University-NBER (Ed.), *Demographic and Economic Change in Developed Countries*. Columbia University Press, pp. 209–240 chapter 7.
- Bertrand, M., 2011. New perspectives on gender. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics* 4B. Elsevier, pp. 1543–1590 chapter 17.
- Bertrand, M., 2019. The gender socialization of children growing up in nontraditional families. *AEA Pap. Proc.* 109, 115–121.
- Bertrand, M., 2020. Gender in the twenty-first century. *AEA Pap. Proc.* 110, 1–24.
- Bertrand, M., Goldin, C., Katz, L.F., 2010. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *Am. Econ. J. Appl. Econ.* 2 (3), 228–255.
- Bertrand, M., Kamenica, E., Pan, J., 2015. Gender identity and relative income within households. *Q. J. Econ.* 130 (2), 571–614.
- Besnes, S., Huitfeldt, I., Leuven, E., 2023. Reconciling estimates of the long-term earnings effect of fertility. *IZA Discussion Paper* 16174.
- Bindler, A., Ketel, N., 2022. Scaring or scarring? Labor market effects of criminal victimization. *J. Labor. Econ.* 40 (4), 939–970.

- Bisin, A., Verdier, T., 2001. The economics of cultural transmission and the dynamics of preferences. *J. Econ. Theory* 97, 298–319.
- Blau, F.D., Winkler, A.E., 2021. *The Economics of Women, Men, and Work*. Oxford University Press..
- Blau, F.D., Kahn, L.M., 2017. The gender wage gap: extent, trends, and explanations. *J. Econ. Lit.* 55 (3), 789–865.
- Blau, F.D., Kahn, L.M., Liu, A.Y.-H., Papps, K.L., 2013. The transmission of women's fertility, human capital, and work orientation across immigrant generations. *J. Popul. Econ.* 26 (2), 405–435.
- Bloom, N., Han, R., Liang, J., 2022. How hybrid working from home works out. NBER Working Paper 30292.
- Blundell, R., MacCurdy, T., 1999. Labor supply: a review of alternative approaches. In: Orley, C., Ashenfelter, Card, David (Eds.), *Handbook of Labor Economics* 3A. Elsevier, pp. 1559–1695 chapter 27.
- Bögl, S., Moshfegh, J., Persson, P., Polyakova, M., 2024. The economics of infertility: evidence from reproductive medicine. NBER Working Paper 32445.
- Borra, C., Sevilla, A., 2019. Competition for university places and parental time investments: evidence from the United Kingdom. *Econ. Inq.* 57 (3).
- Breda, T., Grenet, J., Monnet, M., Van Effenterre, C., 2023. How effective are female role models in steering girls towards STEM? Evidence from French High Schools. *Econ. J.* 133 (653), 1773–1809.
- Brenøe, A.A., Brothers increase women's gender conformity. *J. Population Econ.*, forthcoming.
- Bronars, S.G., Grogger, J., 1994. The economic consequences of unwed motherhood: using twin births as a natural experiment. *Am. Econ. Rev.* 84 (5), 1141–1156.
- Buchmann, N., Meyer, C., Sullivan, C.D., 2024. "Paternalistic Discrimination," Technical Report.
- Bursztyn, L., Gonzalez, A.L., Yanagizawa-Drott, D., 2020. Misperceived social norms: women working outside the home in Saudi Arabia. *Am. Econ. Rev.* 110 (10), 2997–3029.
- Bursztyn, L., Cappelen, A.W., Tungodden, B., Voena, A., Yanagizawa-Drott, D., March 2023. How are gender norms perceived. Working Paper 31049, National Bureau of Economic Research.
- Bursztyn, L., Fujiwara, T., Pallais, A., 2017. Acting wife: marriage market incentives and labor market investments. *Am. Econ. Rev.* 107 (11), 3288–3319.
- Caldwell, S., Oehlson, E., 2023. Gender, outside options, and labor supply: experimental evidence from the gig economy. University of California, Berkeley Working Paper.
- Caldwell, S., Dube, A., Naidu, S., 2024. Monopsony makes it big. Technical Report.
- Cameron, L., Contreras Suarez, D., Setyonaluri, D., 2024. Leveraging women's views to influence gender norms around women working: evidence from an online intervention in Indonesia. Policy Research Working Paper 10681, The World Bank.
- Campa, P., Serafinelli, M., 2019. Politico-economic regimes and attitudes: female workers under state socialism. *Rev. Econ. Stat.* 101 (2), 233–248.
- Card, D., Cardoso, A.R., Kline, P., 2015. Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women. *Q. J. Econ.* 131 (2), 633–686.
- Carlana, M., 2019. Implicit stereotypes: evidence from teachers' gender bias. *Q. J. Econ.* 134 (3), 1163–1224.
- Casarico, A., Lattanzio, S., 2023. Behind the child penalty: understanding what contributes to the labour market costs of motherhood. *J. Popul. Econ.* 36, 1489–1511.
- Chiplunkar, G., Goldberg, P.K., 2024. Aggregate implications of barriers to female entrepreneurship. *Econometrica* (forthcoming).

- Chiplunkar, G., Kleineberg, T., 2024. Gender barriers, structural transformation, and economic development. Technical Report, World Bank.
- Cialdini, R.B., Wosinska, W., Barrett, D.W., Butner, J., Gornik-Durose, M., 1999. Compliance with a request in two cultures: the differential influence of social proof and commitment/consistency on collectivists and individualists. *Person. Soc. Psychol. Bull.* 25 (10).
- Cook, C., Diamond, R., Hall, J.V., List, J.A., Oyer, P., 2020. The gender earnings gap in the gig economy: evidence from over a million rideshare drivers. *Rev. Econ. Stud.* 88 (5), 2210–2238.
- Cools, A., Patacchini, E., 2019. The brother earnings penalty. *Labour Econ.* 58 (C), 37–51.
- Corradini, V., Lagos, L., Sharma, G., 2024. Collective bargaining for women: how unions create female-friendly jobs. Technical Report.
- Correll, S.J., Benard, S., Paik, I., 2007. Getting a job: is there a motherhood penalty? *Am. J. Sociol.* 112 (5), 1297–1338.
- Cortés, P., Pan, J., 2019. When time binds: substitutes for household production, returns to working long hours, and the skilled gender wage gap. *J. Labor. Econ.* 37 (2), 351–398.
- Cortés, P., Kosar, G., Pan, J., Zafar, B., Should mothers work? How perceptions of the social norm affect individual attitudes toward work in the U.S. *Rev. Econ. Stat.* forthcoming.
- Cortés, P., Pan, J., Pilossoph, L., Reuben, E., Zafar, B., 2023. Gender differences in job search and the earnings gap: evidence from the field and lab. *Q. J. Econ.* 138 (4), 2069–2126.
- Costas-Fernandez, J., Sebastian, F., Raute, A., Schönberg, U., 2024. Family friendly workplace policies. Technical Report.
- Croson, R., Gneezy, U., 2009. Gender differences in preferences. *J. Econ. Lit.* 47 (2), 448–474.
- Dahl, G.B., Loken, K.V., Mogstad, M., 2014. Peer effects in program participation. *Am. Econ. Rev.* 104 (7), 2049–2074.
- D'Angelis, I., 2023. The search for parental leave and the early-career gender wage gap. Technical Report 2023–01, Department of Economics, University of Massachusetts, Boston.
- Del Carpio, L., Guadalupe, M., 2021. More women in tech? Evidence from a field experiment addressing social identity. *Manag. Sci.* 68 (5), 3196–3218.
- Delfino, A., 2024. Breaking gender barriers: experimental evidence on men in pink-collar jobs. *Am. Econ. Rev.* 114 (6), 1816–1853.
- Dhar, D., Jain, T., Jayachandran, S., 2022. Reshaping adolescents' gender attitudes: evidence from a school-based experiment in India. *Am. Econ. Rev.* 112 (3), 899–927.
- Doeppke, M., Zilibotti, F., 2017. Parenting with style: altruism and paternalism in intergenerational preference transmission. *Econometrica* 85 (5), 1331–1371.
- Egan, M., Matvos, G., Seru, A., 2022. When harry fired sally: the double standard in punishing misconduct. *J. Political Econ.* 130 (5), 1184–1248.
- Farré, L., Vella, F., 2013. The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica* 80 (318), 219–247.
- Farre, L., Gonzalez, L., 2019. Does paternity leave reduce fertility? *J. Public. Econ.* 172, 52–66.
- Farre, L., Felfe, C., Gonzalez, L., Schneider, P., 2023. Changing gender norms across generations: evidence from a paternity leave reform. IZA Discussion Paper 16341.
- Felfe, C., 2012. The willingness to pay for job amenities: evidence from mothers' return to work. *ILR Rev.* 65 (2), 427–454.
- Fernández, R., 2011. Does culture matter? In: Jackson, M.O., Benhabib, J., Bisin, A. (Eds.), *Handbook of Social Economics 1A*. North-HollandBody, pp. 481–510 chapter 11.
- Fernández, R., 2013. Cultural change as learning: the evolution of female labor force participation over a century. *Am. Econ. Rev.* 103 (1), 472–500.
- Fernández, R., Fogli, A., Olivetti, C., 2004. Mother and sons: preference development and female labour force dynamics. *Q. J. Econ.* 119 (4), 1249–1299.

- Fernández, R., Fogli, A., Olivetti, C., 2009. Culture: an empirical investigation of beliefs, work, and fertility. *Am. Econ. J. Macroecon.* 1 (1), 146–177.
- Flabbi, L., 2010. Gender discrimination estimation in a search model with matching and bargaining. *Int. Econ. Rev.* 51 (3), 745–783.
- Fogli, A., Veldkamp, L., 2011. Nature or nurture? Learning and the geography of female labor force participation. *Econometrica* 79 (4), 1103–1138.
- Folke, O., Rickne, J., 2020. All the single ladies: job promotions and the durability of marriage. *Am. Econ. J. Appl. Econ.* 12 (1), 260–287.
- Folke, O., Rickne, J., 2022. Sexual harassment and gender inequality in the labor market. *Q. J. Econ.* 137 (4), 2163–2212.
- Francesconi, M., 2002. A joint dynamic model of fertility and work of married women. *J. Labor. Econ.* 20 (2), 336–380.
- Gallen, J., 2023. Motherhood and the gender productivity gap. *J. Eur. Econ. Assoc.* 22, 1055–1096.
- Gallen, Y., Joensen, J.S., Johansen, E.R., Veramendi, G.F., 2023. The labor market returns to delaying pregnancy. Available SSRN. <https://doi.org/10.2139/ssrn.4554407>
- Giommoni, T., Rubolino, E., 2022. The cost of gender identity norms: evidence from a spouse tax credit. SSRN Working Paper 4267936.
- Giuliano, P., Nunn, N., 2021. Understanding cultural persistence and change. *Rev. Econ. Stud.* 88, 1541–1581.
- Goldin, C., 1990. Understanding the gender gap: an economic history of American women. National Bureau of Economic Research.
- Goldin, C., 2006. The quiet revolution that transformed women's employment, education, and family. *Am. Econ. Rev.* 96 (2), 1–21.
- Goldin, C., 2014. A grand gender convergence: its last chapter. *Am. Econ. Rev.* 104 (4), 1091–1119.
- Goldin, C., 2014. A pollution theory of discrimination: male and female differences in occupations and earnings. *Human Capital in History: The American Record*. University of Chicago Press, pp. 313–348.
- Goldin, C., Mitchell, J., 2017. The new life cycle of women's employment: disappearing humps, sagging middles, expanding tops. *J. Econ. Perspect.* 31 (1), 161–182.
- Goldin, C., Katz, L.F., 2002. The power of the pill: oral contraceptives and women's career and marriage decisions. *J. Political Econ.* 110 (4), 730–770.
- Goldin, C., Kerr, S., Olivetti, C., 2024. The parental pay gap over the life cycle: children, jobs, and labor supply. *J. Econ. Dyn. Control* (forthcoming).
- Goldin, C., Kerr, S.P., Olivetti, C., 2020. Why firms offer paid parental leave: an exploratory study. NBER Working Paper 26617.
- Greenberg, K., Wasserman, M., Weber, A., 2024. The effects of gender integration on men: evidence from the U.S. military. Technical Report.
- Greenwood, J., Sheshadri, A., Yorukoglu, M., 2005. Engines of liberation. *Rev. Econ. Stud.* 72 (1), 109–133.
- Groes, F., Houstecka, A., Iorio, D., Santaeulàlia-Llopis, R., 2024. The unequal battle against infertility: theory and evidence from IVF success. CEPR Discussion Paper 18766.
- Grosjean, P., Khattar, R., 2019. It's raining men! Hallelujah? The long-run consequences of male-biased sex ratios. *Rev. Econ. Stud.* 86 (2), 723–754.
- Hansen, C.W., Jensen, P.S., Skovsgaard, C.V., 2015. Modern gender roles and agricultural history: the Neolithic inheritance. *J. Econ. Growth* 20, 365–404.

- Hara, H., Rodriguez-Planas, N., 2023. Long-term consequences of teaching gender roles: evidence from desegregating industrial arts and home economics in Japan. *J. Labor. Econ.* (forthcoming).
- Hazan, M., Maoz, Y.D., 2002. Women's labor force participation and the dynamics of tradition. *Econ. Lett.* 75 (2), 193–198.
- Healy, A., Malhotra, N., 2013. Childhood socialization and political attitudes: evidence from a natural experiment. *J. Politics* 75 (4), 1023–1037.
- Heathcote, J., Storesletten, K., Violante, G.L., 2010. The macroeconomic implications of rising wage inequality in the United States. *J. Political Econ.* 118 (4), 681–722.
- Heckman, J.J., MacCurdy, T.E., 1980. A life cycle model of female labour supply. *Rev. Econ. Stud.* 47 (1), 47–74.
- Hirsch, B., Schank, T., Schnabel, C., 2010. Differences in labor supply to monopsonistic firms and the gender pay gap: an empirical analysis using linked employer-employee data from Germany. *J. Labor. Econ.* 28 (2), 291–330.
- Ho, L., Jalota, S., Karandikar, A., 2024. Bringing work home: flexible work arrangements as gateway jobs for women in West Bengal. Technical Report.
- Hotz, V.J., Johansson, P., Karimi, A., 2018. Parenthood, family friendly workplaces, and the gender gaps in early work careers. NBER Working Paper 24173.
- Hotz, V.J., McElroy, S.W., Sanders, S.G., 2005. Teenage childbearing and its life cycle consequences: exploiting a natural experiment. *J. Hum. Resour.* 40 (3), 683–715.
- Hsieh, C.-T., Hurst, E., Jones, C.I., Klenow, P.J., 2019. The allocation of talent and U.S. economic growth. *Econometrica* 87 (5), 1439–1474.
- Hyde, J.S., 2005. The gender similarity hypothesis. *Am. Psychol.* 60 (6), 581–592.
- Hyde, J.S., 2014. Gender similarities and differences. *Annu. Rev. Psychol.* 65 (1), 373–398.
- Ichino, A., Olsson, M., Petrongolo, B., Thoursie, P.S., Taxes, childcare, and gender identity norms. *J. Labor Econ.*, forthcoming.
- Johnson, J.V., Hyejin Ku, Salvanes, K.G., 2024. Competition and career advancement. *Rev. Econ. Stud.* 91 (5), 2954–2980.
- Juhn, C., McCue, K., 2017. Specialization then and now: marriage, children, and the gender earnings gap across cohorts. *J. Econ. Perspect.* 31 (1), 183–204.
- Kaestner, R., Malamud, O., 2023. Headstrong girls and dependent boys: gender differences in the labor market returns to child behavior. *ILR Rev.* 76 (1), 112–134.
- Killewald, A., 2013. A reconsideration of the fatherhood premium: marriage, coresidence, biology, and fathers' wages. *Am. Sociol. Rev.* 78 (1), 96–116.
- Killingsworth, M.R., Heckman, J.J., 1986. Female labor supply: a survey. In: Ashenfelter, O.C., Layard, R. (Eds.), *Handbook of Labor Economics* 1. Elsevier, pp. 103–204 chapter 2.
- Kleven, H., Landais, C., 2017. Gender inequality and economic development: fertility, education, and norms. *Economica* 84 (334), 180–209.
- Kleven, H., Landais, C., Leite-Mariante, G., 2024. The child penalty atlas. *Rev. Econ. Stud.*
- Kleven, H., Landais, C., Søgaard, J.E., 2019. Children and gender inequality: evidence from Denmark. *Am. Econ. J. Appl. Econ.* 11 (4), 181–209.
- Kleven, H., Landais, C., Søgaard, J.E., 2021. Does biology drive child penalties? Evidence from biological and adoptive families. *Am. Econ. Rev. Insights* 3 (2), 183–198.
- Kleven, H., Landais, C., Posch, J., Steinhauer, A., Zweimüller, J., 2019. Child penalties across countries: evidence and explanations. *AEA Pap. Proc.* 109, 122–126.
- Korenman, S., Neumark, D., 1991. Does marriage really make men more productive? *J. Hum. Resour.* 26 (2), 282–307.

- Kuka, E., Shenhav, N., 2024. Long-run effects of incentivizing work after childbirth. *Am. Econ. Rev.* 114 (6), 1692–1722.
- Kuziemko, I., Pan, J., Jenny, S., Washington, E., 2018. The mommy effect: do women anticipate the employment effects on motherhood? NBER Working Paper 24740.
- Landers, R.M., Rebitzer, J.B., Taylor, L.J., 1996. *Human resources practices and the demographic transformation of professional labor markets. Broken Ladders: Managing Careers in the New Economy*. Oxford University Press, New York, pp. 215–246.
- Landsman, R. 2019. GenderDifferences in executive departure. https://economicsdept.blogs.bucknell.edu/files/2019/11/Exec_Departure.pdf.
- Liu, T., Makridis, C.A., Ouimet, P., Simintzi, E., 2022. The distribution of nonwage benefits: maternity benefits and gender diversity. *Rev. Finan. Stud.* 36 (1), 194–234.
- Lundberg, S., 2022. Gender economics and the meaning of discrimination. *AEA Pap. Proc.* 112, 588–591.
- Lundberg, S., Rose, E., 2000. Parenthood and the earnings of married men and women. *Labour Econ.* 7 (6), 689–710.
- Lundberg, S.J., Startz, R., 1983. Private discrimination and social intervention in competitive labor market. *Am. Econ. Rev.* 73 (3), 340–347.
- Lundborg, P., Plug, E., Rasmussen, A.W., 2017. Can women have children and a career? IV evidence from IVF treatments. *Am. Econ. Rev.* 107 (6), 1611–1637.
- Maestas, N., Mullen, K.J., Powel, D., von Wachter, T., Wenger, J., 2023. The value of working conditions in the United States and the implications for the structure of wages. *Am. Econ. Rev.* 113 (7), 2007–2047.
- Manning, A., 2003. Monopsony in motion: imperfect competition in labor markets. Princeton University Press.
- Manning, A., 2021. Monopsony in labor markets: a review. *ILR Rev.* 74 (1), 3–26.
- Manning, A., Swaffield, J., 2008. The gender gap in early-career wage growth. *Econ. J.* 118 (530), 983–1024.
- Mas, A., Pallais, A., 2017. Valuing alternative work arrangements. *Am. Econ. Rev.* 107 (12), 3722–3759.
- Maurin, E., Moschion, J., 2009. The social multiplier and labor market participation of mothers. *Am. Econ. J. Appl. Econ.* 1 (1), 251–272.
- Mertz, M., Ronchi, M., Salvestrini, V., 2024. Female representation and talent allocation in entrepreneurship: the role of early exposure to entrepreneurs. Available SSRN. <https://ssrn.com/abstract=4920176>.
- Miceli, A.D., 2019. Horizontal vs. vertical transmission of fertility preferences. *J. Comp. Econ.* 47 (3), 562–578.
- Miyajima, T., Yamaguchi, H., 2017. I want to but I won't: pluralistic ignorance inhibits intentions to take paternity leave in Japan. *Front. Psychol.* 8, 1508.
- Moffitt, R., 1984. Profiles of fertility, labour supply and wages of married women: a complete life-cycle model. *Rev. Econ. Stud.* 51 (2), 263–278.
- Montgomery, M., Trussell, J., 1986. Models of marital status and childbearing. In: Ashenfelter, O.C., Layard, R. (Eds.), *Handbook of Labor Economics* 1. Elsevier, pp. 205–271 chapter 3.
- Morchio, J., Moser, C., 2023. The gender pay gap: micro sources and macro consequences. Available at SSRN: <https://ssrn.com/abstract=3176868> or <http://dx.doi.org/10.2139/ssrn.3176868>.
- Mota, N., Patacchini, E., Rosenthal, S.S., 2016. Neighborhood effects, peer classification, and the decision of women to work. IZA Discussion Paper 9985.

- Ngai, L.R., Petrongolo, B., 2017. Gender gaps and the rise of the service economy. *Am. Econ. J. Macroecon.* 9 (4), 1–44.
- Ngai, R., Olivetti, C., Petrongolo, B., 2024. Gendered change: 150 years of transformation in US hours. NBER Working Paper 32475.
- Nicoletti, C., Salvanes, K.G., Tominey, E., 2018. The family peer effect on mothers' labor supply. *Am. Econ. J. Appl. Econ.* 10 (3), 206–234.
- Nollenberger, N., Rodriguez-Planas, N., Sevilla, A., 2016. The math gender gap: the role of culture. *Am. Econ. Rev.* 106 (5), 257–261.
- Noonan, M.C., Corcoran, M.E., Courant, P.N., 2005. Pay differences among the highly trained: cohort differences in the sex gap in lawyer's earnings. *Soc. Forces* 84 (2), 853–872.
- Nosek, B.A., Smyth, F.L., Sriram, N., Greenwald, A.G., 2009. National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl Acad. Sci.* 106 (26), 10593–10597.
- Olivetti, C., Petrongolo, B., 2008. Unequal pay or unequal employment? A cross-country analysis of gender gaps. *J. Labor. Econ.* 26, 621–654.
- Olivetti, C., Patacchini, E., Zenou, Y., 2020. Mothers, peers, and gender-role identity. *J. Eur. Econ. Assoc.* 18 (1), 266–301.
- Petrongolo, B., Ronchi, M., 2020. Gender gaps and the structure of local labor markets. *Labour Econ.* 64, 101819.
- Porter, C., Serra, D., 2020. Gender differences in the choice of major: the importance of female role models. *Am. Econ. J. Appl. Econ.* 12 (3), 226–254.
- Prentice, D., Paluck, E.L., 2020. Engineering social change using social norms: lessons from the study of collective action. *Curr. Opin. Psychol.* 35, 138–142.
- Ramey, G., Ramey, V.A., 2010. The rug rat race. *Brook. Pap. Econ. Act.* 41 (1), 129–199.
- Ransom, M., Sims, D., 2010. Estimating the firm's labor supply curve in a "new monopsony" framework: schoolteachers in Missouri. *J. Labor. Econ.* 28 (2), 331–355.
- Rao, N., Chatterjee, T., 2018. Sibling gender and wage differences. *Appl. Econ.* 50 (15), 1725–1745.
- Robinson, J., 1933. The Economics of Imperfect Competition, Macmillan.
- Rosen, S., 1986. The theory of equalizing differences. In O.C. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, Vol. 1, Elsevier, pp. 641–692.
- Rosenzweig, M.R., Wolpin, K.I., 1980. Life-cycle labor supply and fertility: causal inferences from household models. *J. Political Econ.* 88 (2), 328–348.
- Sani, G., Dotti, M., Treas, J., 2016. Educational gradients in parents' child-care time across countries, 1965–2012. *J. Marriage Family* 78 (4), 1083–1096.
- Sarsons, H., 2024. Interpreting signals in the labor market: evidence from medical referrals, Working Paper.
- Schaede, U., Mankki, V., 2024. Quota vs quality? Long-term gains from an unusual gender quotas. Technical Report.
- Senik, C., Lippmann, Q., Georgieff, A., 2020. Undoing gender with institutions: lessons from the german division and reunification. *Econ. J.* 130 (629), 1445–1470.
- Sharma, G., 2024. Monopsony and gender. Technical Report.
- Siminski, P., Yetsgena, R., 2022. Specialization, comparative advantage, and the sexual division of labor. *J. Labor. Econ.* 40 (4), 851–887.
- Skandalis, D., Philippe, A., 2024. Motherhood and the cost of job search. CEPR Discussion Paper 18727.
- Sokolova, A., Sorensen, T., 2021. Monopsony in labor markets: a meta-analysis. *ILR Rev.* 74 (1), 27–55.

- Tô, L., 2024. The signalling role of parental leave, Technical Report.
- Van Kleef, G.A., Gelfand, M.J., Jetten, J., 2019. The dynamic nature of social norms: New perspectives on norm development, impact, violation, and enforcement. *J. Exp. Soc. Psychol.* 84 (103814).
- Wasserman, M., 2022. Hours constraints, occupational choice, and gender: evidence from medical residents. *Rev. Econ. Stud.* 90 (3), 1535–1568.
- Webber, D., 2016. Firm-level monopsony and the gender pay gap. *Ind. Labor. Relat. Rev.* 55 (2), 323–345.
- Weisshaar, K., 2018. For opt out to blocked out: the challenges for labor market re-entry after family-related employment lapses. *Am. Sociol. Rev.* 83 (1), 34–60.
- Wiswall, M., Zafar, B., 2018. Preference for the workplace, investment in human capital, and gender. *Q. J. Econ.* 133 (1), 457–507.
- Wiswall, M., Zafar, B., 2021. Human capital investments and expectations about career and family. *J. Political Econ.* 129 (5), 1361–1424.
- Xue, M.M., 2023. High-value work and the rise of women: the cotton revolution and gender equality in China. Working Paper.

This page intentionally left blank

Chapter 9

Crime and the labor market[☆]

Randi Hjalmarsson^a, Stephen Machin^{b,*}, and Paolo Pinotti^c

^aDepartment of Economics, University of Gothenburg, Sweden, ^bDepartment of Economics and Centre for Economic Performance, London School of Economics, London, United Kingdom,

^cDepartment of Social and Political Sciences and CLEAN Research Unit on the Economic Analysis of Crime, Bocconi University, Milan, Italy

*Corresponding author. e-mail address: S.J.Machin@lse.ac.uk

Chapter Outline

1 Introduction	680	4.3 Policies to improve labor outcomes for workers with criminal records	703
2 Descriptive statistics and stylized facts: an international perspective	683	5 Education and crime	705
3 Labor market impacts on crime	687	5.1 Causal impacts of education on crime	740
3.1 Wages and income	687	5.2 Incapacitation	741
3.2 Unemployment	689	5.3 Schooling quantity and quality	742
3.3 Youth labor markets: summer youth employment programs	694	5.4 Crime impacts on education	743
3.4 Returns to crime: earnings and prices	696	5.5 Crime and education policies	744
4 Criminal record impacts on the labor market	698	6 Future directions	744
4.1 Effects of a record on labor market outcomes	700	6.1 Future direction 1: Victimization	744
4.2 Firm willingness to hire workers with criminal records	703	6.2 Future direction 2: Gangs and organized crime	748
		7 Conclusions	750
		References	751

☆ Prepared for the Handbook of Labor Economics. We thank Christian Dustmann and Thomas Lemieux for their thoughtful comments and wise editorial judgment, and conference participants from the Rockwool Foundation gathering in Berlin for helpful feedback. Randi Hjalmarsson gratefully acknowledges financial support from the Foundation for Economic Research in West Sweden, Stephen Machin is grateful to the Economic and Social Research Council for support through the LSE Centre for Economic Performance, Paolo Pinotti gratefully acknowledges financial support from the European Research Council (ERC) grant CoG 866181.

1 Introduction

High rates of crime are a major societal concern in countries worldwide, and there are many active debates on both the causes of criminal behavior and what policy reforms should be introduced to lower crime. Academic contributions to this debate have traditionally been from fields like criminology, sociology, and law. Over the last 30 years, however, a new tradition and field has emerged – *the economics of crime*. Economists are increasingly directing their attention to studying the causes and consequences of criminal behavior. The economics of crime can definitely be characterized as a ‘growth’ field in economics: as seen in Fig. 1 below, the annual number of crime papers published in top general interest and field journals, including labor journals, has sky-rocketed from just three in 1990 to 93 in 2023. Moreover, the vast majority of this explosion happened after the publication of Richard Freeman’s first Handbook chapter on crime in 1999: 58 crime articles (in total) were published from 1990 to 1999, while 837 have been published since. A new handbook chapter on the economics of crime is long overdue.

Labor economists have played a significant role in the growth of this field for two central reasons. First, there is a substantive overlap between crime and topics that are more traditionally perceived as labor economics (e.g., education, wages, unemployment, discrimination). As will be seen throughout this chapter, criminal justice populations are negatively selected in many dimensions (e.g., worse family backgrounds, less education, higher unemployment rates, lower earnings) than the general population. Moreover, the theoretical framework put forward by Gary Becker in 1968 to study criminal behavior highlights that criminal decision making is not just determined by the perceived probability and severity of punishment but also by the opportunity costs of committing crime – namely legitimate labor market opportunities. We briefly review this framework below.

Second, labor economists have brought the “credibility revolution”, i.e., a methodological toolkit to disentangle correlation from causation, to the study of crime. Fig. 1 below decomposes the annual number of crime publications into three categories: theoretical, empirical but non-causal, and empirical and causal. Before 1999, there were almost no causally identified empirical papers; but more than 70 % of the 895 total papers included in this figure in fact fit into this category. In other words, the vast majority of the growth in the economics of crime research is driven by empirical research that we classify as causal. And much of it has occurred during the last ten years, which is later than when the credibility revolution diffused through many other sub-fields of applied economics. Why? As will be seen throughout this chapter, the study of crime poses many challenges to researchers. These range from traditional identification issues, like omitted variable bias and simultaneity, to less common issues like measurement error due, for instance, to the ‘non-random’ under-reporting of crime or

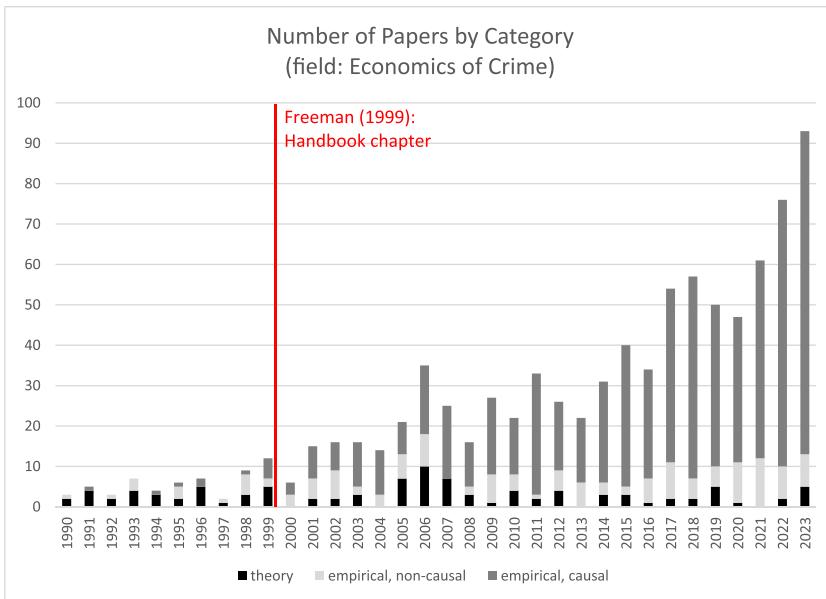


FIG. 1 The growth of the economics of crime. Note – This figure plots the annual number of publications in the Economics of Crime field, including a decomposition in the style of research: theoretical, non-causal empirical, and causal empirical. The publication data are sourced from the database maintained by Jennifer Doleac; see <https://jenniferdoleac.com/resources/> for a link to the database. The database includes crime and criminal justice system related papers in general interest and top field economics journals: AER, QJE, RESTUD, JPE, Econometrica, RESTAT, JEEA, AEJ: Applied, AEJ: Policy, AER: Insights, EJ, JOLE, JPUBE, JLE, JURBANE, JPAM, JHR, JDE, JEH, EEH, JHE. Classification of the type of research is done by the authors of this handbook chapter.

discrimination by the police or judiciary. Moreover, research ethics make one of the economics discipline's go-to tools for causal identification – randomized control trials – a rarely viable alternative in the study of crime.¹ However, increased access to individual level crime and victimization data, a necessary ingredient for many causal quasi-experimental research designs, has helped contribute to the take-off of this revolution. From the perspective of helping policy makers implement evidence-based reforms, the credibility revolution in the economics of crime has the potential to have large real-world societal impacts.

The economics of crime literature is quite broad in scope, and has been greatly influenced by the [Becker \(1968\)](#) framework (and that of [Ehrlich, 1973](#))

¹ [Stevenson \(2023\)](#) provides a recent review of randomized controlled trials conducted to study crime.

where individuals decide whether or not to engage in crime by carrying out a cost-benefit calculation under uncertainty. A stylized version of this framework can be depicted in the following equation:

$$(1 - \pi)U(W_c) - \pi S > U(W_L),$$

where π is the probability of getting caught, S is the sanction, and W_c and W_L are the economic returns to criminal and legitimate work, respectively. In this framework, an individual decides to commit a crime by rationally comparing the expected costs and benefits from criminal and legal activities. The left side of the expression displays the expected benefits of crime, which is the expected returns from illegal work offset by the expected costs or punishment. The right side of the expression includes the expected returns from legal work, which we often think of as the opportunity cost of committing crime. This framework makes clear that the criminal justice system can play a significant role in deterring crime by impacting the probability and severity of punishment. The earliest economics of crime research, and a still growing empirical literature, indeed focuses on testing these relationships. Though beyond the scope of this handbook chapter, we point the interested reader to some recent surveys of this literature ([Nagin, 2013](#); [Chalfin and McCrary, 2017](#); [Owens and Ba, 2021](#); [Apel, 2022](#)).

Rather, this handbook chapter focuses on that empirical research inspired by the other terms in the above equation: the returns from legal and illegal ‘work’. The returns from legal work are of course a function of one’s education and the labor market. Much of the empirical literature thus evaluates the implications of the economic model of crime that improved educational attainment (e.g., quantity and quality) and better labor market conditions (e.g., wages, income, employment) decrease crime. This contrasts to date a much smaller literature on the impact of the economic returns to criminal work, including, for instance, the value of the loot. [Sections 3 and 4](#) of this chapter review the literature on the impact of labor markets on crime and crime on labor market outcomes, respectively. [Section 5](#) reviews the literature linking education and crime.

This chapter highlights not only that there is more economics of crime research, but that it has changed in a multitude of ways since the [Freeman \(1999\)](#) handbook review of crime and the labor market. There are a number of recurring themes throughout the chapter, each of which has very much altered the take-aways from what the research says, at least relative to when [Freeman \(1999\)](#) was drawing his general conclusions. First, there has been a significant shift from using aggregated data (e.g., at the US state level) to highly disaggregated data – with an increasing prominence for individual level micro data and individual decisions. Second, the literature is increasingly non-US centric, in part because of the availability of micro-level crime data in a number of international contexts. Third, there has been an increasing shift from only focusing on the role for criminal justice policy (e.g., police and prisons) to affect crime to also the role for social policy (e.g., schools and labor market barriers). Fourth, the literature is

increasingly considering that social interactions and spillovers may be important in explaining crime behavior and measuring the costs and benefits of policies impacting crime. Fifth, the new economics of crime literature pays careful attention to the causal identification challenges and quasi-experimental solutions. Sixth, new doors are being opened in the field to study questions beyond the Becker framework, like the costs of crime victimization (including labor market costs) and the role of criminal organizations; we offer a discussion of these especially new literatures in the conclusion.

The remainder of this chapter proceeds as follows. Given the international scope of the research surveyed, [Section 2](#) begins by discussing some of the similarities and differences in, for instance, criminal justice systems around the world. We then survey the labor market-crime literature in [Section 3](#), with sections on the roles of wages and income, unemployment, summer youth employment programs, and the returns to illegitimate labor. [Section 4](#) presents the literature concerned with how criminal records impact labor market outcomes, the willingness of firms to hire workers with records, and the effect of policies aimed at improving labor market outcomes. [Section 5](#) reviews the education – crime literature, including that on causal impacts, on incapacitation, the quantity and quality of schooling, productivity and on education policy. [Section 6](#) concludes by highlighting some of the new directions in which the labor and crime research is going, highlighting the economics of victimization and organized crime.

2 Descriptive statistics and stylized facts: an international perspective

This chapter surveys research on crime and the labor market in a wide range of international contexts, including, for instance, the United States, Brazil, United Kingdom, France, Italy, Sweden, Norway, Finland, Denmark, and the Netherlands. The international breadth of this literature speaks to (i) societal concerns about crime being a worldwide phenomenon, even if the nature of crime varies across countries, and (ii) the implications of [Becker's \(1968\)](#) economic model of crime, e.g., that increased legitimate labor market opportunities should decrease crime, not being country specific.

One fundamental challenge in studying the causes and consequences of criminal behavior is accurately observing and measuring crime. There are many potential sources of measurement error in official crime statistics. For instance, aggregated arrest data do not only reflect crime incidents but also reporting rates and police clearance rates. Moreover, these sources of measurement error can vary systematically across crime categories and international contexts. This issue is highlighted in [Fig. 2](#). Homicide rates, which are the most accurately recorded both within and across countries, are negatively related to GDP per capita. Burglary rates, which are more sensitive to under-

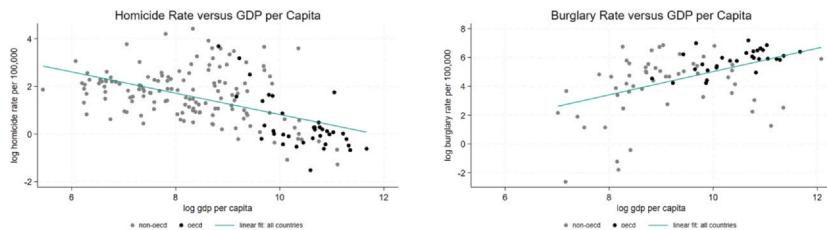


FIG. 2 GDP per capita and crime rates across countries. Note – These statistics are based on the following sources. Burglary rates per 100,000 across countries in 2018 were obtained from <https://dataunodc.un.org/dp-crime-corruption-offences>. GDP per capita in 2018 and the homicide rate per 100,000 were obtained from the World Bank and World Health Organizations, respectively.

reporting, legal system differences, and police effort, are positively related to GDP. Similar patterns are seen for other crime categories.

See: <https://databank.worldbank.org/reports.aspx?source=2&series=NY.GDP.PCAP.CD&country=#> and <https://www.who.int/data/gho/data/indicators/indicator-details/GHO/estimates-of-rates-of-homicides-per-100-000-population>.

A significant advancement of the new crime and labor market literature has been the shift away from many of the problems associated with aggregate statistics to micro-level data. This is in fact one practical reason for the non-US focus of this research, and the over-representation of research in a Scandinavian context; in selected countries, population-wide crime and victimization registers can be matched to other national education, labor market, and health registers.

At the same time, studying the link between crime and the labor market in an international context is motivated by the fact that the US is an outlier in many dimensions, three of which are depicted in Fig. 3, related to crime and the criminal justice system. The 2018 US homicide rate per 100,000 persons is 5.78, which is more than three times larger than neighboring Canada and nearly ten times that of countries like Spain and Norway. Though cross-country comparisons of crime rates are typically complicated by differential crime definitions and reporting rates, such concerns are minimal for the case of homicides. Second, incarceration rates (depicted by the black squares in Fig. 3) are also markedly higher in the US than Canada or Western Europe. In fact, the 2021 US incarceration rate of 531 inmates per 100,000 persons is even higher than that of Brazil, which had a rate of 390 in 2022. Though these international differences in incarceration rates in part reflect differential crime rates, this is not the only explanation; after all, the Brazilian homicide rate was more than five times larger than that in the US. Rather, there are also international differences in punishment severity. At the extreme, this is seen in Fig. 3 with respect to the death penalty. All depicted states but the US have abolished capital punishment for all crimes in both peacetime and wartime (the complete abolition date is listed on the x-axis).

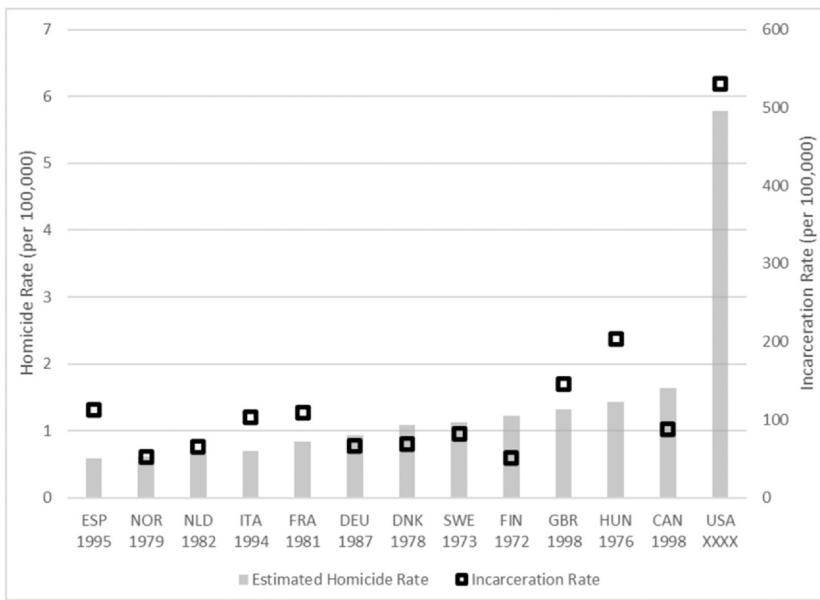


FIG. 3 International homicide and incarceration rates. Note – For each country, we present the year the death penalty was completely abolished, including for war related crimes under the country name. This information was sourced for most countries from <https://deathpenaltyinfo.org/policy-issues/international/countries-that-have-abolished-the-death-penalty-since-1976>. For other countries (Sweden, Finland, Spain, and Canada), we conducted web searches. The death penalty was abolished for peacetime crimes even earlier in many countries. Homicide rate statistics (gray bars) are estimates provided by the World Health Organization for 2018. <https://www.who.int/data/gho/data/indicators/indicator-details/GHO/estimates-of-rates-of-homicides-per-100-000-population>. Incarceration rate statistics (black squares) were sourced from the most recent available year (2021–2024) from <https://www.prisonstudies.org/world-prison-brief-data>.

But there are also many less extreme dimensions in which ‘incarceration’ differs across countries, including both the rate at which offenders are sentenced to prison (especially for more minor drug and property offenses) as well as the sentence length. For instance, [Hjalmarsson and Lindquist \(2022\)](#) note that the average time spent in prison for inmates convicted in Sweden from 1991 to 2001 was 4.7 months, which compared to about 8 months in Western Europe on average in 2001 and more than 30 months in US state and federal prisons. Moreover, the prison conditions or experience vary substantially across countries, and even across regions or states within countries. Prison conditions are known to be especially good in Scandinavian countries. In 2015, Sweden and Norway spent more per prisoner than any other country (nearly 150,000 US dollars) and about four times that of the US ([Hjalmarsson and Lindquist, 2022](#)). This translates into significant cross-country differences in

the inmate to staff ratio, in-prison treatment programs and rehabilitation conditions, healthcare, overcrowding, and job training and education opportunities.

The impact of the labor market on crime or that of a criminal record on employment outcomes may depend on more than the nature of the criminal justice system and sanctions. In particular, international differences in the labor market and social welfare systems may play just as significant a role. Public spending on social expenditures was about 20 % of GDP in the OECD in 2015. But this masks significant heterogeneity across member countries, ranging from 30 % or more of GDP in France, Finland, and Denmark to less than 20 % of GDP in the US, Canada, and Netherlands.²

Other international policies and norms may also be especially salient to the crime and labor market question. These include: (i) whether having a criminal record actually disqualifies one from welfare eligibility or particular jobs, (ii) the availability of criminal records online, (iii) the extent to which criminal background checks are mandated and/or used during hiring, and (iv) the culture of whether offenders should be given a second chance.

This section has thus far highlighted the many criminal justice and societal differences across countries faced by potential or former offenders. These differences are of two-fold importance. First, they motivate the need for the replication of research across different countries and contexts. Second, understanding these institutional differences may play a fundamental role in explaining heterogeneous results.

In light of the discussion, it is natural to ask whether one can at all generalize or apply the results of one study or context to another. We believe the answer to this question is yes, given that, despite these many differences, there are also common stylized facts that characterize criminal justice populations worldwide. For instance, in terms of demographics, offenders are disproportionately young and male. Moreover, the age-crime profile, in which crime tends to increase until young adulthood and then decrease, is a pattern seen across countries and throughout history (Bindler and Hjalmarsson, 2017). Criminals also tend to be negatively selected in many dimensions. With regards to education, for instance, 41 % of US prisoners in 1997 had not completed high school (compared to 18 % of the general population) while more than 75 % of Italian convicts in 2001 had not completed high school (Harlow, 2003; Buonanno and Leonida, 2006). Similar findings are documented in Sweden (Hjalmarsson et al., 2015) and the UK (Machin, Marie and Vujic, 2011). Finally, offenders often have mental health and/or substance abuse problems themselves and disproportionately come from disadvantaged family backgrounds, including single parent households, having criminal fathers, or parents with alcohol or mental health problems (Hjalmarsson, 2022).

² See https://stats.oecd.org/Index.aspx?datasetcode=SOCX_AGG.

3 Labor market impacts on crime

The prime research focus over the years has been to study whether, and how, the labor market impacts crime. Over time, the research has been framed in different, but mutually compatible ways where empirical connections between crime and different labor market measures have been studied. These have evolved considerably since the previous handbook chapter and on this aspect, reflecting the very recent work in this area of the past ten years or so, and since some other more recent reviews of work on crime and the labor market (e.g., Draca and Machin, 2015).

That said, through time, the crime-unemployment relation is the one that has received most attention. Reaching strong conclusions in the early work was hampered both by data limitations and by biases from the use of aggregated cross-sectional data and reverse causation. At the time of Freeman's (1999) economics of crime Handbook chapter, it was becoming clear that the mostly aggregate studies he reviewed did not find much evidence of unemployment being important. Indeed, in his review, Freeman (1999) summarized the evidence then as follows: "unemployment is related to crime, but if your prior was that the relation was overwhelming, you were wrong. Joblessness is not the overwhelming determinant of crime that many analysts and the public a priori expected it to be".

Around that time, some new studies emerged that placed a directed focus on wages, earnings or income rather than unemployment, and it turns out they had more empirical success. In this section, we therefore begin with a discussion of these studies, and then turn to the newer literature that revisits a lack of work as a determinant of criminality. Interestingly, and for a variety of reasons to be discussed, this work reaches a quite different view in comparison to the take-away from the Freeman (1999) chapter. The new, more voluminous literature on crime and the labor market – both the new studies of unemployment and that on wage effects – is much more supportive of the economics of crime model predictions that labor market outcomes matter for crime. This is especially so for the impact of joblessness, but also for wages and income; moreover, the methodological advances in some studies show such effects to be causal.

3.1 Wages and income

In the basic economics of crime framework, legitimate income opportunities represent the "opportunity cost" of crime and, as such, are one of the main deterrents to criminal behavior. From a theoretical perspective, differences in actual and potential earnings may explain, for instance, why the rich typically commit less crime than the poor, or why more educated individuals commit fewer crimes than the less educated. At the same time, individuals with different income opportunities may differ along several other dimensions, such as family background, the areas where they live, the peers they are

exposed to, and so on. Moreover, involvement in crime may itself affect income opportunities in legitimate markets, as previous offenders typically face implicit or explicit barriers to accessing legitimate economic activities, as discussed at length in [Section 4](#) below. For all these reasons, empirically identifying the effect of income opportunities on crime requires plausibly exogenous variation in legitimate income, particularly labor earnings which is the main source of income for poorer individuals at the margin between committing or not committing crimes.

As discussed already, the first round of crime and labor market research primarily focused on the issue of whether crime rates, in particular property crime rates, were related to unemployment rates in a variety of settings. It is on the basis of these studies that [Freeman \(1999\)](#) reported to find any relationship to be “fragile, at best”. Around the same time, or just after, the publication of the [Freeman \(1999\)](#) review, a first set of studies looking at earnings effects on crime emerged. Typically, though not always, these studies also looked at unemployment effects and concurred with the earlier work that these effects were hard to detect. But they did find that wages, earnings or income mattered for crime.

[Groger \(1998\)](#) is the first of these studies. Using the National Longitudinal Survey of Youth (NLSY) cohort data, he shows that many people who self-report some criminal activity are also working in the labor market. This both makes them sensitive to wage changes along an extensive margin between legal and illegal work, and also gives a rationale for weak effects from lack of work. Indeed, in estimates of a probit model of crime incidence there is a strong negative association of crime with wages and less of an effect from unemployment.

[Gould et al. \(2002\)](#) provide evidence based on a US panel of counties, using the wages for non-college educated males as their earnings measure. They include wage and unemployment measures contemporaneously, which allows for some benchmarking of effects. For example, over the 1979–1993 period the recorded 23.3 % fall in unskilled wages predicted 43 % of the total increase in property crime while the 3.05 % point increase in unemployed predicted 24 % of the change. Wages also dominated the results for violent crime (predicting 53 % of the increase versus 8 % for unemployment). They address potential problems related to the endogeneity of crime and economic conditions using an instrumental variables strategy that interacts fixed state-level characteristics with aggregate economic shocks (following the logic of [Bartik, 1991](#)). They find that the instrumented estimates are larger than those estimated by least squares for the wage measure, but are lower for unemployment.

[Machin and Meghir \(2004\)](#) analyze a 20-year panel of police force areas of England and Wales. They use a wage measure based on the 25th percentile of the distribution and empirically find that the marginal effect of a 10 % increase in the wage measure corresponds to 0.7 % point fall in the crime rate. Similarly, [Entorf and Spengler's \(2000\)](#) analysis of data on German regions over

time uncovers significant associations between crime and income, again in line with the notion that changing economic incentives in the labor market matter for crime.

These panel studies all find that wages matter for crime, and more so than unemployment. Another way to consider the relationship between crime and wages is the longitudinal analysis of arrests and wages of [Grogger \(1995\)](#). In a study that is arguably ahead of its time in utilizing administrative data (in this case, for California on criminal histories and labor market earnings), his empirical strategy includes fixed effects in a longitudinal earnings model to enable tracking out the wage effects of arrest over a number of quarters. Whilst the effects are moderate – equal to around 4 % of earnings in the quarter contemporaneous with arrest and falling to an average of around 2–3 % over the next 5 quarters before fading out to a zero statistical effect – they are also suggestive of a wage-crime empirical connection.

A small body of work has taken a rather different approach, by looking at what happens to crime when minimum wage floors are increased. For police force areas of England and Wales, [Hansen and Machin \(2002\)](#) report that crime fell in relative terms in places where the UK introduction of a national minimum wage in 1999 boosted low wages by more. In the US, there is evidence that reductions in recidivism result from minimum wage increases in [Agan and Makowsky \(2023\)](#); [Corman and Mocan \(2005\)](#) also report a negative time series relation in New York City between crime and the real minimum wage. However, other US work does not find evidence of crime reduction ([Beauchamp and Chan, 2014](#); [Fernandez et al., 2014](#); [Fone et al., 2023](#)). Thus, it seems reasonable to conclude that the evidence from this small number of studies proves mixed.

Finally on wage and income effects, a study that attempts to move to causal effects is the interesting economic history paper of [Bignon et al. \(2018\)](#). They track the progressive influx of the phylloxera virus that diffused across areas in France in the 19th century. In doing so it hugely damaged wine production, and both reduced the incomes and livelihoods of wine producers and impacted crime. In an instrumental variable setting, they show the phylloxera crisis caused a substantial increase in property crime rates and a significant decrease in violent crimes, thereby generating causal effects of income on crime. Their interpretation of the findings offers interesting insight, as they conclude the property crime increase arises from what the economic model of crime suggests as income opportunities degenerated, but that the violent crime increase in this setting comes about from reduced alcohol consumption as wine was both less available and more expensive.

3.2 Unemployment

Returning to the question of how unemployment and lack of work affect crime, it is evident that job loss is possibly the biggest (negative) shock to labor earnings. Workers losing their jobs experience an immediate drop in legitimate income and,

in case they struggle to find a new job, they may experience even greater income losses in the medium to long run. Though the severity and timing of these income losses depend on the specific institutional context, particularly the generosity and duration of unemployment insurance, job loss always brings significant risks of prolonged periods of unemployment with long-lasting consequences for human capital accumulation, health, and earning potential (see, e.g., [Pissarides, 1992](#), and [Von Wachter, 2020](#)). These insights underpin the direction to which some of the more recent work on crime and unemployment has moved.

Any economic model of crime would predict that job loss increases the probability of committing crimes. However, as already noted, the earlier empirical studies found only mixed evidence of an effect of job loss on crime. For instance, [Freeman and Rodgers \(1999\)](#) estimate that, across US states, (youth) crime increases by as little as 1.5 % for a one-point increase in unemployment, and [Cullen and Levitt \(1999\)](#) even estimate a null effect across US cities. One big limitation of these earlier studies is that they rely mostly (or exclusively) on correlational evidence. In most cases, they regress crime rates on unemployment rates across geographical areas – states, counties, or cities – without a clear strategy for identifying causal effects. More recent papers exploiting plausibly exogenous variation in economic shocks across areas, as predicted by shift-share instruments interacting national-level sectoral shocks with local sectoral shares, tend to find stronger evidence of a causal effect of unemployment on crime.

[Raphael and Winter-Ebmer \(2001\)](#) estimate the effect of unemployment on crime across US states over the period 1971–1996 using state exposure to oil shocks and military spending directed to each state as an instrument for state-level unemployment rates. The results reveal a statistically significant, positive effect of unemployment on property crime, while the relationship with violent crime is much weaker. Findings that unemployment matters for crime have also been obtained for European countries characterized by quite different labor market institutions than the US – notably, more generous welfare support to displaced workers. [Öster and Agell \(2007\)](#) and [Fougère et al. \(2009\)](#) use a shift-share approach to identify the effect of unemployment on crime across, respectively, Swedish municipalities and French departments during the 1990s. Both papers find that unemployment increases property crime. [Bell et al. \(2018\)](#) estimate the same relationship in the UK and the US using a different, previous unexplored empirical strategy for crime, namely comparing cohorts entering the labor market in different periods. They show that individuals entering the labor market during recessions in the UK and the US are, respectively, 4 % and 10 % more likely to be arrested.

These studies begin to make it clear that the question of whether a lack of work matters for crime requires more consideration of the way in which unemployment can affect crime than the available data and methodologies could address in the early, first round work on crime and unemployment. Two important insights emerge. First, studying unemployment rates of the appropriate population who may commit crime – mostly young men – does find

evidence of an empirical relation between youth crime and youth unemployment. Second, formulating the question in terms of entry unemployment rates (i.e., when individuals enter the labor market) produces strong evidence of a crime-unemployment relation. People who enter the labor market in bad times (local and national recessions) are strongly and persistently more likely to engage in crime. [Bell et al. \(2018\)](#) argue therefore that criminal careers can indeed be ‘made’ according to initial labor market conditions and so recessions can act as a turning point for the onset of criminal careers.

Outside Europe and the US, [Dix-Carneiro et al. \(2018\)](#) exploit the asymmetric impact of the 1990s trade liberalization across Brazilian regions to identify the impact of unemployment on homicides which, they argue, is the only crime that can be consistently measured across regions during their sample period (1980–2010). Regions that experienced a negative trade shock (in the form of higher foreign competition) saw a temporary increase in local homicide rates compared to the national average. Therefore, in contexts with high levels of violence (as is certainly the case for Brazil), violent crimes also seem to respond to unemployment. This is also the case in Mexico where, according to [Dell et al. \(2019\)](#), municipalities experiencing higher increases in unemployment induced by greater competition from China also witness an increase in drug-related homicide rates.

Overall, the evidence emerging from more recent papers leveraging plausibly exogenous economic shocks across local areas is consistent with the theoretical prediction that unemployment increases crime. However, it may be hard to defend the exclusion restriction that economic shocks impact crime only through unemployment rates. For instance, economic crises that hit severely not only the private sector but, also, local public finances, may reduce expenditures on police and other public services, in which case the shift-share approach (or other approaches leveraging exogenous variation from economic shocks) would over-estimate the impact of unemployment on crime. Conversely, local authorities anticipating negative effects on crime may decide to allocate larger budgets to law enforcement, or they could invest in other policy tools aimed at counteracting the expected crime increase, such as assistance to poor families; in this case, the instrumental variable approach would underestimate the impact of unemployment on crime. In sum, estimates obtained using shift-share instruments and aggregate data at the local level may conflate the effect of unemployment with that of other local factors. Therefore, the plausibility of the exclusion restriction would depend on the specific context.

A second, related problem with local-level data is that they may prevent us from assessing the effectiveness of some policy remedies such as unemployment benefits or conditional cash transfers, as the latter are typically uniformly defined at the national level.

Finally, in all countries (including the most violent ones), crimes remain relatively rare events that are committed by a low number of serial offenders,

so averaging data on crime and unemployment within local areas may hide the relationship between the two variables. Put differently, aggregate data may be ill-powered to detect the effect of unemployment on crime, even in cases in which the effect size is large. This problem is particularly severe for violent crimes, particularly murder and other most serious offenses, because they are much rarer than property crimes.

For all the reasons just discussed, there are some inherent limitations to what we can learn from aggregate data. Individual-level data may overcome some of these limitations, and they have been used to study the impact on unemployment on crime since the 1980s. [Witte \(1980\)](#) and Schmidt and Witte used individual-level data on former prison inmates in North Carolina to study several determinants of recidivism, including employment. [Witte and Tauchen \(2000\)](#) focus on a 10 % random sample of the 1945 birth cohort in Philadelphia surveyed over multiple periods of time, which allows them to use panel data methods. These papers find stronger support for the relationship between unemployment and crime compared to contemporaneous papers using only aggregate data. At the same time, the empirical analysis remains correlational, so the estimated coefficient of unemployment cannot be given a causal interpretation. Moreover, in all these studies, the sample remains quite small and, in the case of [Witte \(1980\)](#) and [Schmidt and Witte \(2013\)](#), it includes only former offenders and, thus, is not representative of the national population.

Several recent papers address these limitations by leveraging large registry data and plausibly exogenous variation in job loss. In particular, [Bennett and Ouazad \(2020\)](#) focus on Danish (high-tenure) workers who lost their job in mass layoff events during the period 1990–1994. They show that the probability of committing crime increases by 0.57 % points in the year of displacement (+32 % over the baseline probability of committing crime), and that such an effect is primarily driven by property crimes. The dynamics of the effect is non-monotonic, spiking at 0–1, 3–4, and 6 years after job loss. The authors attribute this pattern to transitions out of unemployment benefits, determined by the design of unemployment insurance schemes. Using a similar research design, [Rege et al. \(2019\)](#) detect a 20 % increase in crime among Norwegian workers displaced in mass layoffs during the period 1992–2008, but the effect is transitory and vanishes before the fourth year after job loss.

Both of these previous papers look at developed countries with relatively low crime rates and generous welfare systems, but the implications of job loss for crime are likely more severe in developing countries in which crime and poverty are much more widespread, and safety nets are less generous. Focusing on Colombian workers in the city of Medellin during the period 2006–2015, [Khanna et al. \(2021\)](#) find that job loss increases the probability of committing a crime by 46 % in the first year after layoff, and by 35 % the second year after layoff. The effect is concentrated on property crimes, and is attenuated by greater access to credit, driven by a financial reform that expanded the number of bank branches. Finally, [Britto et al. \(2022\)](#) estimate the impact of job loss on

crime using employer-employee data on the universe of workers in Brazil matched with judicial records over the period 2009–2017. The probability of committing a crime increases by over 20 %, an effect that persists for at least four years after job displacement. The large sample size allows the authors to precisely identify significant increases for almost all types of offenses, including violent and non-economically motivated crimes; property crimes and other economically motivated crimes, however, increase the most, in line with prior evidence. [Britto et al. \(2022\)](#) then estimate the mitigating role of unemployment insurance by comparing displaced workers that obtained and did not obtain three-to-five months of unemployment benefits after the layoff. Since in Brazil eligibility for benefits varies discontinuously with the timing of previous layoffs, they compare displaced workers that are barely eligible and non-eligible for benefits. Receiving unemployment benefits causes a decrease in crime that completely offsets the increase due to job loss, but the effect is transitory and disappears immediately as benefits expire.

The results in [Britto et al. \(2022\)](#) refer to all types of crime except domestic violence because identifying the impact of job loss on the latter is particularly challenging. Differently from other types of crime, the dynamics of domestic violence and its reporting may depend on the bargaining power of both the offender and the victim. Specifically, a higher incidence of domestic violence after the (male) offender loses his job may reflect either an actual increase in violence, or a higher willingness of the (female) victim to report it, or both. To address this issue, [Bhalotra et al. \(forthcoming\)](#) combine the data used in [Britto et al. \(2022\)](#) with additional data on mandatory reports by health providers on suspect cases of domestic violence, which attenuate concerns about biased reporting. According to their results, both male and female job loss increase the risk of domestic violence – by 32 % and 56 %, respectively. These findings are in contrast with household bargaining models ([Aizer, 2010](#)), which predict that the risk of domestic violence increases with female job loss and decreases with male job loss. They are also in contrast with male backlash models ([Macmillan and Gartner, 1999](#)), which predict exactly the opposite, namely that domestic violence increases with male job loss and decreases with female job loss. Therefore, both these models predict opposite effects of male and female job loss, while the results of [Bhalotra et al. \(forthcoming\)](#) suggest that both male and female job loss increase domestic violence. They show that this result can be reconciled with a simple model in which domestic violence depends negatively on income and positively on time spent together in the couple, so both male and female job loss increase domestic violence through both mechanisms. Interestingly, unemployment benefits attenuate the income effect but increase time spent together, through the reduction in labor supply; the former effect prevails while benefits are paid, whereas the second effect prevails after they expire, so the overall effect of unemployment benefits is more ambiguous than for other types of crime.

Overall, the available evidence is largely consistent with the Becker-Ehrlich model, which predicts that job loss and unemployment increases crime by reducing its opportunity cost. At the same time, more recent evidence based on large registry data points at significant effects on non-technically motivated crimes such as substance abuse and violent crimes, including domestic violence. These findings suggest that job loss and unemployment may affect criminal behavior through other (non-economic) mechanisms such as psychological stress. Reviewing the position of whether lack of work matters for crime now, as compared to 25 years ago in [Freeman's \(1999\)](#) chapter, paints a much clearer picture. More precise framing of the empirical hypotheses of interest, coupled with significant methodological and data advances, lead to the conclusions that being out of work is an important determinant of criminality, as are shifts in wages and income for individuals on the margins of committing crime.

3.3 Youth labor markets: summer youth employment programs

Crime is over-represented amongst youths and young adults. In fact, similar age-crime profiles – in which crime increases throughout the teenage years, peaks around age 19 or 20, and then gradually decreases – are seen throughout the world and history ([Bindler and Hjalmarsson, 2017](#)). This sub-section emphasizes the impact of youth labor markets on crime, and focuses in particular on one dimension that is specific to youths: summer youth employment programs (SYEP).

[Heller and Kessler \(forthcoming\)](#) highlight the prevalence of these programs in the United States. From 2014–2016, 27 of the 30 largest U.S. cities had a SYEP. These programs vary in many dimensions, including: (i) size, from just 70 participants in San Francisco to 54,000 in New York City per year, (ii) allocation mechanism (random, first-come first-serve, merit based, and criteria based), (iii) content (summer employment, supplementary job preparation training, and/or additional support like cognitive behavioral therapy), and (iv) target group (e.g., student versus non-students). One common aspect of almost every program is over-subscription: there are more applicants (often many times more) than positions. Such SYEP are also not specific to the United States – Swedish programs exist, for instance, and evaluations are underway.

The ultimate goals of SYEP are to: improve the short-term economic conditions of participants, provide work experience that can improve future education and employment outcomes, and keep participants out of trouble ([Gelber, Isen, and Kessler, 2016](#)). Crime can be impacted by SYEP participation through multiple channels, including: a contemporaneous incapacitation effect (in which youths are kept busy during the idle summer months), an improvement in legitimate labor market opportunities, an improvement in other behaviors (e.g., responsibility or societal attitudes), and a lower need for financially motivated crime due to higher earnings.

Given that youths who volunteer for SYEP are likely different than non-participants in observable and unobservable dimensions (e.g., risk preferences, motivation and family background), selection bias will act to confound estimates of the effects of SYEP on crime in the absence of an experimental design (LaLonde, 1986). In contrast to most of the literature surveyed in this chapter, recent studies of SYEP take advantage of a random assignment mechanism in select cities to overcome these challenges.

Heller (2014) conducted the first such evaluation in Chicago, where 1634 disadvantaged high school youth were randomly assigned in 2012 to three groups: (i) control, (ii) jobs-only, with 25 h per week of paid employment, and (iii) jobs plus therapy, with 15 h of work and 10 h of cognitive behavioral therapy per week. For both treatment arms, a 43 % reduction in violent crime, but no effect on property crime, is seen over a 16-month period. What mechanisms underlie these effects? Given that the same effects are seen for groups with and without therapy, it appears to be something about the job and not just the therapy itself. Moreover, most of the effects accrued after the program was completed, suggesting they are not just a mechanical result of incapacitation (though this does not rule out a dynamic incapacitation story).

Davis and Heller (2020) further found that these effects replicated when expanding the Chicago program eligibility to include disadvantaged youths who were no longer enrolled in school. Moreover, these results are not specific to the Chicago program. Modestino (2019) evaluates the crime impacts of the 2015 Boston program, which reaches about 10,000 youth per summer and connects them to 900 local employers. Youths work for 25 h per week for six weeks at minimum wage and receive 20 h of job-readiness training, such as preparing for interviews, job search and job applications. Large effects on violent crime (35 % reduction) were seen for Boston participants, while, in this program, property crime also decreased by 29 %. Most of the effects were once again post-program. Modestino also finds in survey data short-term changes in attitudes and behaviors that could be related to crime, including attitudes towards the community and social skills and behavior (e.g., managing emotions and asking for help).

Gelber, et al. (2016) and Kessler et al. (2022) evaluate the impact of participating in the New York City program for youths aged 14 to 21. From 2005 to 2008, 294,100 applications were received, and 164,641 won a job through the computerized lottery. Gelber et al. (2016) find improvements in employment and earnings during the year of program participation, but by three years post program, there is a modest decrease in average earnings and no effect on college enrollment. Finally, the authors do find a reduction in incarceration in New York State prison (a proxy for serious crimes or criminal history) and the chance of death. Kessler et al. (2022) conduct a follow-up evaluation that digs deeper into the crime effects for SYEP participants aged 16 and older, for whom criminal records on adult arrests and convictions in New York state could be matched on. The authors find that the New York sample is less

negatively selected than the Chicago sample: Just three percent had an arrest prior to the program. And it is just for this sub-sample that the authors find an effect: the chance of felony arrest is reduced by more than 20 % during the program and remains large (though not statistically significant) five years later.

Taken together, this new body of work provides strong and convincing evidence that summer jobs programs for youth—and especially those for high-risk youth—reduce violent crime. Though the channels through which these effects occur are still unclear, the research suggests it is more than a pure incapacitation effect and not driven by the therapy sometimes attached to the program. But will these programs continue to have such effects as they expand, given that expansion can change the marginal participating youths and/or employers? [Heller \(2022\)](#) provides evidence that SYEP may indeed successfully scale up by conducting evaluations in (i) Chicago with three times as many participants ($n = 5405$) as her earlier evaluations and (ii) Philadelphia with more than 50 program providers. Despite variation in providers or expanded participant samples, both evaluations led to large reductions in criminal justice interactions in the first year of participation, with some preliminary evidence of a persistent decline in arrests in subsequent years.

3.4 Returns to crime: earnings and prices

The basic economics of crime framework states that shifts in relative earnings from legal work compared to illegal work shape crime. However, there is far more empirical research on legal earnings as compared to illegal earnings from crime, the latter which can be thought of as the economic returns to crime. This much was true when [Freeman \(1999\)](#) wrote his Handbook chapter and, while discussing the paucity of research, he was able to point to studies on the attempted measurement of the earnings of criminals, conclusions from which are difficult to draw due to their small sample sizes, highly specific nature, and/or measurement difficulties. Subject to these caveats, these earlier survey-based findings suggest average illegal earnings to be close to the average legal earnings faced by criminals. But this is with variations across different crime types (e.g., drug dealing may be more lucrative), and also with only a subset benefitting from sizable crime returns, typically prolific offenders in a skewed distribution of returns.

Though things have moved on since then, there is still not much research in this area, which still to this day likely reflects the difficulty of obtaining good data (or any data) on criminal earnings. One notable exception is the field study of a drug selling gang by [Levitt and Venkatesh \(2000\)](#), which links the issue of illegal earnings to the economics of criminal enterprises, in this case a drug gang whose financial operations were documented over a four-year period. The focus provides some important context for understanding criminal earnings, namely the hierarchical structure of criminal work. Drug-selling is input intensive – the wage bill to revenue share is approximately one-third. Wages

for street-level dealers are low – comparable to the minimum wage – and carry serious risks (the death rate for the sample was 7 % annually). The incentive for gang participation therefore lies in the prospect of moving up the hierarchy within the gang, in line with a tournament model. Rewards at the top are high – with wages between 10 and 25 times higher than ‘foot soldier’ wages. The results do sit quite well with the economics of crime model, but at the same time, the message that emerges is one that crime does not pay much for most participants, but that a few criminals benefit significantly from a highly skewed structure of illegal rewards.

The second recent development on the returns to crime comes from studies that [Draca and Machin \(2015\)](#) refer to as the ‘internal rates of return’ to criminal opportunity. This means the cash flow or return generated by a criminal project – the “loot” from crime – holding the probability of detection or other costs fixed. This concept is most relevant for the case of property theft and there are by now a few empirical studies looking at the relationship between property theft rates and prices (i.e., measuring the value of loot).

[Reilly and Witt \(2008\)](#) examine the relationship between domestic burglaries and the real price of audio-visual goods (a major component of the ‘loot’ obtained in burglaries). They consider an annual time series of UK burglary and price data over the period from 1976–2005, when the retail price of audio-visual goods fell by an average 10 % per annum. Their main specification is an error-correction model (ECM) that includes controls for unemployment and inequality (a Gini-based measure) together with their main price variable. The long-run estimates from this ECM indicate an elasticity of 0.286, such that a 10 % fall in prices is associated with a long-run fall in the volume of domestic burglary of 2.9 %.

The paper by [Draca et al. \(2019\)](#) pushes much further and looks at the relationship between goods prices and crime across a wide range of goods. They use records from the London Metropolitan Police’s (LMP) crime reporting system, which features a property type code that classifies stolen goods as part of theft, burglary and robbery incidents. These property types are then matched by label description to ONS data on retail prices. Results based on a panel of 44 matched goods – covering goods ranging from clothing, drink and foodstuffs, electronic equipment, household goods, and jewelry – indicate an average elasticity with respect to prices of 0.3–0.4. Furthermore, there is a short lag between price changes and crime, with the majority of adjustment occurring within three months of a given price change, limiting the scope for time-varying unobservables to explain the price effect.

Endogeneity concerns are addressed by focusing on a subset of goods – three metals (copper, lead and aluminum), as well as jewelry and fuel – where domestic prices can be plausibly linked to international prices. In the case of metals, they instrument local scrap metal dealer prices with global commodity prices, while fuel is instrumented with oil prices and jewelry with the

price of gold. This approach has the advantage of isolating price changes that are a function of international demand (for example, commodity demand from China) rather than variations due to local demand, which could in turn change the local stock of goods available for theft. The results for this sub-set of goods show higher elasticities that mostly exceed unity, indicating that criminals are highly elastic with respect to prices and the implied value of criminal opportunities. The case of metal crime is also explored in [Kirchmaier et al. \(2020\)](#). Their emphasis is different, as they move on from a focus on returns for criminals by looking at price booms and busts, policing initiatives and regulation of second-hand buyers, specifically scrap metal dealers. But, in doing so, they too report a price elasticity greater than unity in the case of metal theft.

A final paper in this area, by [Braakmann et al. \(forthcoming\)](#), moves in another direction on plausibly exogenous price movements by looking at burglaries when one recognizes that some groups in society keep valuables at home – in their case, UK households of South Asian origin having gold jewelry – and that criminals are aware of this. They very clearly show that as the gold price rises, burgling these households becomes a more lucrative business as the rising value of loot on offer generates increases in criminal earnings returns.

4 Criminal record impacts on the labor market

[Bushway et al. \(2022\)](#) studies the criminality of unemployed men in the 1997 National Longitudinal Survey of Youth (NLSY97): 64 % had been arrested at least once by age 35 for non-traffic offenses and 46 % had been convicted. Underlying these statistics are two potential causal pathways: (i) labor market experiences can impact criminal behavior and (ii) criminal behavior and records can impact employment. Of course, it may also be that there is no causal relationship at all but rather some common (un)observables that explain both criminal behavior and labor market outcomes. Besides being a source of endogeneity that researchers need to overcome to find causal evidence (of either pathway), these non-mutually exclusive simultaneous relationships are of fundamental concern for policy makers thinking about how to break the cycle of crime. This section surveys what we know about whether (and why) having a criminal record reduces labor market outcomes. We start by recognizing that observed employment is an equilibrium outcome of both the criminal worker's supply of labor and the firm's demand for labor.

On the supply side, low employment or earnings outcomes could arise because workers with criminal records are less likely to search for work or because their informal networks, which are generally known to play an important role in job finding, are weaker.³ Having a criminal record can also impact how workers search for jobs: for instance, they might be less likely to search if they know that potential employers will conduct background checks.

Alternatively, a criminal record, especially one that places the offender in prison, can weaken social capital; this possibility is highlighted by [Western et al. \(2001\)](#) as one mechanism through which incarceration can causally impact employment. To our knowledge, there is virtually no empirical evidence in economics on how having a criminal record (or the nature of that record) impacts an individual's job search behavior. Outside of economics, [Smith and Broege \(2020\)](#) find using ten survey rounds of the NLSY97 that interacting with the justice system decreases both overall job search and shifts job search from more active to passive methods. One reason for this sparse evidence is the data demands required. There are few contexts in which one can observe job search behavior (as opposed to outcomes), method of search, networks, and of course criminal records.

We have more evidence on the demand-side of the story – namely that poor employment outcomes for individuals with criminal records arise because firms are less willing to hire such workers. A 2003 survey (conducted by the Institute for Research on Labor and Employment) of California employers found that just 2 % would definitely hire a worker with a criminal record while 37 % would definitely not ([Raphael, 2014](#)). A more recent survey by [Cullen et al. \(2022\)](#), discussed in more detail below, found that just 39 % of US firms hiring temporary workers were willing to hire workers with a criminal record. [Raphael \(2014\)](#) highlights many reasons for being unwilling to hire a worker with a criminal record. (i) Most explicitly, state and federal laws ban individuals with records from certain jobs, such as those working with vulnerable populations or sensitive information.⁴ (ii) A firm may also exercise its own discretion in deciding not to hire an ex-offender due to risk aversion; firms may perceive the risks of loss due to employee theft or dishonesty or liability risks due to a harmed customer or employee to be higher for those with criminal records. Such risks again clearly should vary with the nature of the occupation. (iii) Alternatively, firms may be morally averse to hiring workers with criminal records. (iv) Finally, and in contrast to these examples of taste-based discrimination, firms may be statistically discriminating and using a criminal record as a signal of productivity.

We present three bodies of research. The first studies the impact of having a criminal record or interacting with the justice system on (equilibrium) employment outcomes. These studies largely use administrative records and quasi-experimental research designs. The second literature studies a firm's willingness to hire workers with criminal records using experimental variation in surveys or audit and correspondence studies. A third literature considers how policies, such as ban the box or clean slate reforms, impact employment outcomes for workers with criminal records.

³ See [Topa \(2011\)](#) and [Ioannides and Loury \(2004\)](#) for reviews of the theoretical and empirical evidence on the importance of informal networks in job finding.

⁴ [Bushway and Kalra \(2021\)](#) note that there are more than 4000 state licensing statutes that automatically disqualify people on the basis of their criminal history from obtaining a 'license'.

4.1 Effects of a record on labor market outcomes

The earliest work studying the effect of conviction and incarceration on employment and/or earnings was conducted prior to the first labor handbook chapter on crime (Freeman, 1999). Using data from the 1979 National Longitudinal Survey of Youth, which retrospectively asked about criminal justice involvement in the 1980 round, Freeman (1991) finds that the probability of employment is markedly lower after incarceration. Using administrative records for federally convicted offenders that contain an earnings measure from before and after conviction, Waldfogel (1994) finds that first time conviction significantly reduces employment and earnings. Though within individual comparisons of labor market outcomes go some way to controlling for correlated unobservables, these findings do not lend themselves to a causal interpretation due to the possibility of both simultaneity and/or remaining selection biases. Rather than having just one pre and post earnings measure, Grogger (1995) uses unemployment insurance records to construct a longitudinal data set of quarterly earnings data (from 1980 to 1984) that he merges with police records. With a flavor of contemporary event-study designs, Grogger estimates an individual fixed effects model of the effect of arrest on employment and earnings. He concludes that unobservable heterogeneity explains much of the labor market differences for arrested versus non-arrested individuals and there are only moderate, and short-lived, effects of arrest.

These results are in fairly sharp contrast to those of Agan et al. (2023) who conduct event study analyses of the impact of having a criminal record on taxpayer earnings and filing data using IRS records. Agan et al. use federal income tax data from 2000 to 2019, a period when online criminal record searches were common and feasible (in contrast to the 1980s period studied by Grogger). Agan and co-authors find a sharp drop in filing and reporting earnings after a criminal charge, and even after a charge that does not lead to a conviction.

The credibility revolution combined with increased access to administrative records and registers has led to a new body of work studying these old questions. Besides the modern-day event study design, three broad approaches have been used to overcome identification challenges. The first takes advantage of exogenous variation in record or sanction severity that is generated by randomly assigned judges (or other criminal justice agents). The second takes advantage of discontinuities in sentencing guidelines that lead to observably similar individuals – at least in terms of criminal histories and offenses – having different treatments. The third uses exogenous variation in treatment by the criminal justice system generated by the timing of reforms. Another advancement of the literature is to consider the labor market consequences of a wide range of justice system interactions: diversion, pre-trial detention, conviction, and incarceration.

Mueller-Smith and Schnepel (2021) study the effect of diversion away from the criminal justice system. Diversion is a tool used by justice agents to avoid a

criminal record: defendants avoid a formal conviction by undergoing a period of community supervision (at least in the Texas context studied here). But who gets diverted? To deal with the non-random assignment of diversion, Mueller-Smith and Schnepel study two reforms that led to sharp changes in the use of diversion in Harris County, Texas. The first decreased diversion in 1994 for offenders charged with certain drug and property offenses while the second increased diversion in 2007 for low-risk defendants. Using a fuzzy regression-discontinuity framework that takes advantage of the timing of a charge relative to the reform dates, the authors find that diversion improves long-term outcomes for first time felony defendants: not only do future convictions decrease but employment rates increase.

[Dobie et al. \(2018\)](#) study the impact of holding a defendant in pre-trial detention by matching Philadelphia and Miami-Dade court data to tax data from the Internal Revenue Service. Bail judges decide both whether to offer bail and the amount of bail. These decisions are again clearly not random but based on factors like the degree of evidence, flight risks, and potential danger to society. [Dobie et al. \(2018\)](#) capitalize on the quasi-random assignment of bail judges with different propensities to offer bail to isolate plausibly exogenous variation in pre-trial release. The authors find that pretrial release increases formal sector employment (by 25 % four years after the bail hearing) as well as the receipt of employment and tax-related government benefits.

The remainder of the papers in this literature focus on incarceration at either the extensive margin (any incarceration) or intensive margin (length of incarceration). Over and above the effect of conviction, prison can impact labor market outcomes through multiple channels. Prisons often include education, training and treatment, which may improve an offender's chance of finding (and keeping) a job when released from prison. Time in prison can impact criminal and non-criminal networks. To the extent that criminal networks on the outside are weakened, one can have more incentive to participate in the legitimate labor market; on the flip side, if non-criminal networks are weakened, then one can have less access to the labor market.

[Kling \(2006\)](#) conducted one of the first studies of the labor market effects of incarceration following the credibility revolution: he links quarterly earnings data (predominantly from the 1990s) to federal court and state prison data in California and Florida, respectively, and applies a randomly assigned judge fixed effects design, in which sentence length is instrumented for with the average sentence of all other offenders facing the same judge. [Kling \(2006\)](#) finds no evidence that more time in prison harms earnings and employment. This is seen both in descriptive plots of the data but also instrumental variable specifications; the estimates, however, are fairly imprecise. Despite its imprecision and minimal formal testing of the identifying assumptions that are a prominent part of contemporary 'judge fixed effects' designs, [Kling \(2006\)](#) played a significant role in making the randomly assigned harsh judge or 'judge fixed effects' instrument a go-to tool to overcome the non-random

nature of sanctions in the economics of crime.⁵ Another paper that finds an imprecise null effect of prison on employment using the harsh judge design is [Loeffler \(2013\)](#), which matches Cook County Illinois court records to unemployment insurance records.

The most recent work, however, by [Garin et al. \(2023\)](#) reaches the same conclusion: though employment and earnings are incapacitated while one is in prison, differences in labor market outcomes are not discernible by five years post sentence. These authors in fact reach these conclusions in two contexts – North Carolina and Ohio – using two identification strategies: sentencing guideline discontinuities and judge instruments, respectively. Both of these designs and contexts had been previously used to show crime reducing effects of the prison systems in North Carolina on the offender themselves ([Rose and Shem-Tov, 2021](#)) and in Ohio on the offender's family ([Norris et al., 2021](#)).⁶

But, in some contexts, prison appears to have beneficial labor market effects. Thus far, the common theme to these findings is Scandinavia. In Denmark, [Landersø \(2015\)](#) finds that longer incarceration spells improve employment and earnings outcomes. Intuitively, Landersø estimates a difference-in-difference specification where he compares the change in labor market outcomes for those who offended after a 2002 reform that decreased incarceration lengths by about one month (the treated group) to the change in outcomes for those who offended prior to the reform (the control group). In Norway, [Bhuller et al. \(2020\)](#) also find, using a judge stringency instrument, that incarceration (at the extensive margin) increases employment and earnings while reducing crime. These results are not seen for everyone, however, but rather for those not working prior to incarceration: for these same individuals, the authors find a 34 % point increase in job training program participation. In contrast, for individuals employed prior to prison, incarceration results in the loss of their job, which is not quickly recovered.

What is the effect of prison on labor market outcomes? The answer is not straight-forward. A general take-away from the literature is one of mixed findings: depending on the context and/or research design, researchers find positive, null, or negative effects of incarceration on labor market outcomes. These differences could be driven by heterogeneities in the marginal offender off which the results are identified, heterogeneities in the conditions and characteristics of the prison system studied, and heterogeneous effects across offenders.

⁵ See [Frandsen et al. \(2023\)](#) for a thorough discussion of identification issues, and proposed tests.

⁶ [Mueller-Smith \(2015\)](#) finds prison in Harris County Texas worsens labor outcomes and the analysis highlights potential violations of the exclusion restriction (i.e., the judge only affects outcomes via the prison sentence) and monotonicity assumption (if judges vary in their relative treatment of different types of defendants) in judge stringency designs.

4.2 Firm willingness to hire workers with criminal records

In addition to the firm surveys described above, there is also experimental evidence that at least some firms do not want to hire workers with criminal records. [Pager's \(2003\)](#) seminal paper conducts an audit study in Milwaukee, Wisconsin, in which there are two pairs of auditors (one Black pair and one white) that apply in person to jobs advertised by 350 employers. Both members of the pair apply to the same employer, but one member is assigned to have a criminal record. The results are striking: For White individuals, the chance of a callback for an interview decreased by 50 % when the auditor had a criminal record while, for Black individuals, the chance of a callback decreased by 60 %. The race-differential is even larger when accounting for the fact that Black auditors had a lower call-back rate overall. Another audit study conducted by [Uggen et al. \(2014\)](#) finds that auditors in Minneapolis, Minnesota are less likely to have a call-back even when they are assigned a minor criminal record that did not result in conviction (albeit the effect is smaller than in the original study). Finally, in a correspondence study design, which allows for thousands of job applications and holds constant all else but the criminal record, [Agan and Starr \(2017\)](#) show that employers in New York and New Jersey, whose job ads included a box asking about criminal history, were more than 60 % more likely to interview applicants without a felony conviction.

The best evidence to date on *why* firms are less willing to hire workers with criminal records comes from [Cullen et al. \(2022\)](#) who conduct a field experiment in partnership with a national staffing platform in the US: businesses submit job ads and the platform distributes them to qualified workers, who can accept the job on a first-come first-serve basis. Specifically, Cullen et al. ask hiring managers at nearly 1000 businesses their willingness to hire a worker with a criminal record under various hypothesized treatments in which the treatment intensity is randomized. These treatments speak directly to the various reasons that may underlie a firm's preferences, as well as to potential policies that could mitigate the firm's concerns. In the baseline (with no treatments), 39 % are willing to hire a worker with a criminal record. This statistic, however, varies with the nature of the job: it is 45 % if there is no customer interaction, 51 % if no high-value inventory, and 68 % if the job is hard to fill. What works to increase demand? The firm's willingness to hire workers with criminal records would significantly increase when provided (potentially prohibitively) large wage subsidies, relatively small amounts (up to \$5,000) of insurance coverage, information on satisfactory previous job experience, and a clean record for at least one year.

4.3 Policies to improve labor outcomes for workers with criminal records

Ban the Box policies aim to make it illegal to ask about criminal records on the initial job application (i.e., the box) or interview. The hope is that getting

workers with criminal records through the door will increase the chance of a job upon demonstrating their qualifications in an interview. Such policies were passed in at least 25 states and more than 150 jurisdictions ([Agan and Starr, 2018](#)). The theoretical effect of the ban the box policy on employment is not straight forward; moreover, it may depend on whether one actually has a criminal record. If it does get ex-offenders through the door and convince employers of a potential match, then employment could increase. But, if employers simply defer a background check until later in the process and are not willing to hire ex-offenders, then there may be no effect. Similarly, no effect could arise if workers with records sort across firms and/or industries that do and do not do background checks. Banning the box could even impact the employment outcomes of workers without criminal records. Depending on the group one belongs to (e.g., racial group), employment could increase or decrease via statistical discrimination. If firms do not want to hire workers with criminal records, they may use race, for instance, as a proxy via which to statistically discriminate.

The seminal paper on this question is by [Agan and Starr \(2018\)](#), which conducted an online correspondence study with employers in New Jersey and New York City before and after the adoption of ban the box policies in 2015. They submitted about 15,000 fictitious applications on behalf of young males. Applications were matched on race and randomly varied whether the applicant had a felony conviction. The main outcome studied, as is typical in correspondence studies, is callback rates. Call-back rates for workers with criminal records increases for both black and white applicants, suggesting that banning the box does get ex-offenders in the door. But, consistent with statistical discrimination on the basis of race, black and white call back rates decrease and increase, respectively, for workers without records.

A limitation of the Agan and Starr study, and correspondence studies in general, is that impacts on call backs may not translate into impacts on actual employment. This could happen if, for instance, background checks would have been done later in the process. Another potential limitation is that by construction the experiment has ex-offenders applying to the same jobs as non-offenders, which may not happen in reality. Subsequent papers have studied the impact of ban the box laws on employment in non-experimental contexts. Using variation in the timing of state and local ban the box laws and the Current Population Survey, [Doleac and Hansen \(2020\)](#) find that banning the box decreases employment for young, low-skill Black and Hispanic men. In contrast, using quarterly earnings data linked to statewide arrest and court records for 300,000 ex-offenders, [Rose \(2021\)](#) finds that Seattle ban the box laws did not improve labor market outcomes for ex-offenders. Other studies have found decreases ([Jackson & Zhao, 2016](#)) and increases in ex-offender employment ([Craigie, 2019](#)). The bottom line is that it is not clear what impact ban the box policies have on final labor market outcomes; these effects may depend on the context.

Another policy gaining traction is Clean Slate laws: ten US states have passed such legislation from 2018 to 2022.⁷ Common features of Clean Slate laws are the automation of record clearance, especially minor misdemeanor and/or arrest records. If employment gaps of workers with criminal records are attributable to background check failures, then clean slate laws could improve employment outcomes. On the other hand, if having a record and interacting with the justice system leads to gaps in labor market experiences and/or lower quality experiences, then there can permanent scarring effects on labor market outcomes. With the exception of one paper ([Agan et al., forthcoming](#)), there is minimal research on the labor market impacts of such policies. [Agan et al. \(forthcoming\)](#) study the impact of Proposition 47, which reduced certain felonies to misdemeanors. The authors focus on Joaquin County, where the nature of implementation of Proposition 47 created arguably exogenous variation in the timing of the automatic felony reduction. They also use an RCT in which a subset of individuals were notified about the record reduction to measure whether information plays a role in whether record reductions impact employment outcomes. Overall, the findings suggest little impact of reducing a felony to a misdemeanor on employment outcomes, even when notified about the reduction. There are a few exceptions, including an increase in platform gig work and employment increases for individuals with more recent convictions.⁸

5 Education and crime

One central plank in the significant rise of the economics of crime as a research field over the past 20 years has been the large body of work studying the link between crime and education. In fact, in the upsurge of research described above in [Fig. 1](#), a significant portion of the published papers can be classified into the crime and education area. [Fig. 4](#) shows, for all years covered, the principal subject matter of just under 7 % of the crime papers in the selected journals is about crime and education. Moreover, there were no papers about this when [Freeman's \(1999\)](#) chapter was published. In the selected leading journals, all of the work in this burgeoning area of research is from 2001 onwards. In the later years in the Figure, there are even more crime publications classified in the crime and education area – comprising just under one in ten of the total 438 economics of crime field publications in the set of journals considered from 2017–2023.

This section reviews this now sizable crime and education research area – an area that is growing and expanding in various directions. In offering this review, we first highlight the research that (very much in accordance with the

⁷ See <https://www.cleanslateinitiative.org/states>. Last accessed May 5, 2023.

⁸ Earlier work finds that record expungement leads to increased employment, but this is based on a sample of individuals who voluntarily select into expungement ([Prescott and Starr, 2020; Selbin et al. 2018](#)).

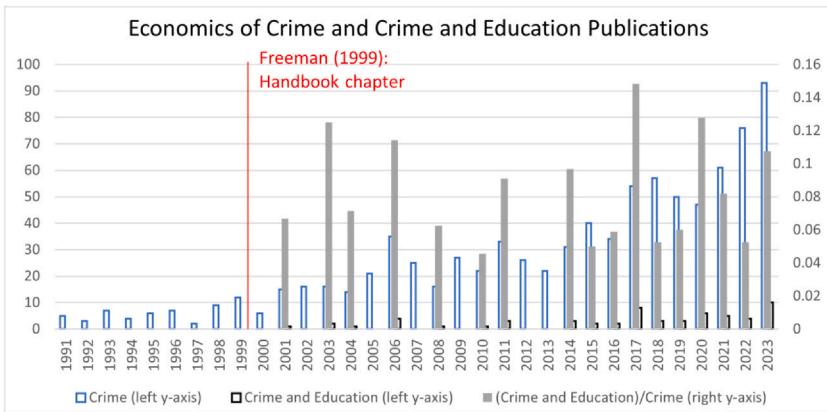


FIG. 4 Economics of crime and crime and education publications (1991 to 2023). Note – As for Fig. 1. From Machin and Sandi (forthcoming).

“credibility revolution” running through applied economics) focused on identifying causal impacts. Various research designs have been used to study both the causal impact of education on crime and of crime on education.

These are important developments as in prior work in a range of social science disciplines (most notably criminology, sociology and psychology, and also in others, though rarely in economics), researchers have used observational data of different sorts to document a (non-causal) negative correlation between education and crime. This has been shown through time in many settings, and for most (but not all) crime measures. But the economics literature on which we focus tackles the causality issue head on, with significant research efforts and intention to move from correlational observation to establishing causality.

The key question on establishing causality is that both the following questions are plausible. Does education reduce crime? And/or does crime reduce education? A series of studies have by now quite convincingly established that there is a causal impact of education on crime by leveraging education policy changes to ensure there is a crime-reducing crime impact of education. Other research designs, to be discussed, have been used to corroborate this and reach a similar conclusion. So has the work showing causation running from crime to education, where juveniles who interact with the justice system, and especially prison, have causally worse education outcomes. The fact that both causal directions are plausible, clearly means that a solid identification strategy is needed, and we will discuss these carefully when reviewing the literature. Besides the simultaneous nature of this relationship, establishing causality is challenging given there are many potential unobservables correlated with both educational attainment and criminal behavior (e.g., ability and family background).

Why can education reduce crime? The first, immediately forthcoming response to this question comes from the economics of crime framework set out earlier in this chapter. It is straightforward to consider crime and education in this setup by making legal earnings W_L a function of education (e.g., through a Mincerian earnings function) so that, coupled with the large literature on positive wage returns to education (Card, 1999), there is a clear prediction that more education reduces crime.

This is not the only route by which a crime reducing impact of education can emerge. Nor need it be an income effect that generates the crime reduction. Other possibilities speak more closely to mechanisms by which an impact may emerge. In the income effect route, the crime reduction arises because extra time spent in school raises productivity and this is rewarded in the labor market by higher earnings, which reduce crime in the usual Becker/Ehrlich fashion. On the other hand, being in the classroom means that juveniles are in school being kept busy in a supervised environment and, thus, off the streets and not committing crime. This is an incapacitation effect, which need have no productivity implication, and thus offers a different crime reduction mechanism. And another possibility, with a more behavioral aspect to it, is that potential future criminals differ in their discount rates (rather like in the Card, 1999, decisions on whether to invest in education or not), so that they value the present more than the future. Schooling therefore not only increases economic returns but may also increase a youth's patience, and lead them to put more weight on their potential future earnings (as in Becker and Mulligan, 1997).

Table 1 collates 64 crime and education publications from the past 20 years, ending in 2023, into four Panels that reflect the way the empirical literature has evolved over time and conveniently define the sub-sections below. Panel A alphabetically lists 12 studies that focus upon crime-education connections where crime is related to completed schooling levels. For the most part, these were the first round of studies on crime and education that aim to pin down a causal impact of education. Some of the more well-known studies implement research designs that exploit school dropout age reforms to ensure the direction of causation runs from education, specifically attendance at school, to crime. Panel B shows a further 25 studies that move beyond school attendance to look at the impact of other education measures. Some of the contributions in this Panel focus on the productivity and incapacitation aspects of the crime-schooling relation. The others, many of which have been published very recently and are an important feature of the big uptick of research in this area shown in Fig. 4, study connections between criminality and a wider range of education measures, including both school quantity and quality. Panel C lists 15 studies that look at causality running the other way, by studying the impact of crime on education. Panel D contains 12 studies that look at the impact of policies and interventions connected to education on crime.

TABLE 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Panel A. Crime and Education, Completed Education and Dropout Age Laws						
Anderson (2014)	US	FBI's Uniform Crime Reports (UCR) from 1980–2008; the National Center for Education Statistics' Digest of Education Statistics; U.S. Census Bureau; Bureau of Economic Analysis; Youth Risk Behavior Survey	Secondary school attendance	Minimum Dropout Age (MDA) reforms to specify a difference-in-difference-in-differences (DDD) approach	Compared with MDA 16–17 laws, MDA 18 laws correlated with ↑ arrest rates of youth aged 13–15 and 16–18 property and a ↓ 22.5 % violent crime arrest rates of youth aged 16–18	Compared with MDA 16–17 laws, MDA 18 laws lead to a ↓ 17.2 % arrest rates, of which a ↓ 9.9 % property and a ↓ 22.5 % violent crime arrest rates of youth aged 16–18
Beattie et al. (2017)	Queensland, Australia	Queensland administrative data from the Department of Education and Training matched at the individual level with the Queensland Police Service from 2002–13	Secondary school attendance	Minimum Dropout Age (MDA)	Exposure to the MDA reform led to a ↓ 10.3 % crime for all 15–21 year olds, to a ↓ 10.8 % crime for males and to a ↓ 8.9 % crime for females	Upper secondary school completion leads to a ↓ 23 pp crime and to a ↓ 9.8 pp conviction for males, as well as to a ↓ 9 pp crime and to a ↓ 3.2 pp conviction for females
Bennett (2018)	Denmark	Administrative Danish Register Data on twins born in 1965–82	Upper secondary school completion	OLS on sample of twins and with twin fixed effects		

<p>Bell et al. (2022)</p>	<p>US</p>	<p>FBI Uniform Crime Report (UCR) from 1974 onwards for males aged 15–24 years old</p>	<p>Secondary school attendance</p>	<p>Minimum Dropout Age (MDA) to define a Regression Discontinuity (RD) design</p>	<p>Cohorts exposed to the MDA reforms face a ↓ 4.6% arrest rate</p>	<p>Exposure to the MDA reforms leads to a ↓ 6.1 % log arrest rates for young adults</p>
			<p>Secondary school attendance</p>	<p>Minimum Dropout Age (MDA) to define an Instrumental Variable (IV) model in a difference in difference specification</p>		

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Buonanno and Leonida (2006)	Italy	Annual panel dataset for Italian regions (NUTS2) from 1980-95; Centre for North South Economic Research (CRENOS); ISTAT (Italian Statistics Bureau) population data	Secondary school attendance	Instrumental variable approach for panel data using the GMM-system estimator	Education, measured as the average years of schooling of the population, is linked with ↓ crime rate	
Buonanno and Leonida (2009)	Italy	Annual panel dataset for Italian regions (NUTS2) from 1980-95; Centre for North South Economic Research (CRENOS); ISTAT (Italian Statistics Bureau) population data and ISTAT Quarterly Labor Force Surveys	Secondary school attendance	Region and time fixed effects as well as region-specific time trends	Education, measured as the average years of schooling of the population, is linked with ↓ property crime rate	

Cano-Urbina and Lochner (2019)	US	US Census data from 1960–1980; 1960–90 FBI's Uniform Crime Reports	Educational attainment and school quality	Minimum Dropout Age (MDA) to define an Instrumental Variable (IV) model	One additional year of schooling leads to ↓ 0.04 pp incarceration rates for White women and ↓ 0.08 pp incarceration rates for black women
Gilpin and Pennig (2015)	US	School Survey on Crime and Safety (SSOCS, 2004, 2006, 2008, 2010)	Secondary school attendance	Reduced form analysis of Minimum Dropout Age (MDA) laws in a difference in difference specification	High schools in states that raise their MDA law to 18 experienced ↑ 21.4 % school crimes
Hjalmarsson et al. (2015)	Sweden	25 % random sample from Sweden's Multigenerational Register on all persons born from 1943–54 who have lived in Sweden at any time since 1961 matched with data on parents, siblings and children; Sweden's Education Register and 1970 Census of Sweden; Sweden's National Council for Crime Prevention records of all criminal convictions from 1973–2007 for each individual	Educational attainment	Minimum Dropout Age (MDA) to define an Instrumental Variable (IV) model	One additional year of schooling leads to ↓ 6.7 % risk of conviction of men and to ↓ 15.5 % incarceration of men, while estimates for women are similar in magnitude but not statistically significant

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Lochner and Moretti (2004)	US	US Census data from 1960–1980 on males aged 20–60; 1960–90 FBI's Uniform Crime Reports; National Longitudinal Survey of Youth	Educational attainment and school quality	Minimum Dropout Age (MDA) to define an Instrumental Variable (IV) model in a difference in difference specification	One additional year of schooling leads to ↓ 0.1 pp incarceration for White men and ↓ 0.37 pp incarceration for black men	One additional year of schooling leads to ↓ 0.1 pp incarceration for White men and ↓ 0.3–0.5 pp incarceration for black men
Machin et al. (2011)	England and Wales	Offenders Index Database (OID) on criminal histories of offenders aged 18–40, born in 1946–70 and convicted of standard list offences from 1963 onwards; Office for National Statistics (ONS) population data by age cohort and year, separately for men and women; the UK General Household Survey (GHS) for the 1972–96 years; 2001–08 British Crime Survey	Secondary school attendance and school qualification	Minimum Dropout Age (MDA) to define a Regression Discontinuity (RD) design	A 10 % increase in age left school is correlated with ↓ 2.1 % for males	A 10 % increase in age left school leads to ↓ 2.1 % crime for males

Panel B. Crime and Education, Quantity/Quality					
Barrett et al. (2021)	Louisiana, US	Student-level data provided by the Louisiana Department of Education (LDOE) from 2000–14	Student discipline disparities by race and family income	Black (poor) students are ↑ 13.9 pp more likely to be suspended in a given year than white (non-poor) students	An increase of 20 pp in the municipal share of full-day high schools leads to a ↓ 3.3 % in the probability of motherhood in adolescence and to a ↓ 11 % to 24 % in the juvenile crime rate
Berthelon and Kruger (2011)	Chile	Chile's National Socio-economic Characterization Survey, CASEN since 1990; administrative data from the Defensoría Penal Pública, i.e., the Chilean equivalent of a Public Defender's office	Length of the school day	School reform that lengthened the school day from half to full-day shifts	Schools' impacts on high-stakes tests are weakly related to impacts on arrest, dropout, teen motherhood, and formal labor market participation
Beuermann et al. (2023)	Trinidad and Tobago	Official administrative Secondary Entrance Assessment (SEA) data for all applicants to a public secondary school in Trinidad and Tobago from 1995–2012; data on the NCSE exams (age 14), the CSEC exams	Exogenous school assignments	Whether schools' impact on test scores measure their overall impact on students	

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Billing et al. (2014)	Charlotte-Mecklenburg, North Carolina, US	(age 16) and the CAPE exams; official arrests records from the Trinidad and Tobago Police Service; official birth records from the Trinidad and Tobago Registrar General; official registry of active contributors to the national retirement fund by May 2017 of National Insurance Board	Administrative Charlotte-Mecklenburg school records (1995-2011); administrative records of adult arrests and incarcerations in Mecklenburg County (1998-2011); National Student Clearinghouse records of college attendance	School share minority	Discontinuous school boundary change	10 pp increase in assigned school share minority leads to ↑ 8% arrest and incarceration among minority males

			A 5 pp increase in school peers linked to parental arrest leads to ↓ 0.016 standard deviation in school achievement and to ↑ 5 % in adult arrest rates		<i>Continued</i>
Billings and Hoekstra (2023)	Charlotte-Mecklenburg, North Carolina, US	Pupil records for Charlotte-Mecklenburg schools from 1998-2011; Mecklenburg County arrest records from 1998-2014; Mecklenburg County jail records from 1998-2014; North Carolina state prison records from 1998-2014	School and neighborhood peers whose parents have been arrested	Cohort variation	

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Cullen et al. (2006)	Chicago, US	Chicago public schools (CPS) administrative data on applications in spring 2000/01; achievement and attainment of CPS students, student survey on degree of satisfaction with school, students treatment by teachers and peers, college expectations, arrest	High-achieving school attendance	Admission lottery into school		Lottery win to high-achieving school leads to nearly ↓ 60% self-reported arrest rates relative to lottery losers
Denning (2011)	Charlotte-Mecklenburg, North Carolina, US	Administrative Charlotte-Mecklenburg school records for students in grades 6–11 in 2002 and age 17–23 in 2009; administrative records of adult arrests and incarcerations in Mecklenburg County (2006–2009); NC Department of Corrections from 2006 on	Better-achieving school attendance	Admission lottery into school		Lottery win to better-achieving school leads to roughly ↓ 50% criminality among high-risk youth relative to lottery losers

Depew and Eren (2016)	Louisiana, US	Administrative records from the Louisiana Department of Education from 1997-2012; administrative data from the Louisiana Department of Public Safety and Corrections, Youth Services, Office of Juvenile Justice	School entry age	Parametric fuzzy Regression Discontinuity (RD) Design	ITT estimates show that being born right after the school entry cutoff leads black females to a 1.3 pp in likelihood to commit a juvenile crime
Dobbie and Fryer (2015)	New York City, US	Survey data from Youth entered in the 2005/06 Promise Academy sixth grade admissions lotteries in the Harlem Children's Zone; administrative data on high school test-taking from the New York City Department of Education and college enrollment data from the National Student Clearinghouse	High-performing charter school attendance	Admission lottery into school	Lottery win to high-performing charter school leads to ↓ 10.1 pp teenage pregnancy and ↓ 4.4 pp male incarceration

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Eren et al. (2022)	Louisiana, US	Administrative records of the Louisiana Department of Education (LDOE); Louisiana Department of Public Safety and Corrections, Adult Services, from 1996–2012		Regression discontinuity design to study test-based promotion policy		Grade retention leads to ↑ 1.05pp in the likelihood of conviction by age 25
Eriksson (2020)	US	Linked census data set of incarcerated and nonincarcerated men	Childhood access to primary schooling	School construction		Exposure to one new primary school built as part of the Rosenwald programme leads to ↓ 1.9 pp risk of incarceration

Figlio (2006)	Florida, US	Administrative dataset on every disciplinary suspension, both in-school and out-of-school, during the four school years from 1996–97 through 1999–2000, i.e., following the introduction of the Florida Comprehensive Assessment Test (FCAT)	High stakes testing	Interaction between the testing calendar, the grade level of the student, and the expected performance level of the student	A one standard deviation increase in the test window manipulation measure is associated with a ↑ 1.2 pp in the likelihood that a student will attain level 2 or better on the FCAT reading exam and a ↑ 1.7 pp in the likelihood that a student will attain level 2 on the FCAT mathematics exam		
García et al. (2023)	US			Newly collected data on the original High Scope Perry Preschool Project participants through late / middle age and on their children into their mid-twenties	High Scope Perry Preschool Project	The intervention led to long-lasting ↑ 0.2–0.4 standard deviation in the original participants' skills, ↑ 10 pp more likely to be married at age 30, ↑ \$10,000 average annual earnings, ↓ 1	

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Gaviria and Raphael (2001)	US	National Education Longitudinal Survey (NELS) 1988–94	Peer group influences	IV estimates using average background characteristics as IV	Drug use by parents correlated with ↑ probabilities of drug, alcohol, and tobacco consumption by their children by 19.4%, 13.2%, and 10.2%, respectively	Drug use by parents correlated with ↑ probabilities of drug, alcohol, and tobacco consumption by their children by 19.4%, 13.2%, and 10.3%, respectively

Gray-Lobe et al. (2023)	Boston, US	<p>All preschool applicants from fall 1997 to fall 2003 from the Boston Public Schools district; National Student Clearinghouse (NSC) data; administrative data from the Massachusetts Department of Elementary and Secondary Education (DESE)</p> <p>Public preschool attendance</p> <p>Preschool enrollment leads to ↑ 18% in college attendance, ↑ 9 pp in SAT test-taking and ↑ 6 pp in high school graduation. Preschool also leads to ↑ 0.17 standard deviation in disciplinary measures including juvenile incarceration, but with no detectable impact on state achievement test scores</p>	<i>Continued</i>
----------------------------	------------	---	------------------

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Heckman et al. (2006)	US	National Longitudinal Survey of Youth 1979	Boosting cognitive and noncognitive skills	Simulations		Moving males in lowest decile of cognitive distribution from lowest to highest decile of noncognitive distribution substantially ↓ incarceration. Moving same males in lowest deciles of both distributions to highest decile of cognitive distribution only slightly ↓ incarceration

Hutunen et al. (2023)	Finland	<p>Finnish joint application registry for cohorts who graduated from compulsory schooling from 1996–2003 and applied to further education immediately upon graduation; population-wide administrative registers from Statistics Finland from 1995–2013; the Finnish Longitudinal Employer-Employee Data (FLEED); Student Register and the Register of Completed Education and Degrees; Prosecutions, Sentences and Punishments based on the district court rulings</p> <p>Secondary education</p> <p>Admission cut-offs in over-subscribed programmes to generate Regression Discontinuity (RD) designs</p> <p>Admission of men to secondary schools leads to ↓ 52% risk of conviction in a district court within 10 years after admission compared with men who are not admitted</p>	Continued
--------------------------	---------	---	-----------

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Jacob and Lefgren (2003)	US	29 urban jurisdictions in NIBRS data, detailed school calendar	School attendance	Teacher in-service days		Youth property crime ↓ 14% on days when school is in session, while violent crime ↑ 28% on such days
Jackson et al. (2020)	Chicago, US	Administrative data from Chicago public schools on 133 public high schools for cohorts of ninth grade students who attended one of these schools in 2011–17	Attendance of schools with high socio – emotional development (SED) value added	Covariance of school value added across outcomes		Higher social value added leads to ↓ 0.728 pp risk of arrest; greater work hard value added leads to ↓ 0.766 pp risk of arrest; greater test score value added leads to ↓ 0.523 pp risk of arrest

Johnson and Jackson (2019)	US	<p>Panel Study of Income Dynamics (PSID); National Archives Record Administration, Inter-university Consortium for Political and Social Research, and Surveillance, Epidemiology, and End Results population data</p> <p>Early childhood exposure to investments designed to promote school readiness among disadvantaged children</p>	<p>Head Start and K–12 spending</p> <p>Start led to ↑ 0.59 additional years of education, ↑ 14.8 pp likely to graduate high school, ↑ 17% higher wages, ↓ 4.7 pp likely to be incarcerated, and ↓ 12 pp less likely to be poor as an adult</p>	<p>For poor children exposed to a 10% increase in K–12 spending,</p> <p>DEE leads to ↓ criminality at (by) all ages until age 19 (22) for boys and at (by) age 15 (19) for girls</p> <p>Roughly 30% (3.3%) of young men with < 10 (≤ 11) years of schooling earn income from crime. Among high school graduates, 24% of</p>
Landersø et al. (2017)	Denmark	<p>Danish register-based data for children born in mid–1981 to mid–1993</p>	<p>Delayed Entry Eligibility (DEE)</p>	<p>Discontinuous minimum school-entry age</p>
Lochner (2004)	US	<p>National Longitudinal Survey of Youth 1979 and FBI's Uniform Crime Report (UCR)</p>	<p>High school dropout</p>	<p>OLS</p>

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Luallen (2006)	Washington State, US	Administrative Washington Juvenile Court Case Records 1980–2001, Census 2000 Summary File 3 and 1990 Summary Tape File 3, Public Employee Strikes in Washington 1967–2003, news articles from the Associated Press and the Seattle Times	School attendance	Teacher strikes	Youth crime ↑ 21.4% on days when strikes occur. In particular on such days, mischievous crime ↑ 48%, property crime ↑ 28.8% and violent crime ↓ 31.5%	

McEachin (2020)	North Carolina, US	Administrative data of the North Carolina Department of Public Instruction for all students in North Carolina public schools in 2004–16; administrative data from the North Carolina Department of Public Safety and population-level records from the North Carolina Board of Elections	Charter school attendance	Doubly-robust inverse probability weighted approach	Compared with students who attended a traditional public school in both 8th and 9th grade, charter school entrants face ↓ 0.9 pp risk to commit any crime, and ↓ 0.7 and ↓ 0.4 pp risk to be convicted for a misdeme- anor and felony off bases of 3, 0.2, 1.3 pp	
Panel C. Crime Effects on Education						
Aizer and Doyle (2015)	Chicago, US	Chicago Public Schools Student Database (1990–2006); Juvenile Court of Cook County Delinquency Database (1990–2006); Illinois Department of Corrections Adult Admissions and Exits Database (1993–2008)	Youth incarceration	High school graduation ↓ 39 pp, adult incarceration ↑ 41 pp	Youth incarceration leads to ↓ 13 pp high school graduation and ↑ 23 pp adult incarceration	

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Ang (2021)	Los Angeles, US	Incident-level data on the universe of officer-involved killings in LA County, California, from 2002–16; home addresses and individual-level panel data for all high school students enrolled in the LA Unified School District	Police killings	Granular variation in how close students live to a killing		Exposure to police violence leads to ↓ 0.04 points in GPA, ↑ 15% incidence of emotional disturbance, ↓ 3.5% rates of high school completion and ↓ 2.5% college enrollment
Arteaga (2023)	Colombia	Colombia's census of potential beneficiaries of welfare (SISBEN); Attorney General's Office records; internet records scraped by the author	Parental incarceration	Judge IV		Parental incarceration leads to ↑ 0.78 years in educational attainment for children of convicted parents
Brown and Velásquez (2017)	Mexico	INEGI monthly homicide reports at the municipal level and Mexican Family Life Survey	Drug-related violence		Surge in drug-related crime	Increased local violence leads to ↓ 0.3 years of education, ↓ 3 pp likelihood of compulsory school completion, and ↑ likelihood of employment

Brück et al. (2019)	West Bank	<p>MOEHE administrative records from 2000–06; Israeli NGO B'Tselem (Israeli Information Center for Human Rights in the Occupied Territories); Palestinian Labor Forces Survey (PLFS) for the period 2000–06</p>	<p>Effect of the Israeli–Palestinian conflict on education outcomes</p>	<p>The conflict leads to ↓ probability of passing the final exam, ↓ in the total test score, and ↓ in the probability of being admitted to university</p>
				<p>Within school variation in the number of conflict-related Palestinian fatalities in the academic year</p> <p>Family problems, as signaled by a request to the court for protection from domestic violence, used as exogenous source of variation in peer quality</p>

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Carrell et al. (2018)	Florida, US	Administrative records from Alachua County (Florida) primary schools from 1995–1996 and 2002–2003 from the Florida Department of Education (FLDOE). Domestic violence cases filed in civil court in Alachua County between January 1, 1993 and March 12, 2003; National Student Clearinghouse (NSC) records from 2012	Exposure to a disruptive peer in elementary school	Variation in cohort composition across time within school	Grade retention ↑ 3 pp the propensity of a student to drop out of school	Exposure to a disruptive peer in classes of 25 in elementary school leads to ↓ 3% earnings aged 24–28
Eren et al. (2017)	Louisiana, US	Administrative records of the Louisiana Department of Education (LDOE) from 1999–2012; Louisiana Department of Public Safety and Corrections, Youth Services, Office of Juvenile Justice	Summer school and grade retention	Regression discontinuity design to study test-based promotion policy		

Eren and Mocan (2021)	Louisiana, US	Louisiana Department of Public Safety and Corrections, Youth Services, Office of Juvenile Justice from 1996–2012;	Louisiana Department of Public Safety and Corrections, Adult Services from 1996–2012	Impact of juvenile punishment on adult criminal recidivism and high school completion	Judge IV	Negative effect on high school completion for earlier cohorts, but no impact on later cohorts; juvenile incarceration leads to 1.27 pp in the probability of adult conviction of a drug offense; null effect for violence	
Foureaux- Koppensteiner and Menezes (2021)	São Paulo, Brazil	Brazilian school census data collected by Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP,	National Institute for Educational Studies and Research “Anísio Teixeira” on behalf of the Brazilian Ministry of Education; Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (SARESP; the education evaluation system of the state of São Paulo); and individual attendance records in all state schools	Effect of exposure to homicides around schools, students' residences, and on way to school	Variation in homicides within a 25-meter radius around schools	Violence leads to ↓ 5% standard deviation test scores and ↑ 20% dropout rates	

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
	from the São Paulo State Secretariat of Education; Brazilian Ministry of Health records					
Harlow (2003)	US	Survey of Inmates in State and Federal Correctional Facilities 1991 and 1997; Survey of Inmates in Local jails 1989 and 1996; Survey of Adults on Probation 1995; Current Population Survey 1997; National Adult Literacy Survey 1992	Education achievement of inmates		Inmates' education links with ↓ youth sentencing, as roughly 40% without high school diploma, 45% with a GED, 26% with a high school diploma, 21% with some college had prior youth sentences either to a facility or probation	
Hjalmarsson (2008)	US	National Longitudinal Survey of Youth 1979	Arrest, charge, conviction, incarceration at age 16 or younger	OLS	Arrests lead to roughly ↓ 11 pp likelihood of graduation, and incarcerations lead to roughly ↓ 26 pp likelihood of graduation	When correcting for unobservables, incarcerations still lead to roughly ↓ 13 pp likelihood of graduation

Michaelsen and Salardi (2020)	Mexico	<p>Ministry of Health (Secretaría de Salud) records of violence; ENLACE - Evaluación Nacional del Logro Académico en Centros Escolares - data on performance of primary school students on national standardized exams from 2006–11</p>	<p>Exposure to violence</p>	<p>Exogenous variation generated by “War on Drugs”</p> <p>Exposure to at least three homicides within a 2 km radius in the week immediately prior to exams leads to ↓ 0.1 in standard deviation of test scores</p>
Padilla-Romo and Peluffo (2023)	Mexico		<p>ENLACE (National Assessment of Academic Achievement in Schools) data from 2005–13; INEGI (National Institute of Statistics and Geography of Mexico) data; CONAPO (National Population Council) data</p>	<p>Out-migration from violence-affected areas and peer exposure to violence</p> <p>Mexican war on drugs</p> <p>Adding a new peer who was exposed to local violence to a class of 20 students leads to ↓ 2% standard deviation in incumbents' academic performance</p>

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Pope and Zuo (2023)	Los Angeles, US	Student-level administrative data from the Los Angeles Unified School District	School suspension	Changes in school suspension policies	10 pp lower suspension rates lead to ↓ 0.040 standard deviation in math and ↓ 0.065 standard deviation in English test scores, ↓ 0.07 standard deviation in grade point averages and ↑ 15.1% absences	Improvements to early childhood education led to ↑ 20% reductions in the likelihood of a serious criminal conviction in adulthood
Panel D. Policy and Interventions						
Anders et al. (2023)	North Carolina, US	Administrative conviction data from North Carolina's Department of Public Safety, 1972–2018; Head Start and Smart Start funding information from the National Archives and Records Administration (NARA)	Early childhood education	Rollout of Head Start and Smart Start		

Anderson and Sabia (2018)	US	Youth Risk Behavior Surveys (YRBS) from 1993–2013	Youth gun carrying and mass shootings	Child access prevention (CAP) gun controls laws	CAP laws lead to a ↓ 13% in the rate of past month gun carrying and a ↓ 18% in the rate at which students report being threatened or injured with a weapon in school
Blattman et al. (2017)	Liberia	Survey data collected for the evaluation of the intervention	Cognitive behavioral therapy to foster self-regulation, patience, noncriminal identity, lifestyle, and \$200 grants	Randomized allocation of treatment	Cash after therapy led to ↓ 0.31 standard deviation antisocial behavior for over a year
Foged et al. (2023)	Denmark	Administrative records on demographics and school and crime records of youth born in 1990 to 2001, still in Denmark at age 18, and with at least one parent granted asylum in Denmark within four years around 1st January 1999	Parental language training for refugees	Reform to expand language training for adult refugees	Parental language training of refugees leads to ↓ 72.7% convictions and ↓ 80.8% charges of male children aged 15–18

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Guleisci et al. (2021)	Bolivia	Data collected for the purpose of the intervention	Impact of a youth empowerment programme on the reported prevalence of violence against girls			The youth empowerment programme led to ↓ 10pp in the reported prevalence of violence against girls during the COVID-19 lockdown
Heller (2022)	Chicago, Philadelphia, US	Chicago and Philadelphia administrative police records of arrests; in Philadelphia, service records from the City's integrated data system, known as CARES, to measure juvenile incarceration (including both detention and prison) and related court ordered services	Summer youth employment programmes (SYEP)	Admission lotteries		In Philadelphia [Chicago], being offered the programme leads to ↓ [9] arrest per 100 youth (i.e., 36%) [i.e., 52%]. Due to the size of first stage, the effect on compliers in Philadelphia is ↓ 3 arrests per 100 youth (i.e., 65%)

Heller et al. (2017)	Chicago, US	Longitudinal student-level CPS records, Illinois State Police (ISP) records and arrest data from the Chicago Police Department (CPD)	Becoming A Man (BAM) programme Set of randomized controlled trials (RCTs)	The BAM programme reduced total arrests during the intervention by 28–35%, reduced violent-crime arrests by 45–50%, raised school engagement and graduation rates by 12–19%
Owens (2017)	US	National Incident Based Reporting System (NIBRS) from 1997–2007; COPS office from 1994–2007; Uniform Crime Reports Law Enforcement Officers Killed and Assaulted (LEOKA); Law Enforcement Management and Administrative Statistics (LEMAS)	Use of sworn School Resource Officers (SROs)	Federal hiring grant to place law enforcement in school The average grant is linked with ↓ 1.1–1.9% disruptive criminal incidents in school

Continued

Table 1 Literature in economics on the link between crime and education in the last 20 years (up to 2023).—Cont'd

Study	Setting	Data	Treatment	Research design	Correlation	Causal impact
Rees et al. (2022)	US	Youth Risk Behavior Surveys (YRBS) data from 2009–17; National Vital Statistics System (NVSS) from 1993–2016	Bullying victimization	State-level anti-bullying laws (ABLs)		State-level anti-bullying laws (ABLs) lead to ↓ 2.6 pp in bullying victimization, ↓ 1.9 pp in depression, and ↓ 1.7 pp in suicidal ideation
Sabates and Feinstein (2008)	England and Wales	Home Office Offenders Index database (OI) records of criminal convictions in England and Wales for 1996–2002 samples	UK government initiatives: Reducing Burglary Initiative; and Educational Maintenance Allowance	OLS with difference in differences design		Areas that introduced both programmes faced ↓ 1.1–1.5 convictions for youth aged 16–18 years old for burglary per 1000 relative to areas where neither programme was introduced

Sorensen et al. (2023)	US	2013/2014 and 2017/2018 waves of the CRDC from all public schools in the US, except preschools and schools with < 25 students; Freedom of Information Act (FOIA) request to the U.S. Department of Justice COPS office	School resource officers (SROs)	Fuzzy regression discontinuity design	SROs lead to ↓ 30% in violence in schools, but do not prevent gun- related incidents
Weisburst (2019)	Texas, US	Texas Education Research Center (ERC) records, Texas Education Agency (TEA) records and Texas Higher Education Coordinating Board (THECB)	Funding for school police on student outcomes	Federal Community Oriented Policing Services (COPS) grants	Federal grants for police in schools lead to ↓ 6% middle school discipline rates, ↓ 2.5% in high school gradu- ation rates and ↓ 4% in college enrollment rate

5.1 Causal impacts of education on crime

The common feature of the 12 published studies shown in Panel A of [Table 1](#) is their focus on the impact of completed education (usually, though not always, measured by years of education/secondary school dropout age) on crime. These include the pioneering causal studies that look at the impact of school attendance on crime by leveraging legislative changes in compulsory school leaving laws to measure a causal impact of completed education on crime.

The seminal US paper of [Lochner and Moretti \(2004\)](#), and the studies of England and Wales by [Machin et al. \(2011\)](#) and of Sweden by [Hjalmarsson et al. \(2015\)](#) are key highly cited papers in this area. The logic they adopt in the empirical research designs to ensure causality runs from education to crime is that raising the school leaving age was implemented as policy to improve education levels, with no obvious connection to crime. Thus, in a standard two-stage causal regression setup, the change in the dropout age measure is used only to predict education and acts as a legitimate instrumental variable for completed schooling in the crime equation. These studies carefully justify this use of education policy – the raising of the school leaving or dropout age – to generate crime reductions from education as an unintended consequence of the compulsory school leaving age legislation.

The US study by [Lochner and Moretti \(2004\)](#) relates prison, arrests and self-reported crime data to education. They leverage state level shifts in the US compulsory school leaving age matched to 1960, 1970 and 1980 Census data, to Uniform Crime Reports arrests data, plus self-reported NLSY crime. In their Census analysis, for example, they look at males aged 20–60. Their study shows a strong first stage, where being exposed to an increase in the school dropout age in the state where you were schooled significantly raises completed years of schooling. The reduced forms show a significant reduction in subsequent crime for treated birth cohorts. In their analysis of age-specific arrest data, their findings imply that an additional year of schooling reduces state level arrest rates by at least 11%, with similar effects for both violent and property crime. Causal crime reduction effects, in the same research design leveraging state-specific changes in compulsory school leaving ages, also feature in the Census imprisonment outcome and for the self-report individual-level NLSY data.

[Machin et al. \(2011\)](#) study causal connections via the raising of the compulsory school leaving age in England and Wales in 1973 from 15 to 16 in a regression discontinuity design. They (mostly) study men aged 18 to 40. By comparing cohorts who turn 15 immediately before and after the reform, they estimate that a one-year increase in average schooling reduces property crime conviction rates by up to 30%. [Hjalmarsson et al. \(2015\)](#) study the causal effect of educational attainment on conviction and incarceration using Sweden's compulsory schooling reform as an instrument for years of schooling. Their

study of men and women aged 19 to 64 reports that one additional year of schooling in Sweden decreased the probabilities of conviction and incarceration for males by about 7% and 15%, respectively.

A crime-reducing impact of completed education is corroborated by the other studies in Panel A of [Table 1](#). These either look at more up to date dropout age reforms in the US ([Anderson, 2014](#); [Bell et al., 2022](#); [Cano-Urbina and Lochner, 2019](#); [Gilpin and Pennig, 2015](#)), or consider crime and education in other settings ([Beaton et al., 2017](#) in Australia; [Bennett, 2018](#), in Denmark; [Brilli and Tonelli, 2018](#), [Buonanno and Leonida, 2006, 2009](#) in Italy). The consensus reached in this research is of a beneficial crime-reducing effect of longer duration school attendance. The extended compulsory school period is particularly effective at reducing property crimes, while little evidence exists of a lasting violence-reducing effect of the extended compulsory school period. But the strong conclusion of a causal crime reducing impact, based on education policy reform, emerges. Interestingly, since the economic costs of crime are high, these crime reduction benefits contribute very strongly to the benefit-cost ratios of the dropout age policies being well above unity, indicating that these reforms generate crime reductions whose socio-economic benefits outweighs their costs (for example, see the calculations presented in the papers by [Bell et al., 2022](#), [Lochner and Moretti, 2004](#), and [Machin et al., 2011](#), for more detail).

5.2 Incapacitation

Other crime and education research looks not at completed education and its impact on adult criminality, but rather studies (i) crime when individuals are still in the compulsory schooling system or (ii) the impact of aspects of schooling that occur prior to education completion. Panel B of [Table 1](#) lists 25 such studies. The first, smaller group, covers the research that looks at the scope for incapacitation effects whilst in the schooling system to underlie the causal impact of education on crime. Six listed studies in the Table place a focus on this question (four listed in Panel B by [Jacob and Lefgren, 2003](#), and [Luallen, 2006](#), and [Cook and Kang, 2016](#), for the US and by [Berthelon and Kruger \(2011\)](#) for Chile, together with the studies by [Anderson, 2014](#), and [Bell et al., 2022](#) that already featured in Panel A). The second group covers a larger number – 21 – of published studies that look at various aspects of education, including quantity and quality, and assess their impact on crime. These are reviewed in sub-[Section 5.3](#) which follows below.

Incapacitation studies focus on the short run effect of education on crime ([Jacob and Lefgren, 2003](#); [Luallen, 2006](#)), whilst individuals' criminality is observed at ages while still in the schooling system. This contemporaneous impact of schooling contrasts with the focus on the long run crime-reducing effect of final educational attainment highlighted in Panel A of [Table 1](#). Causal evidence on incapacitation emerges from work that considers sources of

variation in school attendance coming about from idiosyncratic school closures for teacher training ([Jacob and Lefgren, 2003](#)), teacher strikes ([Luallen, 2006](#)), and school eligibility laws ([Cook and Kang, 2016](#)). These generate short run variation in the day-to-day propensity and possibility of young people committing a crime. A property crime-reducing effect of education arises in these studies, which reflects the short-run incapacitation effect of school on crime. In other words, the contemporaneous effect of compulsory schooling is to keep pupils busy and away from the street. In contrast, in the long run, the construction of human capital and enhanced labor market opportunities emerge as the lasting drivers of the reduction in property crime among more educated people; more educated individuals are able to find more and better jobs, and therefore have a lower crime propensity and a higher opportunity cost attached to criminal activity.

That both effects are relevant for the crime-reducing impact of education is further validated in recent evidence from the US by [Bell et al., 2022](#), which refines theoretical arguments further by introducing the concept of dynamic incapacitation. This work studies how the dropout age reforms alter crime-age profiles of individuals. Increases in the minimum school dropout age incapacitates youths in school at a critical age, beyond which they become less likely to commit crime if they had not engaged in crime onset at an earlier age (see the crime persistence by age estimates before and after dropout age reform as described in [Bell et al., 2022](#)).

Another common theme of the short-run studies of the contemporaneous effect of schooling on crime (e.g. [Jacob and Lefgren, 2003](#); [Luallen, 2006](#)) is that, in contrast to the reduction in property crime, there is an increase in violent crime. This pattern is typically explained by considering the nature of violent crime – they require the in-person interaction of individuals. Incapacitating youths in schools also puts them in close contact with each other, generating a concentration effect and thus increasing the likelihood of violent social interactions.

5.3 Schooling quantity and quality

The studies in Panel B that look at connections between an array of education measures and crime taken overall show that education matters for crime, and that the education impact works through various aspects of education quantity and quality. The former quantity impact – more education reduces crime – is the main finding from the research reviewed in [Sections 5.1 and 5.2](#). The school quality studies featured in Panel B refine this further by showing also that *better schooling* matters for crime. The (mostly US) studies that look at the effects of attending or being admitted to better performing schools all show this. [Cullen et al. \(2006\)](#) for Chicago, [Deming \(2011\)](#) and [McEachin et al. \(2020\)](#) for North Carolina, [Dobbie and Fryer \(2015\)](#) for New York City and [Grey-Lobe et al. \(2023\)](#) for Boston show that winning a lottery to be admitted

to better achieving schools improves crime outcomes. Panel B of the Table also lists studies showing that a number of other aspects of school quality matter for crime, including the social disadvantage structure of the school population and peers (e.g., [Gaviria and Raphael, 2001](#); [Billings et al., 2014](#); [Billings and Hoekstra, 2023](#)), school age entry ([Cook and Kang, 2016](#); [Depew and Eren, 2016](#)), school discipline ([Barrett et al., 2021](#)), grade retention ([Eren et al., 2022](#)) and testing regimes and school accountability ([Figlio, 2006](#)). Some of the research instead places a focus on cognitive and non-cognitive skills, including those acquired in early childhood education, offering evidence that they can help reduce crime ([García et al., 2023](#); [Heckman et al., 2006](#); [Jackson et al., 2020](#); [Johnson and Jackson, 2019](#)). At the time of writing, far fewer studies in this research area are based on evidence from outside the US, but with notable exceptions also showing crime reductions in Denmark by [Landersø et al. \(2017\)](#), in Finland by [Huttunen et al. \(2023\)](#) and in Trinidad and Tobago by [Beuermann et al. \(2023\)](#).

5.4 Crime impacts on education

A smaller, but also growing, area of crime and education research explores a causal relation in the opposite direction, from crime to education. Panel C of [Table 1](#) shows 15 studies. This research looks at crime impacts on subsequent educational attainment. An initial set of US evidence demonstrates that juvenile interactions with the justice system, and especially incarceration, can causally harm educational attainment, as shown in [Hjalmarsson \(2008\)](#) and [Aizer and Doyle \(2015\)](#). Following these studies, various different research designs have further established the impact of crime on subsequent education. Variations in judge leniency in sentencing (“judge fixed effects”) have been used in Colombia ([Arteaga, 2023](#)) and in the US ([Eren and Mocan, 2021](#)). Other US studies look at different dimensions of the crime impact on education, by studying inmate’s education levels as a function of prior youth sentencing ([Harlow, 2003](#)), prior exposure to disruptive peers ([Carrell and Hoekstra, 2010](#); [Carrell et al., 2018](#)), grade retention ([Eren et al., 2017](#)) and school suspensions ([Pope and Zuo, 2023](#)).

As discussed in more detail in the concluding section on future research directions, labor economists have in recent years shifted their focus from studying the human capital consequences of one’s own criminality to that of exposure to criminal environments and incidents. We discuss the papers published to date here because of their relevance to the impact of crime on education, but also feature them below in the conclusion about emerging areas in the economics of crime research area with an increased focus on victimization. Among the set of very recently published papers (one in each of 2017 and 2019, and all others in the 2020 s) about crime exposure, for instance, local violence incidents have also been shown to reduce education, with studies looking at proximity to police killings in Los Angeles ([Ang, 2021](#)), drug

related crimes in Mexico (Brown and Velásquez, 2017; Michaelsen and Salardi, 2020; Padilla-Romo and Peluffo, 2023), conflict fatalities in the West Bank (Brück et al., 2019) and homicides in São Paulo (Foureaux-Koppensteiner and Menezes, 2021).

5.5 Crime and education policies

The final group of crime and education papers, listed in Panel D of [Table 1](#), are a more heterogeneous group that we have collated together owing to their focus on the impact on crime of education-related policies and interventions. They are somewhat more diverse than the core crime and education papers shown in Panels A to C, but are included to complete the picture on the current literature in economics on crime and education. The 12 papers listed in Panel D show that a range of education policies in various settings, including both developed and developing countries, have scope to reduce crime.

6 Future directions

As can be seen from the survey thus far, the literature on crime and the labor market is vast, especially compared to its state at the time of [Freeman's \(1999\)](#) handbook chapter 25 years ago. Some areas, such as the education-crime link, are also clearly more developed than others. We conclude this chapter by highlighting two lines of research that are still in their infancy, but in our opinion, at the crime and labor market research frontier: the economics of victimization and gangs and organized crime.

6.1 Future direction 1: Victimization

Estimates of the cost of crime are regularly used to evaluate the relative costs and benefits of crime control policies, where the benefit is the cost of the crime prevented. Though direct costs (e.g., costs of policing, prison expenses, or the immediate health-related costs of a violent injury) are relatively easy to measure, it is not trivial to observe and/or measure indirect and intangible victim costs.⁹ This is especially the case given that these costs may spill-over onto other individuals besides the victim and may vary with victim characteristics (e.g., youths versus adults). [Anderson \(2021\)](#) estimates a total annual cost of crime in the United States between 4.7 and 5.8 trillion dollars – more than 20% of GDP. Yet, despite this staggering statistic and its importance for policy makers, little attention has historically been given to improving the measurement of these costs.

⁹ Examples of indirect victim costs include lost productivity, costs of precautionary behavior, long-term health consequences while intangible victim costs include, for instance, lower general well-being or the costs of fear. See [Dominguez and Raphael \(2015\)](#) for a discussion of the variety of approaches to estimate the costs of crime.

Economists have only recently branched out beyond Becker's workhorse economic model of *criminal* behavior to study the socioeconomic consequences of victimization and indirect crime exposure. This new literature overcomes two fundamental challenges – data and identification. Victimization registers (in selected countries) and geocoded crime incident data have only recently become available. In terms of identification, individuals select into neighborhoods, jobs, or schools with a non-random allocation of crime; consequently, many unobservable characteristics are related to both crime exposure or victimization and individual outcomes. Especially when studying the labor market effects, simultaneity can also be an issue. One's employment status, type and location of employment, or (as we will see) even peers on the job can impact one's risk of being victimized.

This section surveys the new economics of victimization literature, with an emphasis on the human capital and labor market effects. See [Bindler et al. \(2020\)](#) for a more comprehensive survey of the non-labor market effects of victimization.

Human Capital Costs of Youth Victimization. None of the many line-items in [Anderson's \(2021\)](#) cost of crime estimates include the costs of school-aged juveniles exposed to crime. Because a disruption in human capital accumulation when young can yield a lifetime of employment and earnings losses, a recent international literature considers the human capital costs of youth crime exposure. Much of the recent work uses detailed information on the timing and geography of crime to measure the nature of crime exposure and develop quasi-experimental designs.¹⁰ For example, using temporal variation in exposure to gang violence across cohorts in the same favela or slum of Rio de Janeiro, as well variation in how far the neighborhood is from violence, [Monteiro and Rocha \(2017\)](#) find that students in exposed schools perform worse on math test. This effect increases with conflict intensity, length, and proximity to the exam and can potentially be driven by absent teachers, principal turnover, and temporary school closings. [Foureaux Koppensteiner and Menezes \(2021\)](#) also find that indirect exposure to violence harms educational attainment by exploiting variation in time and space (i.e., whether a homicide is close to a child's school or school route) in Brazil.

The Brazilian context is perhaps extreme, with homicide rates more than 6 and 29 times that of the US and UK, respectively ([Foureaux Koppensteiner and Menezes, 2021](#)). Another extreme but less common event is mass shootings. [Bharadwaj et al. \(2021\)](#) study the 2011 killing of 69 people at a camp for about 600 Norwegian school-aged individuals. Using register data and a difference-

¹⁰In earlier work, [Groger \(1997\)](#) finds, using the High School and Beyond data, that moderate levels of 'local' violence reduce educational attainment, though school and neighborhood violence cannot be disentangled.

in-difference design, in which survivors were matched to similar children not at the camp or from the same school, they show that survivors have significantly worse education outcomes and future labor market outcomes. Higher mental healthcare take-up suggests a role of psychological effects.

A handful of studies also consider the educational impacts of US school shootings using data on the universe of school shootings from the Center for Homeland Defense and Security. Cabral et al. (2020) match 33 Texas school shootings in 1995 to 2016 to administrative data on public school students. Capitalizing on the arguably random timing of the shootings within schools, the authors compare within student education outcomes for the same students before and after a shooting to within student changes at matched control schools. They find that the shootings increased absenteeism and grade repetition and decreased high school graduation and college enrollment.¹¹

Ang (2021) highlights the impact on youth human capital outcomes of a significant but not typically discussed concern related to police use of force. To disentangle the impact of police use of force from selection effects (on where police presence and use of force is prevalent), Ang (2021) uses geocoded data to calculate the distance from each student home to each shooting. Comparing outcomes of those in the same neighborhood who lived close (within 0.5 miles) and slightly further (0.5–3 miles) from a killing, he finds spikes in absenteeism and reductions in both GPA and high school graduation. Moreover, the effects are driven by minority students and victims.

Though homicides, school shootings, mass killings, and police use of force are the most serious forms of violence (with large potential spill-over effects), they account for a relatively small share of crimes. Youths may be exposed to many other less extreme or less publicized crimes. One such category is domestic violence. Even in high income countries, lifetime rates of intimate partner violence are around 25% (Bhuller et al., 2024). But there is little knowledge on how children's exposure to household violence impacts their life outcomes. Bhuller et al. (2024) study this question by matching 22 years of Norwegian domestic violence police reports to identifiers for the victim's children. Using a regression discontinuity design based on the timing of test dates relative to the domestic violence incident, the authors find that domestic violence exposure decreases both exam scores and the chance of completing the first year of high school.¹²

Victimization and the Labor Market. Early papers on the labor market effects of victimization use surveys that ask about both earnings and

¹¹ Levine and McKnight (2020) find worse test scores and chronic absenteeism after the Sandy Hook shooting (28 fatalities), despite spending increases on instruction and support services. Deb and Gangaram (2024) use the 2003–2012 Behavioral Risk Factors Surveillance System surveys but cannot precisely measure shooting exposure.

¹² Exposure to children from households with domestic violence has been shown to result in negative peer effects (Carrell and Hoekstra, 2010; Carrell et al., 2018).

victimization. Given small samples and the rarity of victimization, these studies tended to result in imprecise estimates.¹³ Linked population-wide register data has rejuvenated this literature.

One way to measure victimization is via hospitalization data. [Ornstein \(2017\)](#) studies the effect of hospital-treated assaults using individual-level Swedish registers and a matching estimator that pairs a victim to an individual with comparable pre-assault characteristics. Ornstein finds a 25% (14%) decrease in earnings for female (male) assault victims. The use of hospital data to measure victimization has its limitations: Many crimes do not result in physical injury and, even for violent crimes, only assaults serious enough to require hospitalization can be studied.

[Bindler and Ketel \(2022\)](#) is the broadest study (in terms of offense types and outcomes) of the labor market effects of victimization. They link Dutch victimization registers of all offenses reported to the police to administrative records on employment, earnings, and unemployment insurance, disability, and welfare benefits. Using event study designs with individual fixed effects to control for time-invariant traits (which may be correlated with the risk of victimization), they find a significant reduction in earnings. The effects are immediate and large for violent offenses (robbery and assault) and smaller and more gradual for offenses like threat and burglary that do not result in injury. For most offenses, earnings do not return to pre-event levels within four years, even after health expenditures do. Finally, labor market effects tend to be worse for female victims.

[Bhuller et al. \(2024\)](#) consider the labor market consequences of domestic violence (in the same Norwegian study cited above about the human capital impacts). In a difference-in-difference framework that compares outcomes before and after the domestic violence report (using families who report domestic violence in the future as controls for those who report today), [Bhuller et al. \(2024\)](#) also find that victims' have higher disability insurance and lower earnings and employment.

[Adams-Prassl et al. \(2024\)](#) take advantage of matched Finnish victimization and employment registers to study the labor market consequences of violence against women at work. They identify more than 5000 violent incidents in which both the victim and perpetrator were working at the same plant or firm. A matched difference-in-differences design with individual fixed effects is again used to compare how employment outcomes of affected workers change before and after the workplace violence event relative to unaffected observationally identical workers. For both victims and perpetrators, employment drops immediately after an incident and does not completely recover in the next five years. An important piece of heterogeneity stands out. Employment

¹³ [Velamuri and Stillman \(2008\)](#) conducted one of the first such studies using the 'Household, Income, and Labor Dynamics in Australia' (HILDA).

effects are larger for perpetrators when both the victim and perpetrator are male, but about 60% larger for victims than perpetrators when the victim is female. Finally, there are firm level effects: female representation decreases (due to decreased hiring and retention) at male-managed firms with male-female violence.

Key Take-Aways from the New Economics of Victimization Literature. Three common themes emerge from the new literatures on the human capital and labor market effects of victimization and crime exposure.

- (i) Access to administrative victimization registers has given researchers power to study many dimensions of the human capital and labor market effects of victimization.
- (ii) There are significant human capital costs of indirect crime exposure (in a wide range of violent crime contexts), which are generally excluded from estimates of the cost of crime. Little is still known about the effects of property crime exposure.
- (iii) Victimization significantly reduces employment and earnings. Effects are persistent, increasing in offense severity, not limited to instances of physical injury, and often larger for females, even when not studying domestic violence.

6.2 Future direction 2: Gangs and organized crime

Criminal organizations play in the underworld the same role that large companies play in legitimate markets: they allow their members to pursue complex (but more remunerative) enterprises that would not be feasible for individuals or for smaller groups of associates. For the case of criminal organizations, some of these enterprises include drug-trafficking, racketeering, extortions, tax frauds, infiltrations in procurement contracts, and so on. These crimes arguably have major economic effects, and yet until very recently, criminal organizations have been largely ignored by the economic literature (Pinotti, 2015a). An important reason for this apparent neglect is that measurement issues are even more challenging when studying organized crime than when studying other types of criminal activities. In most countries, “organized crime” is not even defined in the penal code, meaning that members of criminal organizations cannot be prosecuted unless they commit some other type of crime, such as trafficking or violence. In Italy, which is home to some of the oldest and most powerful criminal organizations in the World, it took until 1982 for the national Parliament to punish the *Associazioni di tipo mafioso* (“mafia-type associations”), defined *ex Art. 416-bis* of the Italian Penal Code as organizations “whose members use the power of intimidation deriving from the bonds of membership, the state of subjugation and conspiracy of silence that it engenders to commit offences, to acquire direct or indirect control of economic activities, licenses, authorizations, public procurement contracts and services or to obtain unjust profits or advantages for themselves or others”. This definition

highlights another important source of measurement error: under-reporting of criminal organizations may be more severe when and where such organizations are more powerful and intimidating. As a consequence, most papers use proxies or “intention-to-treat” approaches leveraging events that are likely to affect the presence and strength of criminal organizations, such as targeted policies and enforcement operations, historical events leading to the birth or the move of criminal organizations, and so on.

[Pinotti \(2015b\)](#) estimates that the expansion of organized crime in two Italian regions (Puglia and Basilicata) in the 1970s lowered their GDP per capita by 15–20% over the following decades relative to a “synthetic control” of similar regions. In a mirror natural experiment, [Fenizia and Saggio \(2024\)](#) show that dismissing Italian city councils infiltrated by organized crime increased employment and the number of firms by 17% and 9%, respectively. [Sviatschi \(2022\)](#) looks at the long-term consequences of criminal organizations on human capital accumulation and (criminal) career choice. In particular, she shows that the expansion of coca production in Peru (driven by US anti-drug operations in Colombia) led to a 30% increase in child labor in coca-suitable areas, along with a 26% increase in dropout rate at the beginning of secondary school; in the long run, children grown in coca suitable areas are 30% more likely to be incarcerated when adult (age 18–30). In contrast with these papers, [Murphy and Rossi \(2020\)](#) find that Mexican drug cartels bring an improvement of the socio-economic conditions in the municipalities in which they are present, as measured by average salaries, quality of public services, and (lower) illiteracy rates. To establish causality, they leverage the geographical distribution of Chinese immigrants at the beginning of the 20th century, who used opium as a recreational drug and carried with them poppy seeds and knowledge of production and consumption.

Besides large, structured criminal organizations, smaller criminal groups such as street gangs may also have significant economic consequences. [Melnikov et al. \(2022\)](#) estimate the effect of living in gang-controlled areas that were established in El Salvador after thousands of Salvadoran members of street gangs in the US were deported back to their home country. People living just within the border of gang-controlled areas experience lower income and worse employment conditions than people living just outside the border. The authors attribute these effects to the mobility restrictions imposed by gangs and to the higher dropout rates observed in gang-controlled areas. [Brown et al. \(forthcoming\)](#) document another mechanism through which gangs in El Salvador affect economic activity, namely extortion. Using administrative data on 50,000 extortion payments from a leading wholesale distributor over the period 2012–2019, they show that extortion increases with collusion between gangs and that there is a significant passthrough to retail prices. Turning to other countries, [Dustmann et al. \(2023\)](#) show that refugees in Denmark that are randomly assigned to neighborhoods with significant gang presence are more likely to commit crime before age 19, which in turn affects their working career later in life.

Overall, this evidence suggests that organize crime imposes large economic costs on societies. At the same time, criminal organizations are often deeply intertwined with the official economy, as emphasized by the Art. 416-bis of the Italian Penal Code reported above (“to acquire direct or indirect control of economic activities, licenses, authorizations, public procurement contracts and services”). These connections with the official economy are essential for laundering and re-investing the proceedings from drug-trafficking and other illegal activities. In fact, the most common money laundering schemes, such as false invoicing and ‘smurfing’ (dividing a larger sum of money into multiple tiny transactions), require the collaboration of a large number of firms. In addition, criminal organizations may influence (or even create) firms for other purposes, such as acquiring social and political influence. Using classified data provided by the Financial Intelligence Unit of the Bank of Italy, [Arellano-Bover et al. \(2024\)](#) estimate that about 2% of all Italian firms have links with criminal organizations, and provide a taxonomy of such firms. Using similar data for the Italian region of Lombardy, [Bianchi et al. \(2022\)](#) show that connected firms report higher sales but lower profits than non-connected firms, and are more likely to file for bankruptcy. [Mirenda et al. \(2022\)](#) find similar results relying on a different measure of connection that is based on the family name and area of origin of the firm directors and owners. Both these papers interpret the findings as evidence that connected firms mainly serve the purpose of laundering and re-investing profits from illicit activities. Consistent with this interpretation, [Le Moglie and Sorrenti \(2022\)](#) find that after the credit crunch of 2007–2008, Italian provinces with a greater presence of organized crime displayed more firm creation than provinces with lower organized crime presence, and attribute the differential to the stable supply of (illicit) funds from criminal organizations. [Daniele et al. \(2024\)](#) look instead at the demand side, particularly by financially distressed firms. Using the same data as [Arellano-Bover et al. \(2024\)](#) along with credit score ratings on the universe of Italian firms, they show that being downgraded to a substandard credit risk, thus losing access to bank credit, increases the probability of being infiltrated by criminal organizations. Finally, [Calamunci and Drago \(2020\)](#) show that anti-mafia operations targeting connected firms have positive spillovers on other firms in the market, while [Slutzky and Zeume \(forthcoming\)](#) document increases in innovation activity and competition following such operations.

7 Conclusions

The economics of crime field has substantially advanced our understanding of criminal behavior and its relationship with labor markets. This chapter has documented the profound influence of economic conditions, such as wages, employment opportunities, and educational attainment, on crime rates. It also pays careful attention to the literature on the impact of criminal justice interactions and having a criminal record on human capital attainment and labor

market outcomes. Not only does this latter literature highlight the need for researchers to acknowledge identification challenges, like simultaneity bias, but also the fact that there is an oftentimes reinforcing cycle between crime and inequality. By applying rigorous methodological approaches, economists have been able to disentangle correlation from causation, overcoming this and other identification challenges, thereby providing more reliable insights into the effectiveness of various policy interventions. The shift from aggregated data to individual-level analyses and the incorporation of international perspectives have further enriched what has begun a sizable literature, allowing for more nuanced and context-specific policy recommendations.

Looking ahead, the integration of new data sources and innovative research designs will continue to push the boundaries of what we know about the economics of crime. Future research should focus on exploring the long-term impacts of social, education and criminal justice policies, the role of social networks in criminal behavior, and the effectiveness of rehabilitative versus punitive measures. Additionally, understanding the socio-economic costs of crime victimization and the influence of criminal organizations on labor markets remains a vital area for further investigation. By continuing to bridge the gap between theory and practice, economists can play a pivotal role in shaping policies that not only reduce crime but also enhance overall societal well-being.

References

- Adams-Prassl, A., Huttunen, K., Nix, E., Zhang, N., 2024. Violence against women at work. *Q. J. Econ.* 139, 937–991.
- Agan, A., Garin, A., Koustas, D., Mas, A., Yang, C., 2023. *The Impact of Criminal Records on Employment, Earnings, and Tax Filing. IRS SOI Working Paper*.
- Agan, A., Garin, A., Koustas, D., Mas, A., Yang, C., forthcoming. Labor market impacts of reducing felony convictions. *Am. Econ. Rev.: Insights*.
- Agan, A., Makowsky, M., 2023. The minimum wage, EITC, and criminal recidivism. *J. Hum. Resour.* 58, 1712–1751.
- Agan, A., Starr, S., 2017. The effect of criminal records on access to employment. *Am. Economic Rev.* 107, 560–564.
- Agan, A., Starr, S., 2018. Ban the box, criminal records, and racial discrimination: a field experiment. *Q. J. Econ.* 133, 91–235.
- Aizer, A., 2010. The gender wage gap and domestic violence. *Am. Economic Rev.* 100, 1847–1859.
- Aizer, A., Doyle, J., 2015. Juvenile incarceration, human capital, and future crime: evidence from randomly assigned judges. *Q. J. Econ.* 130, 759–803.
- Anders, J., Barr, A., Smith, A., 2023. The effect of early childhood education on adult criminality: evidence from the 1960s through 1990s. *Am. Econ. J.: Econ. Policy* 15, 37–69.
- Anderson, D., 2014. In school and out of trouble? The minimum dropout age and juvenile crime. *Rev. Econ. Stat.* 96, 318–331.
- Anderson, D., Sabia, J., 2018. Child-access-prevention laws, youths' gun carrying, and school shootings. *J. Law Econ.* 61, 489–524.

- Anderson, D., 2021. The aggregate cost of crime in the United States. *J. Law Econ.* 64, 857–885.
- Ang, D., 2021. The effects of police violence on inner-city students. *Q. J. Econ.* 136, 115–168.
- Apel, R., 2022. Sanctions, perceptions and crime. *Annu. Rev. Criminol.* 5, 205–227.
- Arellano-Bover, J., Simoni, M.D., Guiso, L., Macchiavello, R., Marchetti, D., Prem, M., 2024. CEPR working. Mafias Firms Paper 18982.
- Arteaga, C., 2023. Parental incarceration and children's educational attainment. *Rev. Econ. Stat.* 105, 1394–1410.
- Barrett, N., McEachin, A., Mills, J., Valant, J., 2021. Disparities and discrimination in student discipline by race and family income. *J. Hum. Resour.* 56, 711–748.
- Bartik, T., 1991. Who Benefits from State and Local Economic Development Policies? W.E Upjohn Institute for Economic Research.
- Beauchamp, A., Chan, S., 2014. The minimum wage and crime. *B. E. J. Econ. Anal. Policy* 14, 1213–1235.
- Becker, G., 1968. Crime and punishment: an economic approach. *J. Political Econ.* 76, 169–217.
- Becker, G., Mulligan, C., 1997. The endogenous determination of time preference. *Q. J. Econ.* 112, 729–758.
- Beattie, D., Kidd, M., Machin, S., Sarkar, D., 2017. Larrikin youth: crime and Queensland's earning and learning reform. *Labour Econ.* 52, 149–159.
- Bell, B., Bindler, A., Machin, S., 2018. Crime scars: recessions and the making of career criminals. *Rev. Econ. Stat.* 100, 392–404.
- Bell, B., Costa, R., Machin, S., 2022. Why does education reduce crime? *J. Political Econ.* 130, 732–765.
- Bennett, P., 2018. The heterogeneous effects of education on crime: evidence from danish administrative twin data. *Labour Econ.* 52, 160–177.
- Bennett, P., Ouazad, A., 2020. Job displacement, unemployment and crime: evidence from danish microdata and reforms. *J. Eur. Economic Assoc.* 18, 2182–2220.
- Berthelon, M., Kruger, D., 2011. Risky behavior among youth: incapacitation effects of school on adolescent motherhood and crime in Chile. *J. Public. Econ.* 95, 41–53.
- Beuermann, D., Jackson, C., Navarro-Sola, L., Pardo, F., 2023. What is a good school, and can parents tell? Evidence on the multidimensionality of school output. *Rev. Econ. Stud.* 90, 65–101.
- Bhalotra, S., Britto, D., Pinotti, P., Sampaio, B., forthcoming. Job displacement, unemployment benefits and domestic violence. *Rev. Econ. Studies.*
- Bharadwaj, P., Bhuller, M., Løken, K., Wentzel, M., 2021. Surviving a mass shooting. *J. Public. Econ.* 201, 104469.
- Bhuller, M., Dahl, G., Løken, K., Mogstad, M., 2020. Incarceration, recidivism and employment. *J. Political Econ.* 128, 1269–1324.
- Bhuller, M., Dahl, G., Løken, K., Mogstad, M., 2024. Domestic violence and the mental health and well-being of victims and their children. *J. Hum. Resour.* 59 (S), S152–S186.
- Bianchi, P., Marra, A., Masciandaro, D., Pecchiarini, N., 2022. Organized crime and firms' financial statements: evidence from criminal investigations in Italy. *Account. Rev.* 97, 77–106.
- Bignon, V., Caroli, E., Galbiati, R., 2018. Stealing to survive? Crime and income shocks in nineteenth century France. *Econ. J.* 127, 19–49.
- Billings, S., Deming, D., Rockoff, J., 2014. School segregation, educational attainment and crime: evidence from the end of busing in Charlotte-Mecklenburg. *Q. J. Econ.* 129, 435–476.
- Billings, S., Hoekstra, M., 2023. The effect of school and neighborhood peers on achievement, misbehavior, and adult crime. *J. Labor. Econ.* 41, 643–685.

- Bindler, A., Hjalmarsson, R., 2017. Prisons, recidivism and the age-crime profile. *Econ. Lett.* 152, 46–49.
- Bindler, A., Ketel, N., Hjalmarsson, R., 2020. Costs of victimisation, in Marcotte, D. (Section Editor). In: Zimmerman, K. (Ed.), *Handbook of Labour, Human Resources and Population*. Springer Nature.
- Bindler, A., Ketel, N., 2022. Scaring or scarring? Labour market effects of criminal victimization. *J. Labour Econ.* 40, 939–970.
- Blattman, C., Jamison, J., Sheridan, M., 2017. Reducing crime and violence: experimental evidence from cognitive behavioral therapy in Liberia. *Am. Econ. Rev.* 107, 1165–1206.
- Braakmann, N., Chevalier, A., Wilson, T., forthcoming. Expected returns to crime and crime location. *Am. Econ. J.: Appl. Econ.*
- Brilli, Y., Tonello, M., 2018. Does increasing compulsory education decrease or displace adolescent crime? New evidence from administrative and victimization data. *CESifo Economic Stud.* 64, 15–49.
- Britto, D., Pinotti, P., Sampaio, B., 2022. The effect of job loss and unemployment insurance on crime in Brazil. *Econometrica* 90, 1393–1423.
- Brown, R., Velásquez, A., 2017. The effect of violent crime on the human capital accumulation of young adults. *J. Dev. Econ.* 127, 1–12.
- Brown, Z., Montero, E., Schmidt-Padilla, C., Sviatschi, M., forthcoming. Market structure and extortion: evidence from 50,000 extortion payments. *Rev. Econ. Studies*.
- Brück, T., Di Maio, M., Miaari, S., 2019. Learning the hard way: the effect of violent conflict on student academic achievement. *J. Eur. Economic Assoc.* 17, 1502–1537.
- Buonanno, P., Leonida, L., 2006. Education and crime: evidence from Italian regions. *Appl. Econ. Lett.* 13, 709–713.
- Buonanno, P., Leonida, L., 2009. Non-market effects of education on crime: evidence from Italian regions. *Econ. Educ. Rev.* 28, 11–17.
- Bushway, S., Cabreros, I., Paige, J.W., Schwan, D., Wenger, J., 2022. Barred from employment: more than half of unemployed men in their 30s had a criminal history of arrest. *Sci. Adv.* 8, eabj6992.
- Bushway, S., Kalra, N., 2021. A policy review of employers' open access to conviction records. *Annu. Rev. Criminology* 4, 165–189.
- Cabral, M., Rossin-Slater, M., Bokyung Kim, M., Schnell, M., Schwandt, H., 2020. Trauma at school: the impacts of shootings on students' human capital and economic outcomes. NBER Working Paper 28311.
- Calamunci, F., Drago, F., 2020. The economic impact of organized crime infiltration in the legal economy: evidence from the judicial administration of organized crime firms. *Italian Econ. J.* 6, 275–297.
- Cano-Urbina, J., Lochner, L., 2019. The effect of education and school quality on female crime. *J. Hum. Cap.* 13, 188–235.
- Card, D., 1999. The causal effect of education on earnings. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of labor economics*, vol. 3 North Holland.
- Carrell, S., Hoekstra, M., 2010. Externalities in the classroom: how children exposed to domestic violence affect everyone's kids. *Am. Econ. J.: Appl. Econ.* 2, 211–228.
- Carrell, S., Hoekstra, M., Kuka, E., 2018. The long-run effects of disruptive peers. *Am. Econ. Rev.* 108, 3377–3415.
- Chalfin, A., McCrary, J., 2017. Criminal deterrence: a review of the literature. *J. Econ. Lit.* 55, 5–48.

- Cook, P., Kang, S., 2016. Birthdays, schooling, and crime: regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *Am. Econ. J.: Appl. Econ.* 8, 33–57.
- Corman, H., Mocan, N., 2005. Carrots, sticks, and broken windows. *J. Law Econ.* 48, 235–266.
- Craigie, T.-A., 2019. Ban the box, convictions, and public employment. *Econ. Inq.* 58, 425–445.
- Cullen, Z., Dobbie, W., Hoffman, M., 2022. Increasing the demand for workers with a criminal record. *Q. J. Econ.* 138, 103–150.
- Cullen, J., Jacob, B., Levitt, S., 2006. The effect of school choice on participants: evidence from randomized lotteries. *Econometrica* 74, 1191–1230.
- Cullen, J., Levitt, S., 1999. Crime, urban flight, and the consequences for cities. *Rev. Econ. Stat.* 81, 159–169.
- Daniele, G., Desimoni, M., Marchetti, D., Marcolongo, G., Pinotti, P., 2024. A loan you can't refuse: firms' financial vulnerability organized crime. Mimeo.
- Davis, J., Heller, S., 2020. Rethinking the benefits of youth employment programs: the heterogeneous effects of summer jobs. *Rev. Econ. Stat.* 102, 664–677.
- Dell, M., Feigenberg, B., Teshima, K., 2019. The violent consequences of trade-induced worker displacement in Mexico. *Am. Econ. Rev.: Insights* 1, 43–58.
- Deming, D., 2011. Better schools, less crime? *Q. J. Econ.* 126, 2063–2115.
- Depew, B., Eren, O., 2016. Born on the wrong day? School entry age and juvenile crime. *J. Urban. Econ.* 96, 73–90.
- Dix-Carneiro, R., Soares, R., Ulyssea, G., 2018. Economic shocks and crime: Evidence from the Brazilian trade liberalization. *Am. Econ. J.: Appl. Econ.* 10, 158–195.
- Dobbie, W., Fryer, R., 2015. The medium-term impacts of high-achieving charter schools. *J. Political Econ.* 123, 985–1037.
- Dobbie, W., Goldin, J., Yang, C., 2018. The effects of pre-trial detention on conviction, future crime, and employment: evidence from randomly assigned judges. *Am. Economic Rev.* 108, 201–240.
- Doleac, J., Hansen, B., 2020. The unintended consequences of ban the box: statistical discrimination and employment outcomes when criminal histories are hidden. *J. Labor. Econ.* 38, 321–374.
- Dominguez, P., Raphael, S., 2015. The role of the cost-of-crime literature in bridging the gap between social science research and policy making. *Criminol. Public Policy* 14, 589–632.
- Draca, M., Koutmeridis, T., Machin, S., 2019. The changing returns to crime: do criminals respond to prices? *Rev. Econ. Stud.* 86, 1228–1257.
- Draca, M., Machin, S., 2015. Crime and economic incentives. *Annu. Rev. Econ.* 7, 389–408.
- Dustmann, C., Mertz, M., Okatenko, A., 2023. Neighbourhood gangs, crime spillovers and teenage motherhood. *Econ. J.* 133, 1901–1936.
- Ehrlich, I., 1973. Participation in illegitimate activities: a theoretical and empirical investigation. *J. Political Econ.* 81, 521–563.
- Entorf, H., Spengler, H., 2000. Socio-economic and demographic factors of crime in Germany: evidence from panel data of the German states. *Int. Rev. Law Econ.* 20, 75–106.
- Eren, O., Depew, B., Barnes, S., 2017. Test-based promotion policies, dropping out, and juvenile crime. *J. Public. Econ.* 153, 9–31.
- Eren, O., Lovenheim, M., Mocan, N., 2022. The effect of grade retention on adult crime: evidence from a test-based promotion policy. *J. Labor. Econ.* 40, 361–395.
- Eren, O., Mocan, N., 2021. Juvenile punishment, high school graduation, and adult crime: evidence from idiosyncratic judge harshness. *Rev. Econ. Stat.* 103, 34–47.

- Eriksson, K., 2020. Education and incarceration in the Jim Crow South. *J. Hum. Resour.* 55, 43–75.
- Fenizia, A., Saggio, R., 2024. Organized crime and economic growth: evidence from municipalities infiltrated by the mafia. *Am. Econ. Rev.* 114, 2171–2200.
- Fernandez, J., Holman, T., Pepper, J., 2014. The impact of living-wage ordinances on urban crime. *Ind. Relat.* 53, 478–500.
- Figlio, D., 2006. Testing, crime and punishment. *J. Public. Econ.* 90, 837–851.
- Foged, M., Hasager, L., Peri, G., Nielsen, A., Bolvig, I., 2023. Intergenerational spillover effects of language training for refugees. *J. Public. Econ.* 220, 104840.
- Fone, Z., Sabia, J., Cesur, R., 2023. The unintended effects of minimum wage increases on crime. *J. Public. Econ.* 219, 104780.
- Fougère, D., Kramarz, F., Pouget, J., 2009. Youth unemployment and crime in France. *J. Eur. Economic Assoc.* 7, 909–938.
- Foureaux-Koppensteiner, M., Menezes, L., 2021. Violence and human capital investments. *J. Labor. Econ.* 39, 787–823.
- Frandsen, B., Lefgren, L., Leslie, E., 2023. Judging judge fixed effects: testing the identifying assumptions in judge fixed-effects designs. *Am. Econ. Rev.* 113, 253–277.
- Freeman, R., 1991. Crime and the Employment of Disadvantaged Youths. NBER Working Paper 3875.
- Freeman, R., 1999. The economics of crime. In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics* 3 North Holland.
- Freeman, R., Rodgers III, W., 1999. Area economic conditions and the labor market outcomes of young men in the 1990s expansion. In: Cherry, R., Rodgers, W. (Eds.), *Prosperity for All?? The Economic Boom and African Americans*. Russell Sage Foundation, NY.
- García, J., Heckman, J., Ronda, V., 2023. The lasting effects of early-childhood education on promoting the skills and social mobility of disadvantaged African Americans and their children. *J. Political Econ.* 131, 1477–1506.
- Garin, A., Kousta, D., McPherson, C., Norris, S., Pecenco, M., Rose, E., et al., 2023. The impact of incarceration on employment and earnings. Becker Friedman Institute Working Paper 2023–108.
- Gaviria, A., Raphael, S., 2001. School-based peer effects and juvenile behavior. *Rev. Econ. Stat.* 83, 257–268.
- Gelber, A., Isen, A., Kessler, J., 2016. The effects of youth employment: evidence from New York City lotteries. *Q. J. Econ.* 131, 423–460.
- Gilpin, G., Pennig, L., 2015. Compulsory schooling laws and school crime. *Appl. Econ.* 47, 4056–4073.
- Gould, E., Weinberg, B., Mustard, D., 2002. Crime rates and local labor market opportunities in the United States: 1979–1997. *Rev. Econ. Stat.* 84, 45–61.
- Gray-Lobe, G., Pathak, P., Walters, C., 2023. The long-term effects of universal preschool in Boston. *Q. J. Econ.* 138, 363–411.
- Grogger, J., 1995. The effect of arrests on the employment and earnings of young men. *Q. J. Econ.* 110, 51–72.
- Grogger, J., 1997. Local violence and educational attainment. *J. Hum. Resour.* 32, 659–682.
- Grogger, J., 1998. Market wages and youth crime. *J. Labor. Econ.* 16, 756–791.
- Gulesci, S., Puente-Beccar, M., Ubfal, D., 2021. Can youth empowerment programs reduce violence against girls during the COVID-19 pandemic? *J. Dev. Econ.* 153, 102716.
- Hansen, K., Machin, S., 2002. Spatial crime patterns and the introduction of the UK minimum wage. *Oxf. Bull. Econ. Stat.* 64, 677–697.

- Harlow, C., 2003. Education and Correctional Populations, Bureau of Justice Statistics Special Report, NCJ 195670. U.S. Department of Justice.
- Heckman, J., Stixrud, J., Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *J. Labor. Econ.* 24, 411–482.
- Heller, S., 2014. Summer jobs reduce violence among disadvantaged youth. *Science* 346, 1219–1223.
- Heller, S., 2022. When scale and replication work: learning from summer youth employment experiments. *J. Public. Econ.* 209, 104617.
- Heller, S., Kessler, J., forthcoming. How to allocate slots: the market design of summer youth employment programs. In Kominers, S., Teytelboym, A. (Eds.), *Fair by Design: Economic Design Approaches to Inequality*. Oxford University Press.
- Heller, S., Shah, A., Guryan, J., Ludwig, J., Mullainathan, S., Pollack, H., 2017. Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *Q. J. Econ.* 132, 1–54.
- Hjalmarsson, R., 2008. Criminal justice involvement and high school completion. *J. Urban. Econ.* 63, 613–630.
- Hjalmarsson, R., 2022. Social policies as crime control. SNS Research Report.
- Hjalmarsson, R., Holmlund, H., Lindquist, M., 2015. The effect of education on criminal convictions and incarceration: causal evidence from micro-data. *Econ. J.* 125, 1290–1326.
- Hjalmarsson, R., Lindquist, M., 2022. The health effects of prison. *Am. Econ. J.: Appl. Econ.* 14, 234–270.
- Huttunen, K., Pekkarinen, T., Uusitalo, R., Virtanen, H., 2023. Lost boys? Secondary education and crime. *J. Public. Econ.* 218, 104804.
- Ioannides, Y., Loury, L.D., 2004. Job information networks, neighborhood effects, and inequality. *J. Econ. Lit.* 42, 1056–1093.
- Jacob, B., Lefgren, L., 2003. Are idle hands the devil's workshop? Incapacitation, concentration and juvenile crime. *Am. Econ. Rev.* 93, 1560–1577.
- Jackson, C., Porter, S., Easton, J., Blanchard, A., Kiguel, S., 2020. School effects on socio-emotional development, school-based arrests, and educational attainment. *Am. Econ. Rev.: Insights* 2, 491–508.
- Jackson, O., Zhao, B., 2016. The effect of changing employers' access to criminal histories on ex-offenders' labor market outcomes: evidence from the 2010–2012 Massachusetts CORI reform. Federal Reserve Bank of Boston Research Paper Series 16–30.
- Johnson, R., Jackson, C., 2019. Reducing inequality through dynamic complementarity: evidence from head start and public school spending. *Am. Econ. J.: Econ. Policy* 11, 310–349.
- Kessler, J., Tahamont, S., Gelber, A., Isen, A., 2022. The effects of youth employment on crime: evidence from New York City lotteries. *J. Policy Anal. Manag.* 41, 710–730.
- Khanna, G., Medina, C., Nyshadham, A., Posso, C., Tamayo, J., 2021. Job loss, credit and crime in Colombia. *Am. Econ. Rev.: Insights* 3, 97–114.
- Kirchmaier, T., Machin, S., Sandi, M., Witt, R., 2020. Prices, policing and policy: their role in crime booms and busts. *J. Eur. Econ. Assoc.* 18, 1040–1077.
- Kling, J., 2006. Incarceration length, employment, and earnings. *Am. Econ. Rev.* 96, 863–876.
- LaLonde, R., 1986. Evaluating the econometric evaluations of training programs with experimental data. *Am. Econ. Rev.* 76, 604–620.
- Landersø, R., 2015. Does incarceration length affect labor market outcomes? *J. Law Econ.* 58, 205–234.
- Landersø, R., Nielsen, H., Simonsen, M., 2017. School starting age and the crime-age profile. *Econ. J.* 127, 1096–1118.

- Le Moglie, M., Sorrenti, G., 2022. Revealing Mafia Inc.? Financial crisis, organized crime and the birth of new enterprises. *Rev. Econ. Stat.* 104, 142–156.
- Levine, P., McKnight, R., 2020. Exposure to a school shooting and subsequent well-being. NBER Working Paper 28307.
- Levitt, S., Venkatesh, S., 2000. An economic analysis of a drug-selling gang's finances. *Q. J. Econ.* 115, 755–789.
- Lochner, L., 2004. Education, work and crime: a human capital approach. *Int. Econ. Rev.* 45, 811–843.
- Lochner, L., Moretti, E., 2004. The effect of education on crime: evidence from prison inmates, arrests, and self reports. *Am. Econ. Rev.* 94, 155–189.
- Loeffler, C., 2013. Does imprisonment alter the life course? Evidence on crime and employment from a natural experiment. *Criminology* 51, 137–166.
- Luallen, J., 2006. School's out forever: A study of juvenile crime, at-risk youths and teacher strikes. *J. Urban. Econ.* 59, 75–103.
- Machin, S., Marie, O., Vujić, S., 2011. The crime reducing effect of education. *Econ. J.* 121, 463–484.
- Machin, S., Meghir, C., 2004. Crime and economic incentives. *J. Hum. Resour.* 39, 958–979.
- Machin, S., Sandi, M., forthcoming. Crime and education. *Ann. Rev. Econ.*
- Macmillan, R., Gartner, R., 1999. When she brings home the bacon: labor-force participation and the risk of spousal violence against women. *J. Marriage Family* 947–958.
- McEachin, A., Lauen, D.L., Fuller, S.C., Perera, R., 2020. Social returns to private choice? Effects of charter schools on behavioral outcomes, arrests and civic participation. *Econ. Educ. Rev.* 76, 101983.
- Melnikov, N., Schmidt-Padilla, C., Sviatschi, M., 2022. Gangs, labor mobility and development. Mimeo.
- Michaelsen, M., Salardi, P., 2020. Violence, psychological stress and educational performance during the “war on drugs” in Mexico. *J. Dev. Econ.* 143, 102387.
- Mirenda, L., Mocetti, S., Rizzica, L., 2022. The economic effects of mafia: firm level evidence. *Am. Econ. Rev.* 112, 2748–2773.
- Modestino, A., 2019. How do summer youth employment programs improve criminal justice outcomes, and for whom? *J. Policy Anal. Manag.* 38, 600–628.
- Monteiro, J., Rocha, R., 2017. Drug battles and school achievement: evidence from Rio de Janeiro's Favelas. *Rev. Econ. Stat.* 99, 213–228.
- Mueller-Smith, M., 2015. The criminal and labor market impacts of incarceration. Mimeo.
- Mueller-Smith, M., Schepel, K., 2021. Diversion in the criminal justice system. *Rev. Econ. Stud.* 88, 883–936.
- Murphy, T., Rossi, M., 2020. Following the poppy trail: origins and consequences of Mexican drug cartels. *J. Dev. Econ.* 143, 102433.
- Nagin, D., 2013. Deterrence: A review of the evidence by a criminologist for economists. *Annu. Rev. Econ.* 5, 83–105.
- Norris, S., Pecenco, M., Weaver, J., 2021. The effects of parental and sibling incarceration: evidence from Ohio. *Am. Economic Rev.* 111, 2926–2963.
- Ornstein, P., 2017. The price of violence: consequences of violent crime in Sweden. IFAU Working Pap. 2017 22.
- Öster, A., Agell, J., 2007. Crime and unemployment in turbulent times. *J. Eur. Econ. Assoc.* 5, 752–775.
- Owens, E., 2017. Testing the school-to-prison pipeline. *J. Policy Anal. Manag.* 36, 11–37.
- Owens, E., Ba, B., 2021. The economics of policing and public safety. *J. Econ. Perspect.* 35, 3–28.

- Padilla-Romo, M., Peluffo, C., 2023. Violence-induced migration and peer effects in academic performance. *J. Public. Econ.* 217, 104778.
- Pager, D., 2003. The mark of a criminal record. *Am. J. Sociol.* 108, 937–975.
- Pinotti, P., 2015a. The causes and consequences of organised crime: preliminary evidence across countries. *Econ. J. Features* 125, F158–F174.
- Pinotti, P., 2015b. The economic costs of organised crime: evidence from Southern Italy. *Econ. J. Features* 125, F203–F232.
- Pissarides, C., 1992. Loss of skill during unemployment and the persistence of employment shocks. *Q. J. Econ.* 107, 1371–1391.
- Pope, N., Zuo, G., 2023. Suspending suspensions: the education production consequences of school suspension policies. *Econ. J.* 133, 2025–2054.
- Prescott, J., Starr, S., 2020. Expungement of criminal convictions: an empirical study. *Harv. Law Rev.* 133, 2460–2555.
- Raphael, S., Winter-Ebmer, R., 2001. Identifying the effect of unemployment on crime. *J. Law Econ.* 44, 259–283.
- Raphael, S., 2014. The New Scarlet Letter? Negotiating the U.S. Labor Market with a Criminal Record. W.E. Upjohn Institute for Employment Research., Kalamazoo, MI.
- Reilly, B., Witt, R., 2008. Domestic burglaries and the real price of audio-visual goods: some time series evidence for Britain. *Econ. Lett.* 100, 96–100.
- Rees, D., Sabia, J., Kumpas, G., 2022. Anti-bullying laws and suicidal behaviors among teenagers. *J. Policy Anal. Manag.* 41, 787–823.
- Rege, M., Skardhamar, T., Telle, K., Votruba, M., 2019. Job displacement and crime: evidence from Norwegian register data. *Labour Econ.* 61, 101761.
- Rose, E., 2021. Does banning the box help ex-offenders get jobs? Evaluating the effects of a prominent example. *J. Labor. Econ.* 39, 79–113.
- Rose, E., Shem-Tov, Y., 2021. How does incarceration affect reoffending? Estimating the dose-response function. *J. Political Econ.* 129, 3302–3356.
- Sabates, R., Feinstein, L., 2008. Effects of government initiatives on youth crime. *Oxf. Econ. Pap.* 60, 462–483.
- Schmidt, P., Witte, A., 2013. *An Economic Analysis of Crime and Justice: Theory, Methods, and Applications*. Elsevier.
- Selbin, J., McCrary, J., Epstein, J., 2018. Unmarked: criminal record clearing and employment outcomes. *J. Crim. Law Criminol.* 108 article 1.
- Slutzky, P., Zeume, S., forthcoming. Organized crime and firms: evidence from antimafia enforcement actions. *Manage. Sci.*
- Smith, S., Broege, N., 2020. Searching for work with a criminal record. *Soc. Probl.* 67, 208–232.
- Sorensen, L., Avila-Acosta, M., Engberg, J., Bushway, S., 2023. The thin blue line in schools: new evidence on school-based policing across the U.S. *J. Policy Anal. Manag.* 42, 941–970.
- Stevenson, M., 2023. Cause, effect, and the structure of the social world. *Boston Univ. Law Rev.* 103, 2001–2047.
- Sviatschi, M., 2022. Making a narco: childhood exposure to illegal labor markets and criminal life paths. *Econometrica* 90, 1835–1878.
- Topa, G., 2011. Labor markets and referrals. In: Benhabib, J., Bisin, A., Jackson, M. (Eds.), *Handbook of Social Economics*. North-Holland.
- Uggen, C., Vuolo, M., Lageson, S., Ruhland, E., Whitham, H., 2014. The edge of stigma: an experimental audit of the effects of low-level criminal records on employment. *Criminology* 52, 627–654.

- Velamuri, M., Stillman, S., 2008. The impact of crime victimisation on individual wellbeing: evidence from Australia. In: Morrison, P. (Ed.), Proceedings of the Joint LEW13/ALMRW Conference. Victoria University of Wellington.
- Von Wachter, T., 2020. The persistent effects of initial labor market conditions for young adults and their sources. *J. Econ. Perspect.* 34, 168–194.
- Waldfogel, J., 1994. The effect of criminal conviction on income and the trust reposed in the workmen. *J. Hum. Resour.* 29, 62–81.
- Weisburst, E., 2019. Patrolling public schools: the impact of funding for school police on student discipline and long-term education outcomes. *J. Policy Anal. Manag.* 38, 338–365.
- Western, B., Kling, J., Weiman, D., 2001. The labor market consequences of incarceration. *Crime, Delinquency* 47, 410–427.
- Witte, A., 1980. Estimating the economic model of crime with individual data. *Q. J. Econ.* 94, 57–84.
- Witte, A., Tauchen, H., 2000. Work and crime: an exploration using panel data. In: Fielding, N., Clarke, A., Witt, R. (Eds.), *The Economic Dimensions of Crime*. Palgrave Macmillan UK.

This page intentionally left blank

Chapter 10

Monopsony power in the labor market

José Azar^a, and Ioana Marinescu^{b,*}

^aCentre for Economic Policy Research, London, United Kingdom, ^bUniversity of Pennsylvania, PA, United States

*Corresponding author. e-mail address: ioana.marinescu@gmail.com

Chapter Outline

1 Introduction	761	3.4 Reduced-form approach based on workers' outside options	804
2 New quantitative models of monopsony power	764	3.5 Calibration and simulation	804
2.1 What is monopsony power?	764	3.6 Structural estimation	805
2.2 Oligopsony	767	3.7 Production function approach	806
2.3 Differentiated jobs	769	3.8 Summary and discussion of markdown estimates	810
2.4 Search frictions	781		
2.5 Discussion	791		
3 Empirically measuring monopsony power	792	4 Policy and monopsony power	812
3.1 Definition of the markdown in the empirical literature	792	4.1 Merger control	812
3.2 Elasticity of labor supply	793	4.2 Non-competition agreements	815
3.3 Labor market concentration	799	4.3 Minimum wage	818
		5 Conclusion	820
		References	822

1 Introduction

While the labor market has traditionally been studied under the assumption of perfect competition, a recent theoretical and empirical literature, ¹ – reviewed in Sokolova and Sorensen (2021); Manning (2021); Card (2022) – has challenged this assumption by documenting the significant monopsony power that firms wield. For the purposes of this chapter, we focus on the markdown as the key measure of monopsony power, i.e. firms' market power as buyers of labor services.² The markdown is the gap between workers' marginal

¹ Manning (2011) reviewed the older literature relevant to imperfect competition in the labor market in a previous volume of this Handbook.

² As we discuss below, “competition” and “market power” have complex meanings that cannot be fully captured by the markdown.

revenue product ($MRPL$) and the wage³ w (scaled by the wage), i.e. $(MRPL - w)/w = MRPL/w - 1$.⁴ The markdown measures the percent wage increase that would occur, all other things equal, if employers' monopsony power were eliminated so that the wage is equal to the marginal revenue product of labor, $w = MRPL$.

This chapter summarizes and explains the insights from this burgeoning monopsony literature by focusing on three pivotal theoretical frameworks: oligopsony, job differentiation, and search and matching. These frameworks not only deepen our understanding of the sources of monopsony power but also have implications for public policy, particularly in the areas of merger control, non-competition agreements, and minimum wage regulation.

In the oligopsony framework, we explain how to extend the basic “isolated firm” monopsony model to introduce strategic interactions between firms. Firms compete by selecting employment levels, taking the market-level labor supply curve as given. In this framework, the average markdown in a labor market increases with concentration, and decreases with the labor supply elasticity to the market.

The job differentiation framework posits that monopsony power can arise even in markets with atomistic firms, due to heterogeneous worker preferences over jobs that provide both wages and amenities. The greater the degree of job differentiation, the greater the markdown. This framework is instrumental in understanding how firms can pay wages below the marginal revenue product of labor without losing all of their workers. Workers are willing to accept less than competitive wages when the job provides higher utility – a better bundle of wages and amenities – than the next best alternative. Job differentiation can serve as a microfoundation for the finite labor supply elasticity to the market in the oligopsony model. New theoretical frameworks have added large firms, allowing for the integration of oligopsony and job differentiation. In these hybrid models, both job differentiation and labor market concentration increase the markdown.

The search and matching framework introduces search frictions. Because of these frictions, workers cannot instantaneously meet all available jobs, and thus cannot fully benefit from competition between employers. These frictions enable even atomistic firms to offer wages below what would prevail in a perfectly competitive market. The recent integration of large firms into these

³ In a broader sense, the wage can be understood as full compensation, including costly amenities provided by the firm, as we will discuss.

⁴ More generally, one may define monopsony power as the ability of employers to pay workers less than the *competitive* wage, while recognizing that the competitive wage benchmark may be different from the marginal revenue product, depending on model assumptions. For example, if the solution to the firms' profit maximization problem is not interior, then their level of employment may not satisfy the first-order condition that the wage equals the value of the marginal product of labor.

models offers fresh perspectives on wage-setting dynamics and introduces labor market concentration as a negative determinant of wages. In search and matching models, the markdown typically decreases with the worker's non-employment income (the outside option), and varies with concentration and the nature and degree of search frictions.

Each of these theoretical frameworks has been used to quantify the role of monopsony power in the labor market using constructs like labor market concentration, the labor supply elasticity, and the production function. We summarize the results of this empirical literature and the econometric challenges that it seeks to address. Starting with concentration, the empirical research often focuses on wages as the outcome instead of markdowns. Yet, wages need not decrease with concentration, even if theory predicts that the *markdown* increases with concentration. Wages might increase with concentration if the dominant firm's productivity significantly improves, allowing it to expand its market share. Thus, concentration is a useful proxy for market power (i.e., the markdown), but it may not be as informative about the wage, and ultimately worker *welfare*. Nevertheless, the empirical literature typically finds that higher concentration is associated with lower wages. We then summarize results on the labor supply elasticity and the markdown (see Tables 4, 5 and 6). The literature typically finds a labor supply elasticity to the firm between 2 and 6, a labor supply elasticity to the market between 0.5 and 5, and a markdown between 15 % and 50 %. The markdown estimates thus imply that, all other things equal, wages would increase between 15 % and 50 % if monopsony power could be eliminated. Overall, the literature paints a consistent picture indicating that firms have significant monopsony power in the labor market.

Monopsony power affects our understanding and interpretation of variation in wages across workers, firms, and over time. In general, we should be careful not to interpret the wage as a direct measure of the marginal productivity of labor – a common assumption in economic analysis. Variation in wages may be driven by variation in the markdown as well as variation in productivity. Future research should consider whether the role of monopsony power is significant to understand the economic question at hand.

The models of monopsony power we discuss in this chapter can fundamentally affect our understanding of policy effects in various contexts. We focus on three areas where there exists a significant body of empirical research. First, the oligopsony model elucidates the mechanisms that drive the observed wage suppression in mergers that increase labor market concentration. Second, the search and matching model can inform the policy debate around non-competition agreements, and helps explain the observed negative effects on wages and labor mobility. Lastly, the oligopsony and job differentiation models offer a lens to examine the disparate impacts of minimum wage policies across labor markets with differing levels of monopsony power. Empirical studies of minimum wage effects suggest that markdowns can buffer the labor market

against the negative employment effects predicted by the perfectly competitive model; the minimum wage can even *increase* employment in the least competitive labor markets. Monopsony power thus helps explain why many studies fail to find a negative effect of the minimum wage on employment.

The chapter proceeds as follows. In [Section 2](#), we describe and compare the three theoretical frameworks that help explain monopsony power: oligopsony, job differentiation, and search and matching. We provide formulas for the markdown ([Table 1](#)) and the firm-level and market-level labor supply elasticities ([Table 2](#)) in these different models. In [Section 3](#), we discuss the empirical challenges in estimating monopsony power, and summarize the results from various strands of the literature, including the labor supply elasticity, labor market concentration, structural estimation, and the production function approach. In [Section 4](#), we discuss the role of monopsony power in policy analysis, focusing on areas where the literature is most developed and most well integrated with theory: merger control, non-competition agreements, and the minimum wage.

2 New quantitative models of monopsony power

2.1 What is monopsony power?

The markdown – the difference between workers’ marginal revenue product and the wage – is in our view the key measure of monopsony power. A primary goal of this theory section is to derive formulas for the markdown and its determinants across models, so that we can better understand what drives market power in different models.

Before we get started with our theoretical discussion, it is important to recognize that “market power” and “competition” have a rich set of meanings in policy and business contexts. For example, in 2023, the White House Office of Budget and Management ([OIRA, OMB, 2023](#)) defined these terms as follows:

Competition is, among other things, the process by which individuals or firms vie to win customers’ business for goods or services, to purchase suppliers’ goods or services, or to hire workers for their labor services. Competitive markets are characterized by (but not exclusively characterized by) the presence of independent and rival buyers and sellers such that each market participant has many potential options to turn to. Encouraging competitive markets is an important policy goal. Competitive markets are associated with lower prices for consumers, higher wages for workers, more innovative products and services, more business formation, and greater resilience to unexpected events. When markets are less competitive, we say that certain market participants have market power. Firms with market power have the ability to change their behavior so as to increase their own profits or advance their other interests at the expense of others.

TABLE 1 Markdowns across models (see [Section 2](#) for more details).

Class of monopsony model	Average markdown across firms	Markdown at firm level	Notes
Oligopsony	$\frac{HHI}{\eta}$	$\frac{s_j}{\eta}$	η is the labor supply elasticity to the market. The employment share of firm j is s_j , and the HHI is based on employment shares.
Differentiated jobs			
Monopsonistic competition	$\frac{1}{\theta}$	$\frac{1}{\theta}$	θ goes to infinity means jobs are perfect substitutes
Oligopsony	$\frac{1}{\theta} + \left(\frac{1}{\varphi} - \frac{1}{\theta}\right)HHI$	$\frac{1}{\theta} + \bar{s}_j \left(\frac{1}{\varphi} - \frac{1}{\theta}\right)$	φ is the aggregate elasticity of labor supply. The payroll share of firm j is \bar{s}_j , and the HHI is based on payroll shares.
Discrete choice (logit)	$\approx \frac{1}{\eta} \left[\frac{J(J^2 - 3J + 1)}{(J-1)^3} + \frac{J^3}{(J-1)^3} HHI \right]$	$\frac{1}{\eta(1-s_j)}$	η is the coefficient on log wage in the utility, thus related to wage elasticity of labor supply. The number of firms in the labor market J also affects the firm-level wage elasticity of labor supply.
Search and matching			
Burdett-Mortensen	$\frac{A - b}{b + \frac{\lambda}{\delta} A}$		A and b are the productivity of an employed worker, and the income of a non-employed worker, respectively. λ/δ is the ratio of the job-finding rate and the separation rate.

Continued

TABLE 1 Markdowns across models (see [Section 2](#) for more details).—Cont'd

Class of monopsony model	Average markdown across firms	Markdown at firm level	Notes
Rudanko	$\gamma \frac{A - b}{\gamma b + (1 - \gamma)A}$ (if firms homogeneous)	$\gamma \frac{A - b}{\gamma b + (1 - \gamma)A}$	γ is a function of the matching function elasticity, and takes different values in the constrained and unconstrained case.

TABLE 2 Elasticity of labor supply across models (see [Section 2](#) for more details).

Class of monopsony model	Firm-level elasticity	Market-level elasticity	Notes
Oligopsony	$\frac{\eta}{s_j}$	η	η is the labor supply elasticity to the market. The employment share of firm j is s_j .
Differentiated jobs			
Monopsonistic competition	θ	φ	θ goes to infinity means jobs are perfect substitutes. Firms do not take into account φ when setting wages.
Oligopsony	$\frac{\theta\varphi}{\varphi + \tilde{s}_j(\theta - \varphi)}$	φ	θ goes to infinity means jobs a perfect substitutes. The payroll share of firm j is \tilde{s}_j .
Discrete choice (logit)	$\eta(1 - s_j)$	0	η is the coefficient on log wage in the utility.

If we adopt a broad definition of competition and market power (such as the one above), the markdown does not capture all the nuances of market power. The markdown does not directly measure how the competitive process unfolds in a particular labor market. Instead, we can think of the markdown as a measure of the outcome of this competitive process, which gives us information about firms' ability and willingness to reduce wages below the marginal

revenue product of labor. Our theoretical discussion helps make more precise what mechanisms drive the markdown in specific models. Ultimately, the markdown is a useful summary statistic for market power in the labor market at a given point in time.

2.2 Oligopsony

Oligopsony is the classical theoretical framework where firm size matters, and explains how wage determination is affected by monopsony power. Labor market concentration is a key determinant of the markdown in this class of models.

In the textbook perfectly competitive labor market, infinitely many firms compete for workers, and wages are equal to the worker's marginal revenue product. These firms are assumed to be atomistic, so that they cannot affect market outcomes. At the other extreme of the perfectly competitive model is the literal case of monopsony, in which there is a single employer in a labor market. The simplest model of monopsony, the “isolated firm” model, goes back to [Robinson, \(1933, 1966\)](#). In contrast to the perfectly competitive labor market model where the individual firm faces a constant labor supply curve, this model considers an upward sloping labor supply curve. Oligopsony models extend the basic “isolated firm” monopsony model to allow for multiple large firms operating in a labor market. The term “large” indicates that these firms are significant in size relative to the market; they are not atomistic or infinitesimal. Consequently, these large firms influence market-level outcomes, such as employment levels. This scenario introduces strategic interactions among firms in the labor market. Coupled with the upward-sloping market-level labor supply curve, these strategic interactions form the foundational elements of the oligopsony model. These components differentiate the oligopsony model from a model of perfect competition. The oligopsony model is sometimes dubbed “classical monopsony” in the literature.

In the simplest version of the oligopsony model,⁵ several large firms compete with each other by selecting the level of employment, given their expectations of competing firms' employment levels. The first-order condition for a firm n in the oligopsonistic labor market is:

$$R'(L_n) = w(L) + w'(L)L_n, \quad (1)$$

where $R'(L_n)$ is the marginal revenue product of labor, L is market-level employment, $w(L)$ is the inverse labor supply curve to the market, and L_n is the employment of firm n . In the perfectly competitive model, the employment level of the firm would not influence the equilibrium wage in the market, so the term $w'(L)L_n$ would be set to zero. As a result, we would have a wage that is

⁵ For example, [Azar et al. \(2023\)](#) use this oligopsony model and derive its wage and employment effects.

equal to the marginal revenue product of labor, a well-known result for perfectly competitive markets. Therefore, maximum competition in this model results in wages equal to marginal productivity.

It's worth noting here that the firm is assumed to pay the same wage to all its workers, and this assumption leads to the term $w'(L)L_n$ in [equation \(1\)](#). This equality constraint is a fundamental assumption that leads to lower employment: because the firm has to pay incumbent workers more in order to hire the marginal worker, it chooses a lower level of employment in equilibrium. A firm's markdown (the difference between the marginal revenue product and the wage as a share of the wage) is given by:

$$\frac{R'(L_n) - w}{w} = \frac{s_n}{\eta}, \quad (2)$$

where $R'(L_n)$ is the revenue marginal product of labor of firm n , w is the equilibrium wage in the labor market, s_n is the employment share of firm n , and η is the market-level elasticity of labor supply. If the firm is a literal monopsonist, the firm's share is equal to one and the market-level elasticity of labor supply is also the firm's elasticity of labor supply. In an oligopsony labor market, each firm's labor supply elasticity is larger than the monopsonist's, since each firm's share is less than one, and firms compete with each other.

The average markdown or Pigou's rate of exploitation is equal to the market-level inverse elasticity of labor supply times the employment Herfindahl-Hirschman Index (HHI) in the labor market ([Boal and Ransom, 1997](#)):

$$\sum_{n=1}^N s_n \frac{R'(L_n) - w}{w} = \frac{\sum_{n=1}^N s_n^2}{\eta} = \frac{HHI}{\eta}, \quad (3)$$

where the HHI is defined as the sum of the squares of employment shares for each firm: $HHI = \sum_{n=1}^N s_n^2$.

An alternative model of oligopsony would feature firms competing in wages (i.e., analogous to competing à la Bertrand in product markets) as opposed to firms competing in levels of employment (i.e., the analog of competing à la Cournot in product markets). With undifferentiated jobs and symmetric firms, competition in wages leads to a "Bertrand paradox" in which the Nash equilibrium wage is equal to the marginal product of labor as long as there are two or more firms competing in the labor market ([Bertrand, 1883](#)). The reason is that, as long as there is a gap between the marginal revenue product of labor and the wage, any firm in the market has an incentive to offer a wage infinitesimally above the prevailing wage, and hire all the workers in the labor market. This implies that the only equilibrium wage is equal to the marginal product of labor, and thus the wage outcome is the same as under perfect competition. This model is generally considered unrealistic for modeling oligopsony in labor markets but has some application in models with

search frictions, which we discuss later. A model of firms competing in wages is more appealing when jobs are differentiated, as job differentiation can allow for equilibrium wages that are below the marginal revenue product of labor.

2.3 Differentiated jobs

In the Cournot oligopsony model, there is a finite labor supply elasticity to the whole labor market (as opposed to the individual firm). One way to justify this assumption is to posit that jobs in a labor market are differentiated from jobs in other labor markets. Therefore, at least implicitly, the homogeneous jobs Cournot model already hints at the idea of differentiation as a source of monopsony power. However, one can also add differentiation explicitly both across markets and across jobs in the same labor market.

In a differentiated jobs model, the key source of monopsony power is not the finite number of firms as in oligopsony models, but rather the fact that workers have heterogeneous preferences over jobs that differ in wages and amenities. When a worker's job is better (provides higher utility) than their next best option, the firm is able to pay the worker less than the marginal revenue product of labor.

When jobs are differentiated and workers' preferences are heterogeneous, a firm that marginally increases its wage above competitors' levels will not attract all employees from the market. Since workers weigh wages against other job amenities differently, only a negligible number of them will transition to the firm that decides to offer a marginally higher wage. Consequently, even amidst wage-setting competition among firms, equilibrium wages can remain below the marginal product of labor. Introducing job differentiation can therefore resolve the Bertrand paradox. The models we explore in this section are typically premised on the notion that amenities are exogeneously provided by firms at no cost, sidestepping the specifics of how amenities are determined. While the seminal Rosen compensating differentials model (Rosen, 1974, 1986) accounts for the cost of amenities, its integration with monopsony power is not straightforward and remains a subject of active inquiry (Lavetti, 2023). The Rosen framework assumes that amenities are a cost that negatively impacts productivity, and that firms optimally choose the level of amenities depending on the marginal cost of providing them. It also further assumes that wages are equal to marginal productivity. Assuming that wages are competitive is important in this literature to draw inferences about the amenities of different jobs from data on wage differentials (Lavetti, 2023). However, with monopsony power, wages are not equal to marginal productivity. This makes it difficult to use Rosen-style models to estimate the value of amenities in the presence of monopsony power.⁶

⁶ Dube et al. (2022) illustrate the limits of the Rosen model to understand amenity provision. They show that, when firms have monopsony power, minimum wages need not decrease the provision of costly amenities, contrary to this model's predictions. Empirically, they find no effect of Walmart's company-imposed minimum wage on amenities provided by the firm.

Here, we delve into the details of differentiated jobs models with exogenous amenities, which can more easily accommodate monopsony power. These models fall into two categories: representative agent models (Bhaskar and Manning, 2002; Berger et al., 2022a) and discrete choice models of labor supply (Card et al., 2018; Azar et al., 2019). While discrete choice models are more widely used in empirical work, we start with a discussion of the representative choice model for several reasons. First, it is a more straightforward theoretical extension of the oligopsony model, as it assumes Cournot competition, while discrete choice models typically assume Bertrand competition. Additionally, the representative agent model with Cournot competition is more tractable, resulting in a simple formula for the markdown that includes a parameter for job differentiation. Within the representative choice model, we will first assume atomistic firms so that markdowns arise from job differentiation alone rather than from a finite number of firms. We then assume a finite number of firms, enriching the job differentiation model with oligopsony interactions.

2.3.1 *Monopsonistic competition with a representative household and differentiated jobs*

In the representative household model of Berger et al. (2022a), there is one household with a utility function that represents preferences for differentiated jobs, and it supplies some of its labor to each of the various jobs in the economy. This is analogous to a consumer choice model in which the representative household chooses the share of its budget that it will allocate to each good. If a job becomes more attractive, for example due to a higher wage, the representative household increases the share of its labor endowment that it supplies to that job. Again, as in the consumer choice model, unless the choices are perfect substitutes (in this case, perfectly substitutable jobs), a small change in the price (i.e., wage) will not induce a discrete jump in the market share of that choice. Thus, job differentiation induces an upward sloping labor supply to the firm.⁷

Consider a representative household that has preferences over consumption C and a bundle of labor supply to differentiated jobs L , and l_i is the labor supply to firm i . The utility U of the representative household is:

$$U(C, L) = C - \frac{1}{\varphi^{\frac{1}{\varphi}}} \frac{L^{1+\frac{1}{\varphi}}}{1 + \frac{1}{\varphi}}, \quad (4)$$

⁷ Common specifications for the representative agent model include quadratic utility, which leads to a linear system of labor supplies, and constant-elasticity of substitution, which leads to a constant elasticity of labor supply.

where

$$L = \left[\int_0^1 l_i^{\frac{1}{\theta}} di \right]^{\frac{1}{\theta+1}} \quad (5)$$

This preference specification generates constant-elasticity market-level and firm-level labor supply functions. The φ parameter determines the aggregate wage elasticity of labor supply, while θ determines the level of job differentiation in the market, which affects the firm-level elasticity of labor supply ($\bar{\varphi}$ is a labor supply shifter). When θ goes to infinity, jobs become perfect substitutes, and the model becomes the same as a Cournot oligopsony with homogeneous labor (Section 2.2). The household's budget constraint is $pC = \int_0^1 w_i l_i di + b$, where b is nonlabor income.

The first-order condition with respect to l_i yields the inverse labor supply function to firm i (which takes aggregate labor supply L as given):

$$w_i = \frac{1}{\bar{\varphi}^{\frac{1}{\theta}}} L^{\frac{1}{\theta}} \left(\frac{l_i}{L} \right)^{\frac{1}{\theta}} \quad (6)$$

We can rewrite this in terms of the wage index $W = \left[\int_0^1 w_j^{1+\theta} dj \right]^{\frac{1}{1+\theta}}$ as follows:

$$w_i = \left(\frac{l_i}{L} \right)^{\frac{1}{\theta}} W \quad (7)$$

In the equation above, the wage of the individual firm depends on its market share l_i/L and the level of job differentiation θ . However, the wage does not depend on the elasticity of labor supply at the market level φ , because the firm does not take into account its effect on aggregate employment L . In the next subsection, with oligopsonistic competition, firms *will* take into account their effect on aggregate employment L .

Firm i chooses l_i (or, equivalently, w_i) to maximize profits π_i :

$$\pi_i = pF(l_i) - w_i(l_i)l_i \quad (8)$$

The firm takes as given the price of its product (there is therefore an assumption of perfect competition in the product market), as well as the aggregate wage and labor supply. However, while the firm cannot affect the overall level of wages, it has monopsony power over its own wage, because the jobs it offers are differentiated from other firms' jobs. As in the classical monopsony model, the first-order condition with respect to l_i implies that firms equate the marginal revenue product of labor to the marginal cost of labor, with the latter being above the wage:

$$pF'(l_i) = w_i \left(1 + \frac{1}{\theta} \right) \quad (9)$$

Note that in this case, $pF'(l_i)$ is the value of the marginal product of labor and also equal to the marginal revenue product of labor $R'(l_i)$ (see [equation \(1\)](#)). As we will discuss in [Section 2.3.6](#) below (see specifically [equation \(40\)](#)), when there is product market power in addition to labor market power, the value of the marginal product of labor is different from the marginal revenue product of labor.

The markdown is given by:

$$\mu_i \equiv \frac{pF'(l_i) - w_i}{w_i} = \frac{1}{\theta} \quad (10)$$

Thus, the markdown is positive even if, in this case, the firms are atomistic, and the level of labor market concentration is equal to zero because there is an infinite number of firms. As θ goes to infinity, jobs become homogeneous, and the markdown goes to zero because there is no longer any job differentiation. Further, the markdown does not depend on the labor supply elasticity to the market φ because, as we noted above, the firm does not take into account its effect on aggregate employment.

2.3.2 Oligopsony with a representative household and differentiated jobs

If the number of firms is finite, then we can write a similar model but with firms having strategic interactions, and the markdown will depend on the HHI. In this case, firms do not take as given the aggregate wage and aggregate employment. If there are N firms, the labor supply to a firm i is:

$$w_i = \left(\frac{l_i}{L} \right)^{\frac{1}{\theta}} W, \quad (11)$$

$$\text{where } L = \left[\sum_{j=1}^N l_j^{\frac{\theta+1}{\theta}} dj \right]^{\frac{\theta}{\theta+1}} \text{ and } W = \left[\sum_{j=1}^N w_j^{1+\theta} dj \right]^{\frac{1}{1+\theta}}.$$

The first-order condition of firm i again equalizes the marginal revenue product of labor to the marginal cost of labor. However, because firms now perceive the effect of their actions over aggregate employment and wages, the slope of the inverse labor supply with respect to labor now has an extra term that depends on firm i 's payroll market share⁸ $\tilde{s}_i = w_i l_i / (WL)$. The first order condition becomes:

$$pF'(l_i) = w_i \left(1 + \frac{1}{\theta} + \left(\frac{1}{\varphi} - \frac{1}{\theta} \right) \tilde{s}_i \right) \quad (12)$$

⁸ If firms are identical, the payroll market shares are the same as the employment shares and equal to $1/N$.

This implies that the markdown of firm i is given by:

$$\mu_i \equiv \frac{pF'(l_i) - w_i}{w_i} = \frac{1}{\theta} + \left(\frac{1}{\varphi} - \frac{1}{\theta} \right) \tilde{s}_i \quad (13)$$

Taking a weighted average, weighted by payroll shares, yields an expression in terms of the (payroll-share) Hefindahl-Hirschman Index:

$$\sum_{i=1}^N \tilde{s}_i \mu_i = \frac{1}{\theta} + \left(\frac{1}{\varphi} - \frac{1}{\theta} \right) HHI \quad (14)$$

Relative to the markdown without job differentiation (see [equation \(3\)](#)), there are two main differences: (i) the $HHI = \sum_{i=1}^N \tilde{s}_i^2$ is in terms of payroll shares instead of employment shares s_i , and (ii) there is a constant term $1/\theta$ that reflects the fact that, with differentiation, firms have a baseline level of market power even if concentration is zero (i.e., even if they are atomistic). The fact that the HHI is based on payroll instead of employment shares only matters if the firms do not pay the same wage (e.g. because they have different levels of productivity as in [Berger et al. \(2022a\)](#)).

2.3.3 Monopsonistic competition and oligopsony with discrete choice

In contrast to representative household models, workers in a *discrete* choice model only pick one job rather than spreading themselves across jobs. Discrete choice models can be used to analyze how workers make decisions about job choice. These models assume that workers pick the job that gives them the highest utility. Here again, the markdown will depend on firm shares, and on the HHI.

Consider a labor market with J firms offering differentiated jobs. Firms compete by setting wages (Bertrand) and forming expectations about the wages of their competitors. There is a continuum of workers of measure L that have random utility over the jobs. In particular, worker i 's utility for firm j 's job is:

$$u_{ij} = \delta_j + \epsilon_{ij}, \quad (15)$$

where δ_j is the job's mean utility, and ϵ_{ij} is the random component of utility, which has an extreme value distribution. δ_j allows for jobs to be different in a deterministic way, and these differences are valued in the same way by all workers. This can be labeled as vertical differentiation. The random component of utility adds more job differentiation into the model, as a given worker values two jobs differently even if they have the same δ_j . This can be labeled as horizontal differentiation.⁹ Because ϵ_{ij} is random across jobs *and* workers, the ranking of jobs is different across workers. The mean utility is $\delta_j = \alpha + \eta \log(w_j)$. Because the wage is a component of δ_j , two jobs with the same wage can be valued differently by the same worker due to the random

⁹ For example, [Lamadon et al. \(2019\)](#) use this terminology.

component ϵ_{ij} . With these assumptions, it can be shown that the labor market share of firm j is:

$$s_j = \frac{e^{\delta_j}}{\sum_{k=1}^J e^{\delta_k}} \quad (16)$$

This implies that employment at firm j is equal to $s_j \times L$, so we can rewrite this as:

$$\frac{l_j}{L} = \frac{w_j^\eta}{\sum_k w_k^\eta} \quad (17)$$

Firm j chooses its wage to maximize its profits:

$$\max_{w_j} (A_j - w_j) s_j, \quad (18)$$

where A_j is the marginal revenue product of labor.

The first-order condition of firm j is:

$$-s_j + (A_j - w_j) \frac{\partial s_j}{\partial w_j} = 0, \quad (19)$$

where

$$\frac{\partial s_j}{\partial w_j} = \frac{\eta s_j (1 - s_j)}{w_j} \quad (20)$$

From the first-order condition, we obtain the formula that equates the markdown and the inverse-elasticity of labor supply, which in this case is:

$$\frac{A_j - w_j}{w_j} = \frac{s_j}{w_j} \frac{1}{\partial s_j / \partial w_j} = \frac{1}{\eta(1 - s_j)} \quad (21)$$

This formula is similar to [equation \(13\)](#) in the prior [Section 2.3.2](#), with the labor supply elasticity parameters (respectively φ above, and η here) decreasing the markdown while the shares (payroll share \tilde{s}_i above, and share s_j here) increase the markdown. However, the formulas are not identical because here we have Bertrand rather than Cournot competition.

In the prior [Section 2.3.2](#), there was an explicit job differentiation parameter θ . The reader may be wondering how job differentiation is parameterized in this model. Imagine there was no random component of utility ϵ_{ij} in [equation \(15\)](#). In that case, all jobs would be exactly the same, except maybe for the wage. Thus, if a firm reduces the wage below the maximum that is paid by other firms in the labor market, it would get zero workers. Therefore, firms would have no market power without the random component. The fact that the random component has a positive variance is what introduces worker-specific job differentiation, and through it a finite labor supply elasticity to the firm.¹⁰

If firms are symmetric in the sense that $A_j = A$ for all j (and the deterministic component of utility is also equal across firms except maybe for the wage, i.e., $\delta_j = \alpha + \eta \log(w_j)$ as we have assumed so far), there is an exact relation between the markdown and the HHI:

$$\frac{A - w}{w} = \frac{1}{\eta(1 - HHI)}, \quad (22)$$

where $HHI = \frac{1}{J}$, and there is a closed-form solution for the wage as well:

$$w = \frac{A}{1 + \frac{1}{\eta(1 - HHI)}} \quad (23)$$

If firms are not symmetric so that each firm has its own productivity A_j , the markdown in [equation \(21\)](#) doesn't imply an exact relationship between the average markdown and market concentration. However, we can derive an approximate relationship if we assume that firms have similar but not identical market shares.¹¹

Under some assumptions about hours worked (in particular, if workers target a given level of income), it can be shown that the logit model in this section is equivalent to the CES model developed in the previous [Section 2.3.2](#) ([Berger et al., 2019](#)).¹² If instead of Bertrand competition in wages, we assumed that firms compete in employment (Cournot), then the logit model would imply the same relationship between the average markdown and the payroll share HHI as in the representative household CES model, because the

¹⁰ If there were a variance parameter for the random component ϵ_{ij} , this parameter would play a role similar to the θ in the model in the previous section. However, it is convention to fix the variance of the random component in discrete choice models, because, econometrically, this variance cannot be identified separately from the deterministic component parameters (i.e. the coefficients on the observed variables that predict worker choice, such as the wage); indeed, what really matters is the relative magnitude of the random component and the deterministic component. Therefore, with a fixed variance of the random component, it is the magnitude of the deterministic component parameters (relative to that fixed variance) that drives differentiation in this model.

¹¹ The expression using a second-order Taylor approximation around $\frac{1}{J}$ is:

$$\sum_j s_j \frac{A_j - w_j}{w_j} \approx \frac{1}{\eta} \left[\frac{J(J^2 - 3J + 1)}{(J - 1)^3} + \frac{J^3}{(J - 1)^3} HHI \right] \quad (24)$$

¹² [Berger et al. \(2019\)](#) show this is true in the nested logit and nested CES case. Since the non-nested versions are special cases of the nested imposing specific parameter values, their proof also works for the non-nested case. The analogous arguments for the product market logit and CES and nested logit and nested CES are made by [Anderson et al. \(1988\)](#) and [Verboven \(1996\)](#), respectively. Specifically, what these papers show is that a “fictitious” representative consumer with (nested) CES preferences can represent a population of consumers with heterogeneous (nested) logit preferences.

logit and CES models are equivalent in the sense that they imply the same labor supply system.

2.3.4 Nested logit

One can build more flexible discrete choice models by using nesting.¹³ In a nested model, the worker first makes a decision whether to work in a given labor market or not; if the worker decides not to work in that given labor market, the worker chooses the “outside option”. The outside option represents either nonemployment or working in other labor markets. If the worker chooses to work in the market, they next choose which firm to work for within the market. Thus, we have a nested logit framework with two nests: a top-level nest for the decision of whether to work in the market or not, and a bottom-level nest for the decision of which firm in the market to work for. Azar et al. (2022) estimate such a nested model.

Formally, the utility of worker i for working for firm j is:

$$u_{ij} = \alpha + \eta \log(w_j) + v_i(\lambda) + \lambda \epsilon_{ij} \quad (25)$$

Here, $\delta_j \equiv \alpha + \eta \log(w_j)$ is the deterministic component of utility. There are two error terms in the utility function: an extreme value error term ϵ_{ij} , which captures random shocks to the utility of worker i from working at firm j , and a group preference term $v_i(\lambda)$, which represents the worker's overall preference for the group, i.e. the labor market they are considering. This group term has a unique distribution that ensures that the term $v_i(\lambda) + \lambda \epsilon_{ij}$ is also distributed as an extreme value. The nesting parameter $\lambda \in [0,1]$ determines how differentiated the jobs inside the market are from the outside option. When $\lambda = 1$, the model simplifies to a standard logit model, where the outside option is treated as part of the same nest as the focal labor market. This means that jobs within the focal labor market compete just as much with jobs outside the market as within the market, implying no differentiation between the two markets in terms of the worker's choice. Conversely, when $\lambda = 0$, then the labor market is fully segmented from the outside option, and a worker either always works in the market, or always chooses the outside option.

The labor market share of firm j in the nested logit model is:

$$s_j = \underbrace{s_j|_g}_{\text{within-group share}} \times \underbrace{s_g}_{\text{group share}} = \frac{e^{\frac{\delta_j}{\lambda}}}{e^I} \times \frac{e^{\lambda I}}{1 + e^{\lambda I}}, \quad (26)$$

where

¹³ Similarly, nesting can be used to obtain more flexible versions of the representative household CES model.

$$I = \log \sum_{k=1}^J e^{\frac{\delta_k}{\lambda}} \quad (27)$$

is the “inclusive value”, which measures the expected maximum utility of working in the specific labor market group. The group share is the share of workers that work at any job in this particular labor market. The within-group share is the share of firm j within the labor market g .

Assuming that firms compete in wages, the first-order condition of firm j is the same as in the simple logit case, except that the formula for the slope of the market share with respect to the wage now also depends on the nesting parameter and is given by:

$$\frac{\partial s_j}{\partial w_j} = \frac{\eta}{w_j} \frac{1}{\lambda} s_j [1 - (1 - \lambda)s_{j|g} - \lambda s_j] \quad (28)$$

The own-wage elasticity of labor supply to the firm is:

$$\frac{\partial \log s_j}{\partial \log w_j} = \frac{\eta}{\lambda} [1 - (1 - \lambda)s_{j|g} - \lambda s_j] \quad (29)$$

The markdown of firm j is the inverse of the elasticity right above:

$$\frac{A_j - w_j}{w_j} = \frac{\lambda/\eta}{1 - (1 - \lambda)s_{j|g} - \lambda s_j} \quad (30)$$

There is no closed-form solution for the equilibrium markdown, given that the market shares depend on the wage, and that, even in the symmetric case, market shares are not equal to $1/J$ because of the outside option. However, empirically, if we know the elasticity and the market shares, we can calculate the markdown without numerically solving the nonlinear system of first-order conditions.

2.3.5 Nesting with monopsonistic competition

Lamadon et al. (2022) develop an equilibrium model of the labor market with two-sided heterogeneity where workers view firms as imperfect substitutes because of heterogeneous preferences over non-wage job characteristics, i.e. amenities. Such heterogeneous preferences lead to a finite labor supply elasticity to the firm. They define markets as an industry by commuting zone. They allow for correlation of worker idiosyncratic preferences for firms within each nest, with the degree of correlation varying across nests. Each nest is assumed to include many firms, and firms therefore are assumed not to act strategically, that is, firms do not take into account the impact of changing their own wages on the market-level wage. Still, firms do exercise market power by taking into account the finite labor supply elasticity to the firm.

The assumption that there are many firms in each nest simplifies the analysis: it implies that firms take the inclusive value as given when taking

derivatives of their market share with respect to the wage in their first-order condition. In this case, the elasticity of labor supply to the firm in equation (29) becomes simply

$$\frac{\partial \log s_j}{\partial \log w_j} = \frac{\eta}{\lambda} \quad (31)$$

This is equivalent to setting the market share of the firm to zero within the market, reflecting the assumption that the firm is infinitesimal relative to the size of the labor market. In this sense, this is a model of monopsonistic competition as in Section 2.3.1 above, but with discrete choice instead of the representative household framework.

2.3.6 Simultaneous labor and product market power

In the monopsonistic competition and oligopsony models in the previous sections, we have assumed that the product market was competitive. If we assume instead that firms produce differentiated goods, then firms have market power simultaneously in the labor market *and* the product market. This kind of model is derived in a representative household framework by Deb et al. (2022), and in a discrete-choice framework by Kroft et al. (2022).¹⁴

The simplest model to illustrate how this works is a model of simultaneous monopsonistic and monopolistic competition. We use the same notation as in Section 2.3.1 above. Consider a representative household that has preferences over a bundle of differentiated consumption goods C and a bundle of labor supply to differentiated jobs L :

$$U(C, L) = C - \frac{1}{\varphi^{\frac{1}{\varphi}}} \frac{L^{1+\frac{1}{\varphi}}}{1 + \frac{1}{\varphi}}, \quad (32)$$

where

$$C = \left[\int_0^1 c_i^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}} \quad (33)$$

and

$$L = \left[\int_0^1 l_i^{\frac{\theta+1}{\theta}} di \right]^{\frac{\theta}{\theta+1}} \quad (34)$$

Product market power is introduced through the assumption that consumption goods are differentiated, with σ as the parameter that governs the differentiation. When σ goes to infinity, consumption goods become perfect substitutes

¹⁴ Azar and Vives (2021) derive a model of simultaneous product and labor market power, in a general equilibrium oligopoly context instead of monopolistic-monopsonistic competition.

and the product market becomes perfectly competitive; thus we go back to the case of monopsonistic competition in the labor market without product market power in [Section 2.3.1](#). As before in [equation \(11\)](#), when θ goes to infinity, monopsony power disappears. The household's budget constraint is $\int_0^1 p_i c_i di = \int_0^1 w_i l_i di + b$, where p_i is the price of product i , c_i is the consumption of product i , and b is nonlabor income.

Firms take aggregate consumption and aggregate labor as given, and face product demand given by:

$$p_i = \left(\frac{c_i}{C} \right)^{-\frac{1}{\sigma}} P, \quad (35)$$

where $P \equiv \left[\int_0^1 p_i^{1-\sigma} di \right]^{\frac{1}{1-\sigma}}$ is the Dixit-Stiglitz price index. Firm i 's inverse labor supply, as before in [equation \(11\)](#), is given by:

$$w_i = \left(\frac{l_i}{L} \right)^{\frac{1}{\theta}} W \quad (36)$$

Recall that $W = \left[\int_0^1 w_j^{1+\theta} dj \right]^{\frac{1}{1+\theta}}$.

Firm i maximizes profits taking into account the effects of its actions on both its price and its wage. We assume that firm i has a production function that uses only labor, $c_i = F(l_i)$. We express profits in real terms.¹⁵ The firm's profit maximization can be written as:

$$\max_{l_i} \frac{p_i}{P} c_i - \frac{w_i}{P} l_i \quad (37)$$

The first-order condition of firm i sets the marginal revenue from an additional worker equal to marginal revenue product of labor (MRPL, or $R'(l_i)$ in the notation of [equation \(1\)](#)) to the marginal cost of labor (MCL):

$$\left(1 - \frac{1}{\sigma} \right) \frac{p_i}{P} F'(l_i) - \left(1 + \frac{1}{\theta} \right) \frac{w_i}{P} l_i = 0 \quad (38)$$

MRPL *MCL*

Alternatively, we could divide this expression by $F'(l_i)$ and it would read "marginal revenue (MR) equal to marginal cost (MC)" (and this is the

¹⁵ It is helpful to express profits in real terms in order to allow for the general case where firms can influence the overall price level in the economy P . In the case when the firms take the price level as given, as in the model presented here, it does not matter whether profits are expressed in nominal or real terms. However, it does matter in the more general case in which firms can affect the price level in the economy: in that case, if the profits were expressed in nominal terms, the equilibrium would depend on the choice of price normalization ([Gabszewicz and Vial, 1972](#); [Azar and Vives, 2021](#)).

first-order condition that obtains directly when the problem is rewritten for the firm to choose its output level instead of employment l_i):

$$\begin{array}{c} \left(1 - \frac{1}{\sigma}\right) \frac{P_i}{P} - \left(1 + \frac{1}{\theta}\right) \frac{w_i}{P} \frac{1}{F'(l_i)} = 0 \\ \text{MR} \qquad \qquad \qquad \text{MC} \end{array} \quad (39)$$

The equilibrium markdown is the wedge between the MRPL and the real wage, and it is given by:

$$\frac{\frac{R'(l_i)}{P} - \frac{w_i}{P}}{\frac{w_i}{P}} = \frac{1}{\theta} \quad (40)$$

Interestingly, this is the exact same markdown formula as in the case of no product market power (see [equation \(10\)](#)). However, note that in this case the MRPL $R'(l_i)$ includes the firm's product market power. Thus, in this case, unlike in [equation \(10\)](#), MRPL is different from the value of the marginal product of labor (VMPL), which is defined as the price that the firm receives from selling its product times the number of additional units that it produces when it hires an extra worker. The formula for VMPL in this model is $\frac{P_i}{P} F'(l_i)$. The formula for the MRPL includes the term $1 - \frac{1}{\sigma}$ to take into account the fact that, when a firm hires an extra worker and increases output, this reduces its price, which has a negative effect on marginal revenue.

The equilibrium markup is the wedge between the relative price of the firm and the marginal cost (expressed relative to the price level), given by:

$$\frac{\frac{P_i}{P} - MC}{\frac{P_i}{P}} = \frac{1}{\sigma} \quad (41)$$

Under perfect competition in both product and labor markets, the wage is equal to the value of the marginal product of labor, which is also equal to the marginal revenue product of labor. In this model with labor and product market power, we have two forces that create two wedges between the wage and the value of the marginal product of labor:

1. Product market power creates a wedge (markup) between the *value* of the marginal product of labor and the marginal *revenue* product of labor.
2. Labor market power creates a wedge (markdown) between the marginal revenue product of labor and the wage.

In the monopsonistic competition model of the previous section (and also in the oligopsony model), we only had the second force, driven by labor market power.

The overall effect of product *and* labor market power is summarized in the following formula for the equilibrium gap between VMPL and the wage:

$$\frac{\frac{p_i}{P}F'(l_i) - \frac{w_i}{P}}{\frac{w_i}{P}} = \frac{1 + \frac{1}{\theta}}{1 - \frac{1}{\sigma}} - 1 = \frac{\frac{1}{\theta} + \frac{1}{\sigma}}{1 - \frac{1}{\sigma}} \quad (42)$$

Note that, when σ goes to infinity, product market power goes away, and so does the product market power component of the gap (i.e., the denominator goes to one). Similarly, when θ goes to infinity, labor market power goes away. Conversely, when σ decreases, i.e. product market power increases, workers get paid less relative to the value of their marginal product. In this sense, product market power reduces wages beyond the effects of labor market power.

2.4 Search frictions

In search and matching models, the key deviation from perfect competition is search costs, as well as workers' imperfect information. Firms can pay workers less because workers cannot instantaneously meet all available jobs. One example of how imperfect information affects the labor market is the following: workers do not know which jobs other workers have already applied to, and so several workers may end up randomly competing for the same job. Therefore, even when there are more vacancies than job seekers, there is no guarantee that an individual worker will get hired. This assumed lack of coordination is a key mechanism introducing frictions in search and matching models.

2.4.1 Monopsony power with infinitesimal firms: random search, wage posting, and on the job search

Burdett and Mortensen (1998) build a search model in which firms post wages and have monopsony power despite being infinitesimal in the labor market. If workers search for jobs both when they are employed and when they are unemployed, the equilibrium is characterized by a distribution of wages as opposed to a single wage, even when all firms and workers are identical.

In the Burdett-Mortensen model, all firms are infinitesimal relative to the labor market. There is a continuum of workers with mass M_w , and a continuum of firms with mass M_f .

A firm in the labor market offers a wage w , taking as given the distribution of wages across firms in the labor market $F(w)$. Both employed and unemployed workers receive offers, at an exogenous Poisson arrival rate of λ . There is also an exogenous rate of separation to unemployment δ . The productivity of an employed worker is A , and the flow of income in unemployment for the worker is b .

In steady state, the number of workers going into unemployment must be the same as the number of workers leaving unemployment. This implies

that $\lambda u M_w = \delta(1 - u)M_w$. Therefore, the steady state unemployment rate is:

$$u = \frac{\delta}{\delta + \lambda} \quad (43)$$

If a firm increases its wage, then (i) it becomes more successful at hiring the workers it meets, and (ii) it experiences a decrease in the probability that its current employees leave when they receive an offer from other firms. Denote $N(w; F)$ the steady-state level of employment of a firm with wage w when the distribution of wages across firms is F . Then, a firm that sets a wage of w has steady-state profits equal to

$$\pi(w; F) = (A - w)N(w; F) \quad (44)$$

How do we obtain an expression for $N(w; F)$? If a firm pays a wage w , it will lose workers when they go to unemployment, and when they receive offers from other firms that pay a wage higher than w . Thus, if a firm pays w , its separation rate is:

$$s(w; F) = \delta + \lambda(1 - F(w)). \quad (45)$$

The total number of workers the firm loses is equal to the product of the separation rate and the firm's level of employment: $s(w; F)N(w; F)$.

The firm also receives a flow of recruits. If a firm pays a wage w , it recruits a share of the unemployed workers that randomly match with the firm, as well as employed workers that randomly match with the firm, if their current firm pays less than w . Mathematically, the flow of unemployed workers to the firm is $\lambda u \frac{M_w}{M_f}$, and the flow of employed workers to the firm is $\lambda(1 - u)G(w; F) \frac{M_w}{M_f}$, where $G(w)$ is the distribution of wages across workers.

The distribution of wages across workers in the labor market is different from the distribution of wages across firms in the labor market. The reason is that higher wage firms have more workers. To obtain an expression for $G(w)$, consider the set of workers with wages less than or equal to w . There can be transitions among workers between firms in this set, which doesn't affect the fraction of workers in the set. However, there cannot be transitions from workers outside of this set directly to this set, because those workers' wages are higher than w and they would not accept offers from firms paying w or less. Thus, the net transitions into the set are only from unemployment. The rate of transitions into this set is the fraction of unemployed workers who get an offer $u\lambda F(w)$. Transitions out of this set can be both to firms with wages higher than w , at a rate $\lambda(1 - u)G(w)(1 - F(w))$, or to unemployment, at a rate $\delta(1 - u)G(w)$.

In steady state, the rate of entry and the rate of exit from this set have to be equal, implying a distribution of wages across workers:

$$G(w; F) = \frac{\delta F(w)}{\delta + \lambda(1 - F(w))} < F(w) \quad (46)$$

With this expression for the distribution of wages across workers, we can also now express the flow of recruits to the firm as:

$$R(w; F) = \frac{M_f \lambda \delta}{M_w} \left(\frac{1}{\delta + \lambda(1 - F(w))} \right) \quad (47)$$

In steady-state, the flow of recruits to the firm has to be equal to the flows of workers out of the firm, that is $R = sN$. This gives us an expression for $N(w; F)$:

$$N(w; F) = \frac{M_f \lambda \delta}{M_w \{\delta + \lambda[1 - F(w)]\}^2} \quad (48)$$

Substituting into the profit function implies:

$$\pi(w; F) = \frac{M_f \lambda \delta (A - w)}{M_w \{\delta + \lambda[1 - F(w)]\}^2} \quad (49)$$

It can be shown that in equilibrium, firms are indifferent between setting any wage between b and $A - \left(\frac{\delta}{\delta + \lambda}\right)^2(A - b)$. They can choose a lower wage and have higher profit margins per worker, but have a lower level of employment, or they can choose a higher wage and make lower profit margins per worker, but at a higher level of employment. The equilibrium distribution of wages across firms adjusts such that there is indifference between any two wages. The equilibrium distribution of wages across firms is:

$$F^*(w) = \frac{\delta + \lambda}{\lambda} \left(1 - \sqrt{\frac{A - w}{A - b}} \right) \quad (50)$$

Because firms have the same productivity and are indifferent across wage levels, the firm-level markdown is not uniquely defined (this is why there is no entry for the firm-level markdown in our summary [Table 1](#)).

The average wage is in between the utility from unemployment and the value of the marginal product of labor:

$$E(w) = \frac{\delta}{\delta + \lambda}b + \frac{\lambda}{\delta + \lambda}A \quad (51)$$

When unemployment is higher, the wage is lower and closer to the utility of unemployment, and when unemployment is lower, the wage is closer to the value of the marginal product of labor. If the unemployment rate goes to zero (which happens when the job separation rate δ goes to zero), the wage goes to the value of marginal product of labor A . If the job offer arrival rate λ goes to infinity, then the wage also goes to the marginal product of labor. This shows that, in this model, monopsony power is enabled by both a non-zero job loss probability, and a less than one probability of job finding.

One can write the markdown relative to the expected wage $(A - E(w))/E(w)$ as follows.¹⁶

$$\frac{A - b}{b + \frac{\lambda}{\delta}A} \quad (52)$$

When firm productivity is heterogeneous and uniformly distributed between zero and one, [Burdett and Mortensen \(1998\)](#) derive an expression for the equilibrium wage as a function of productivity:

$$w(A) = \frac{\frac{\lambda}{\delta}A^2}{1 + \frac{\lambda}{\delta}} \quad (53)$$

This implies that the firm-level markdown is:

$$\frac{1 + \frac{\lambda}{\delta}(1 - A)}{\frac{\lambda}{\delta}A} \quad (54)$$

[Postel-Vinay and Robin \(2002\)](#) consider a version of [Burdett and Mortensen \(1998\)](#) where firms are able to counter-offer when employed workers receive an outside offer. Indeed, in [Burdett and Mortensen \(1998\)](#), firms are passive, and workers leave when they receive a higher offer than their current wage. In [Postel-Vinay and Robin \(2002\)](#), when an employed worker receives an offer from another firm, the firms engage in Bertrand competition. When all workers have the same opportunity cost of employment and all firms are also identical, there are only two wages in this economy: the "monopsony wage" offered to the unemployed, and the competitive wage offered to the employed who received an outside job offer and who benefited from the all out Bertrand competition between the two firms. When introducing firm productivity dispersion, there is additional wage dispersion, because more productive firms are able to offer higher wages to poach already employed workers. Offers and counter-offers for employed workers thus play an important role in wage growth, and allow workers to escape the "monopsony wage."

2.4.2 Monopsony power with large firms: labor market concentration in a random search model

[Jarosch et al. \(2019\)](#) develop a random search model with multi-vacancy firms, based on the Diamond-Mortensen-Pissarides wage bargaining framework. In contrast to the approach in [Burdett and Mortensen \(1998\)](#), where firms post wages, in the [Jarosch et al. \(2019\)](#) framework, firms do not post wages and bargain over wages with workers after meeting them. The key difference between [Jarosch et al. \(2019\)](#) and the standard random search and bargaining

¹⁶Note that this is not exactly the same as the expected markdown.

model is that, in [Jarosch et al. \(2019\)](#), firms are not atomistic, so there is a positive probability that a worker will encounter the same firm in the future.

Consider a model with N firms in which the firms are granular, i.e., large relative to the market. Job openings occur exogenously, and the probability that a job opening is by firm i is f_i , where $\sum_{i=1}^N f_i = 1$. We refer to f_i as the firm's market share or its "size". There is a fixed cost c_i per job opening. The productivity of a worker is normalized to one.

The matching process involves u workers applying for v vacancies, and the job finding rate is $\lambda \equiv \frac{v}{u}(1 - e^{-\frac{v}{u}})$. Separations occur exogenously at a rate δ . Firms and workers engage in Nash bargaining, with worker bargaining power parameter $\alpha \in [0,1]$. The value of unemployment is U , and the value of employment at firm i is W_i . The value of unemployment is given by:

$$U = b + \beta \left(\lambda \sum_{i=1}^N f_i W_i + (1 - \lambda) U \right), \quad (55)$$

where b is the flow of income to an unemployed worker, and β is the workers' discount factor.

The value of working for firm i is given by:

$$W_i = w_i + \beta(\delta U + (1 - \delta) W_i) \quad (56)$$

The main departure from the standard model is that, if bargaining between a firm and an unemployed worker breaks down, the firm will not hire the worker if the worker matches with the firm again during the same unemployment spell (unless the worker is the only applicant, which happens with probability $\underline{\lambda} \equiv e^{-\frac{u}{v}}$). After the worker finds a new job, the no-rehire policy ends: that is, if the worker finds at least one new job and then becomes unemployed again, the firm does not refuse to hire the worker even if the worker rejected the firm's offer in a prior unemployment spell. In an off-equilibrium path in which the bargaining between an unemployed worker and firm i breaks down, the continuation value for the worker is:

$$U_i = b + \beta \left(\lambda \sum_{j \neq i} f_j W_j + \underline{\lambda} f_i W_i + (1 - \lambda(1 - f_i) - \underline{\lambda} f_i) U_i \right) \quad (57)$$

For this punishment mechanism to work, the firm has to be able to track applicants, and the threat not to hire re-applicants has to be credible. [Jarosch et al. \(2019\)](#) show that, even if firms cannot commit, the firm's threat not to rehire the worker is still operative as long as it is more costly to the worker than to the firm, which occurs when the firm can find a close enough substitute for the worker.

The value of a bilateral relationship for firm i is:

$$J_i = 1 - w_i + \beta(1 - \delta) J_i \quad (58)$$

The value of a job opening is:

$$V_i = -c_i + \beta(1 - e^{\frac{u}{v}}) J_i \quad (59)$$

The joint surplus from a match is:

$$S_i \equiv W_i - U_i + J_i \quad (60)$$

The surplus is split between the worker and the firm, such that the worker's surplus is $W_i - U_i = \alpha S_i$, and the firm's surplus is $J_i = (1 - \alpha)S_i$.

[Jarosch et al. \(2019\)](#) show that, in this model, the average wage $\bar{\omega}$ is negatively related to labor market concentration C , in the following way:

$$\bar{\omega} = 1 - (1 - \alpha) \frac{1 - \beta(1 - \delta)}{1 - \beta(1 - \lambda\alpha(1 - C) - \delta(1 - \tau C))}, \quad (61)$$

where

$$\tau = \alpha \frac{\beta(\lambda - \underline{\lambda})}{1 - \beta(1 - \lambda)} \quad (62)$$

$$C \equiv \frac{\sum_{k=2}^{\infty} \tau^{k-2} f^k}{1 + \tau \sum_{k=2}^{\infty} \tau^{k-2} f^k}, \quad (63)$$

where, in turn, $f^k \equiv \sum_i f_i^k$, such that f^1 is equal to 1, and f^2 yields the labor market HHI. Higher orders of f^k are non-standard measures of concentration that differ from the HHI.¹⁷ In practice, concentration C is very similar to the standard HHI in the Austrian data used by [Jarosch et al. \(2019\)](#). Here, the HHI measures how often a worker who searches randomly encounters a job vacancy from their current employer: intuitively, the higher an employer's market share, the more likely the worker would meet the employer again and thus not be rehired.

Firm-level wages for two firms i and j depend on their sizes f and are characterized by:

$$\frac{1 - w_i}{1 - w_j} = \frac{1 - \tau f_j}{1 - \tau f_i} \quad (64)$$

This model demonstrates that HHI can measure market power not only in an oligopsony model, but also in a search and matching model. In both models, firms with higher market shares have more market power. In the oligopsony model, larger firms face higher wage costs when hiring an additional worker, because they have to raise the wages of their existing inframarginal workers. Therefore, larger firms keep wages low. In [Jarosch et al. \(2019\)](#), larger firms can keep wages low because they can prevent job seekers from accessing their future vacancies. Therefore, larger firms have more leverage in bargaining with workers over the wage. To note, the distribution of firm size is taken as given in [Jarosch et al. \(2019\)](#), and therefore the model is not helpful to understand the impact of monopsony power on employment or the firm size distribution. In

¹⁷This formula is derived in the Appendix of [Jarosch et al. \(2019\)](#).

contrast, it is a useful model to understand workers' bargaining leverage in wage negotiations with firms of different sizes, and to understand how this bargaining leverage may be affected if firms collude.

2.4.3 Monopsony power in a directed search model where firms have multiple employees

Rudanko (2023) explores wage setting in a directed search model of multi-worker firms facing within-firm equalizing constraints on wages. The paper builds a model of directed search as in Moen (1997), but with multi-vacancy firms. When firms have multiple vacancies and equality constraints require them to pay the same wage to their existing workers as they offer new hires, wages are reduced through a mechanism similar to that in the oligopsony model in Section 2.2. However, in contrast to the oligopsony model, paying lower wages leads equality-constrained firms to hire *more* workers rather than fewer workers, and equilibrium employment is higher than without equality constraints. The surprising positive employment effect of equality constraints can be explained as follows: lower equilibrium wages encourage firms to post more vacancies, and workers are willing to apply as long as posted wages are higher than the reservation wage.

In this model, there are a large number I of firms. Firm i begins the period with n_i workers, and needs to decide how many vacancies to create v_i , and at what wage w_i . The labor force is normalized to 1. The total number of employed workers is a share $N = \sum_{i=1}^I n_i$ of the labor force, and there are $1 - N$ unemployed workers. Output per worker at a firm is constant at A , and the flow income for the unemployed is $b < A$. The cost of posting v vacancies (where v can be non-integer) for a firm with employment n is $\kappa(v/n)n$, where κ is increasing in v/n and convex.

Vacancies that have the same wage are grouped into a labor submarket (possibly composed of just one firm). There is a homothetic matching function $m(v, u)$ that determines how many matches are created as a function of the number of vacancies and unemployed workers (there is no on-the-job search in this model). Tightness in a labor submarket is defined as the number of vacancies per unemployed worker searching in that submarket. The probability that a vacancy is filled in a labor submarket i with tightness $\theta_i = v_i/u_i$ is $q(\theta_i) = m(v_i, u_i)/v_i = m(1, 1/\theta_i)$, while the probability that a worker finds a job in a market with tightness θ_i is $\mu(\theta_i) = m(v_i, u_i)/u_i = m(\theta_i, 1)$. The value of search at a market with tightness θ_i and wage w_i is the probability of finding a job multiplied by the market wage w_i , plus the probability of not finding a job multiplied by b . In equilibrium, the value of search has to be the same across all labor markets and equal to U , such that

$$U = \mu(\theta_i)w_i + (1 - \mu(\theta_i))b \quad (65)$$

Equation (65) implicitly defines tightness as a function of the wage, conditional on the value of search U in the labor market. We denote this function $g(w_i; U)$. This function captures the fact that if firm i increases its wage, then

more unemployed workers are attracted to the market, and therefore tightness is lower. Lower tightness in turn implies a higher probability of filling vacancies for the firm, and a lower probability of finding a job for the workers searching in the labor submarket. The firm is aware of its effect on tightness when it chooses its wage and how many vacancies to create. In particular, the firm chooses its wage and the number of vacancies to maximize profits:

$$\max_{w_i, v_i} (n_i + q(\theta_i)v_i)(A - w_i) - \kappa(v_i/n_i)n_i, \quad (66)$$

subject to the constraint in [equation \(65\)](#).

This maximization problem implicitly imposes an equality constraint, namely that the wages of existing employees must also adjust to be equal to the wages of the newly posted vacancies.

This problem is scale-independent and can be rewritten in terms of $x = v/n$, which is the number of vacancies divided by the stock of employed workers of a firm:

$$\max_{x_i} (1 + q(\theta_i)x_i)(A - w_i) - \kappa(x_i) \quad (67)$$

The first-order condition with respect to vacancies is:

$$\kappa'(x) = q(\theta)(A - w) \quad (68)$$

On the left-hand side, we have the marginal cost of increasing vacancies per existing employee. On the right-hand side we have the probability of filling the vacancy multiplied by the profit margin per worker.

The first-order condition with respect to the wage is:

$$1 + q(\theta)x = q'(\theta)g_w(w; U)x(A - w) \quad (69)$$

On the left-hand side, we have the cost of increasing wages by one dollar, which is the expected number of employees after hiring (expressed per existing employee). On the right-hand side, we have the benefit of increasing wages, which is the increase in the probability of filling vacancies when tightness goes down $q'(\theta)$, multiplied by the change in tightness when the wage changes $g_w(w; U)$, multiplied by the number of vacancies per existing employee, multiplied by the profit margin.

If we don't impose the equality constraint that wages for existing employees have to adjust to the newly posted wages, then the first-order condition for vacancies is exactly the same, and the first-order condition for wages is:

$$q(\theta)x = q'(\theta)g_w(w; U)x(A - w), \quad (70)$$

reflecting that the cost of increasing wages is lower compared to [equation \(69\)](#), because new wages do not apply to the existing workers.

An equilibrium in this model is a wage, a level of vacancies, a level of tightness, and a value of search that solve the maximization problem of the firm, and such that each unemployed worker applies to one firm:

$$1 - N = xN/\theta \quad (71)$$

In both the constrained and unconstrained cases, the equilibrium wage can be expressed as a convex combination of the flow of income when unemployed and the output per worker when employed:

$$w = \gamma b + (1 - \gamma)A \quad (72)$$

Note that the expression for the wage is similar to that for the expected wage in the random search model of Burdett-Mortensen (equation (51)), in that the wage is a weighted average of the flow of income when unemployed and productivity A . Of course, the expression for the weights is different in the two models.

The equilibrium markdown in this model can be written as:

$$\frac{A - w}{w} = \gamma \frac{A - b}{\gamma b + (1 - \gamma)A} \quad (73)$$

In the unconstrained case, $\gamma = \epsilon$, where ϵ is the matching function elasticity:

$$\epsilon = \frac{\mu'(\theta)\theta}{\mu(\theta)} \quad (74)$$

In the constrained case, the expression for γ is:

$$\gamma = \frac{\epsilon/\tau}{1 - \epsilon + \epsilon/\tau}, \quad (75)$$

where $\tau = \frac{q(\theta)x}{1 + q(\theta)x}$ is the firm's new workers as a fraction of its overall employment (after recruiting). If τ goes to one, which means that all the employment of the firm is composed of new recruits, then γ goes to ϵ , which takes us back to the unconstrained case. If τ goes to zero, then γ goes to one, and the wage goes down to the flow of income when unemployed b . Intuitively, when the number of new recruits is small relative to the firm's existing workforce, the cost of increasing wages to attract more recruits is very high.

Note that γ in the constrained case is between ϵ and one, and is thus always higher in the constrained case than in the unconstrained case. This also implies that the equilibrium markdown is higher, and the equilibrium wage is always lower in the constrained case relative to the unconstrained case (also see the wage equation (72)).

The first-order condition for the wage is essentially the same as the first-order condition with respect to the wage in the classical monopsony model (equation (1)). In Rudanko (2023), when ϵ – the elasticity of the probability of finding a job with respect to tightness – is higher, the markdown is higher. The

reason is that an increase in ϵ reduces the absolute value of the elasticity of the probability of filling out a vacancy with respect to tightness, which is $1 - \epsilon$. This elasticity is directly related to the firm-level elasticity of labor supply in this model, and a lower elasticity of labor supply implies a higher markdown in equilibrium, just as in the classical monopsony model (see [equation \(2\)](#)).

Because wages are lower in the constrained case than in the unconstrained case, firms optimally post more vacancies in the constrained case. This can be seen using [equation \(71\)](#), which implies that $\kappa'(x)/q(\theta)$ is increasing in x . Knowing this, we can see from [equation \(68\)](#) that an increase in $A - w$ implies an increase in equilibrium vacancy creation x . Given that workers accept jobs as long as the wage is above the flow of income in nonemployment b , an increase in vacancies in this model leads to higher equilibrium employment.

The intuition for a positive equilibrium employment effect is thus similar to that in the generalized oligopsony model of [Azar et al. \(2021\)](#): the firm has two decisions, a wage decision and a hiring intensity decision. The lower equilibrium wage from monopsony raises the incentive to increase employment, by increasing hiring effort as in [Azar et al. \(2021\)](#) or by increasing the number of vacancies as in the [Rudanko \(2023\)](#) model. It is interesting, however, that in the case of directed search à la [Rudanko \(2023\)](#), there is *always* a positive effect of monopsony power on employment, while in the generalized oligopsony model, the effect is ambiguous. The reason for the unambiguous result in [Rudanko \(2023\)](#) is that firms do not anticipate that increasing vacancies will increase market tightness, because firms see themselves as infinitesimal relative to the labor market. In both cases, however, it is crucial for the positive employment effect for firms to be able to hire from a pool of unemployed workers, as opposed to the classical monopsony case in which the equilibrium is always on the aggregate labor supply curve, and therefore there is no unemployment.

There are two opposing forces at play in these models. The first is the upward-sloping market-level labor supply, which implies that when wages are lower, workers supply less labor to the market. The second force is firms' incentives to invest in costly hiring efforts: when firms have a hiring intensity ([Davis et al., 2012](#)) or a vacancy margin, lower wages imply a higher incentive to invest in hiring. This second force can sometimes dominate the first one, leading to an increase in employment when wages are lower. In the case of [Rudanko \(2023\)](#), the first force is shut down because the market-level labor supply is vertical in the relevant wage range (workers accept any wage above the reservation wage). Therefore, in that model with a wage equality constraint and monopsony power, the second force always dominates and employment is higher.

A positive employment effect of monopsony power is an interesting theoretical finding because (i) it helps rationalize some empirical results that find greater concentration leading to lower wages without lower employment, and (ii) it has very different implications for the effects of monopsony power on efficiency compared to the classical model. In the oligopsony model,

equilibrium employment is below the efficient level, while in the [Rudanko \(2023\)](#) model, equilibrium employment is above the efficient level. How could employment be inefficiently high? An excessive employment equilibrium happens when workers have high employment but low wages: in this case, average workers' expected income is lower than in the constrained efficient equilibrium that has lower employment but significantly higher wages. In other terms, it can be more advantageous for workers to cycle more often through unemployment if this maintains sufficiently high wages in equilibrium.

2.5 Discussion

2.5.1 *The role of firm size heterogeneity*

Our theoretical discussion makes it clear that monopsony power can exist even when firms all have very small market shares, due to e.g. job differentiation and search frictions.

While the core theoretical insights about the sources of monopsony power do not hinge on the explicit consideration of firm size heterogeneity within a market, incorporating this aspect reveals additional, nuanced insights into labor market outcomes, monopsony power, and firm size. Here, we use firm size and firm share interchangeably, but strictly speaking the theoretical results are about firm share within a labor market.

Firstly, differences in firm size within a labor market affect the formula for the HHI. When some firms are larger than others, the HHI is no longer just the inverse number of firms, but reflects the firm size distribution.

Secondly, differences in firm size also lead to markdown heterogeneity within a market; whether larger firms have lower or higher markdowns depends on the model. In oligopsony models and job differentiation models, the markdown increases in firm share (see [Table 1](#)). By contrast, in [Burdett and Mortensen \(1998\)](#), larger firms pay higher wages and make lower profits per worker.

Thirdly, when firm size differences are endogenized based on productivity, complex relationships emerge between the markdown, the wage, productivity, and firm size. In [Burdett and Mortensen \(1998\)](#), the markdown still decreases with firm productivity and size, at least as long as productivity is uniformly distributed (see [equation \(54\)](#) above). However, in [Postel-Vinay and Robin \(2002\)](#); [Rudanko \(2023\)](#), the relationship is more complex. The rise in markdown with firm share and productivity dampens the rise in wage with productivity, and, under some conditions, the relationship could be reversed, resulting in more productive firms paying lower wages. For example, in [Postel-Vinay and Robin \(2002\)](#), this reversal can occur because more productive firms can attract new workers with the promise of larger future wage increases, allowing more productive firms to offer lower wages to new recruits. Therefore, one cannot simply conclude that oligospongy models fit the data better than

search models from the fact that there is a positive relationship between markdown and firm size.

Lastly, differences in firm productivity and firm size within a labor market can also help explain labor reallocation after a minimum wage increase. We will explore this further in [Section 4.3](#), drawing on [Dustmann et al. \(2022\)](#).

2.5.2 *Determinants of monopsony power across models*

Theoretically, the wage markdown increases (and hence wages decrease) with (see [Table 1](#)):

- The inverse labor supply elasticity
- Labor market concentration
- The inverse substitutability between jobs
- The opposite of the income flow while unemployed

Recent empirical work on monopsony power has focused on estimating the labor supply elasticity and the impact of labor market concentration on wages. Both of these quantities can affect wages across a variety of models. Labor market concentration becomes particularly relevant when there are relatively few firms in the labor market, as would be the case in markets affected by anticompetitive mergers.

[Table 3](#) organizes the various types of models and mechanisms of monopsony power documented in the literature. For each class of models, we summarize the key deviation from the perfectly competitive labor market model, and in particular what gives firms the ability to pay wages below the marginal revenue product of labor (marginal productivity). In the last column, we summarize the key factors that determine the level of labor market power in each class of models. The Table can help guide researchers in choosing modeling frameworks that best capture key features of the problem they are interested in. Oligopsony models are the most straightforward to derive, and are particularly useful when analyzing changes in firm shares in the labor market, for example as a result of mergers. Search models are useful to analyze job search, job finding and recruitment and retention efforts, e.g. when evaluating the effects of non-competition agreements. Finally, differentiated job models are particularly useful when dealing with non-wage job characteristics and amenities, such as geographic distance. Our policy-oriented monopsony review paper ([Azar and Marinescu, 2024](#)) illustrates how these models may be applied to a large array of policy questions.

3 Empirically measuring monopsony power

3.1 Definition of the markdown in the empirical literature

Based on the theories in [Section 2](#), defining the markdown as $(MRPL - w)/w = MRPL/w - 1$ is most practical. This definition directly connects to the

TABLE 3 Mechanisms of monopsony power across models.

Class of monopsony model	Key deviation from perfect competition	Why firms can pay workers less than their marginal productivity	Determinants of market power
Oligopsony	Finite number of firms that engage in strategic interactions	Workers' labor supply elasticity to the market is limited	Labor market concentration, labor supply elasticity to the market
Search	Matching frictions	Workers cannot instantaneously meet all available jobs	Labor market concentration, workers' outside option, matching frictions, bargaining power (random search)
Differentiated jobs with monopsonistic competition	Heterogeneous preferences over jobs that differ in wages and amenities	The job is better than the next best alternative for the worker	Job substitutability (workers' preferences)

formulas in Table 1, facilitating a smoother transition from the empirical estimate of the markdown to the theoretical framework. The markdown ($MRPL - w$)/ w measures the percent wage increase that would result, all other things equal, if all monopsony power were eliminated so that $w = MRPL$. We will use this definition of the markdown in our tables that summarize empirical estimates (Tables 4, 5 and 6). It is important to know that the empirical literature has also used other definitions of the markdown, including $MRPL/w$, or $(MRPL - w)/MRPL = 1 - w/MRPL$, or $w/MRPL$. All of these alternative definitions of the markdown, except the last one ($w/MRPL$), are monotonically increasing functions of the markdown as we define it. The reader should be especially careful when an article defines the markdown as $w/MRPL$ because this expression is *decreasing* in the other definitions of the markdown, and this means that the signs of relationships with other variables are opposite. In general, readers should double-check what definition of the markdown authors are using.

3.2 Elasticity of labor supply

Authors typically calculate the markdown as the inverse of their estimated firm-level labor supply elasticity, which is consistent with most models, as can be seen in Tables 1 and 2.

TABLE 4 Elasticity of labor supply and markdowns across studies estimating the elasticity of labor supply.

Study	Data set	Identification	Firm-level elasticity	Market-level elasticity	Estimated markdown ^a
Dal et al. (2013)	Regional Development Program (RDP) experiment in Mexico (2011)	Random assignment of wages and job offers to prospective public servants	2.15		not reported
Dube et al. (2020)	Observational and experimental Amazon Mechanical Turk (MTurk) data, 2014 –2017	Plausibly exogenous variation in pay using double machine learning	0.096		Workers are paid less than 13% of their productivity. ^b
Caldwell and Oehlsen (2022)	Uber data from the US, 2016 (Boston) and 2017 (Houston)	Random assignment of Uber driver pay increases; comparison of drivers with/without access to Lyft to get at market vs. firm level elasticity	2	0.8 ^c	30%
Bassier et al. (2022)	Oregon (USA) unemployment insurance data, 2000 –2017	Compare the worker separation probability between high and low wage firms using fixed effects and matching	4.2		20%
Coolsbee and Syverson (2023)	IPEDS data on faculty at US four-year not-for-profit colleges and universities, 2002 –03 to 2016 –17	Estimate inverse labor supply elasticity using lagged college applications as instrument for faculty employment	5.1 ^d , 28.6 ^e		not reported
Amadio and de Roux (2023)	Colombian manufacturing, exports, and exchange rate data, 1994 –2009	Response of wages and employment to plant-specific shocks due to plausibly exogenous exchange rate variation	2.5		40 %

^aAll estimated markdowns are reported as percentages calculated as $(\text{MRPL}-w)/w$, where MRPL is marginal revenue product of labor and w is the wage.

^bThe markdown corresponding to an elasticity of 0.096 would be $1/0.096 = 1042\%$.

^cMarket defined as ride-share driving in Houston; this elasticity includes both extensive and intensive margins.

^dFor tenure-track faculty.

^eFor non-tenure track faculty.

Identifying the elasticity of labor supply requires a plausibly exogenous source of variation in the wage. Further, it is important to consider whether one is measuring the elasticity of labor supply to an individual firm or to a whole labor market (for theoretical implications of the market-level vs. firm-level elasticities see [Tables 1](#) and [2](#)). If the elasticity of labor supply is to the market, then one needs to define which market this is.

Further, the empirical setups used to identify the labor supply elasticity may focus on a specific part of the process leading to a firm's employment level. Some papers focus on job applications and/or hires, while others focus on job separation. This focus may be due to data availability, and/or may enable a stronger identification strategy. Whatever the reasons for this restricted focus, when a paper does not have as an outcome the firm's employment level, one needs to transform the elasticity estimates in order to get to a labor supply elasticity (wage effect on employment levels) rather than e.g. an application elasticity or a separation elasticity. The authors typically perform these transformations, and we report below the labor supply elasticity as calculated by the authors.

3.2.1 Experiments

[Dal et al. \(2013\)](#) examine an experiment conducted as part of Mexico's Regional Development Program (RDP) to increase the federal presence in certain municipalities. In the first stage, two different wage offers were randomly offered to prospective public servants across recruitment sites. In the second stage, eligible applicants were selected at random to receive job offers. A 33 % increase in wages led to a 26 % increase in applications and a 35 % increase in the conversion rate (i.e. the rate at which job offers are converted into filled vacancies), implying a labor supply (arc-)elasticity of around 2.15.

[Dube et al. \(2020\)](#) use Amazon Mechanical Turk (MTurk) data from 2014 to 2017, including data from five MTurk experiments. They use both the experiments, and observational data with machine learning to estimate the labor supply elasticity. They find low labor supply elasticities, with an elasticity of 0.096 based on the double machine learning specification, which is their preferred specification. Experimental estimates give similar elasticities. Based on the range of elasticities estimated, they conclude that workers on MTurk are paid at most 13 % of their productivity.

[Caldwell and Oehlsen \(2022\)](#) analyze Uber field experiments with an "earnings accelerator" feature in the U.S., in Houston (2017) and Boston (2016). This accelerator introduces temporary, exogenous variation in wages among Uber ridesharing drivers. The experiments offered drivers a 10 %–50 % increase in earnings per trip for one week. For a 10 % increase in wages, female drivers work 8 % more hours (elasticity of .8), and male drivers work only 4 % more hours (elasticity of .4). Uber's firm level labor supply

TABLE 5 Elasticity of labor supply and markdowns across studies using structural estimation.

Study	Data set	Identification	Firm-level elasticity	Market-level elasticity	Estimated markdown ^a
Azar et al. (2022)	CareerBuilder.com application and vacancy data from the US, April–June 2012	Nested logit model. Instrument for the posted wage using average predicted wages for vacancies in other markets	4.8	0.5 ^b	21 %
Lamadon et al. (2022)	US employer-employee panel data of workers aged 25–60 based on tax filings, 2001–2015	Nested logit model. Instruments (auctions and Bartik) to recover firm level labor supply elasticity.	6.5	4.6 ^c	15 %
Roussille and Scuderi (2023)	Hired.com data on San Francisco-based job candidates	Very detailed observational data on the hiring process	3.6–5.7		24 % ^d
Kroft et al. (2022)	Employer-employee panel data on US construction firms and procurement auctions, 2001–2015	Estimate inverse labor supply elasticity, with exogenous employment variation arising from firms winning procurement auctions	4.1		25–29 % ^e

^aAll estimated markdowns are reported as percentages calculated as $(\text{MRPL}-w)/w$, where MRPL is marginal revenue product of labor and w is the wage.

^bMarket defined as SOC-6 occupation and commuting zone.

^cMarket defined as two-digit industry and commuting zone.

^dUnder the preferred, monopsonistic model, the average predicted markdown is \$31,640 (or 19.5 % of productivity). So $(\text{MRPL}-w)/\text{MRPL} = 1-w/\text{MRPL} = 0.195$. So $w/\text{MRPL} = 0.805$, $\text{MRPL}/w = 1.24$.

^eThey report markdowns in Table 3. “The first column of Panel A provides our estimate of the wage markdown relative to MRPL , $(1+\theta)-1$, using either the DiD or RDD estimand.” Converting these estimates of .803 and .777 to the $(\text{MRPL}-w)/w$ terms gives $(1-803)/.803$ and $(1-777)/.777$, or 25 % and 29 %.

elasticity is interpreted as a market level labor supply elasticity when Uber is a monopsonist for drivers who could not shift to Lyft, while Uber's elasticity is interpreted as a firm-level elasticity when drivers could shift to Lyft. Based on this approach, the authors find a market level labor supply elasticity of 0.8 and a firm-level elasticity of 1.5. In the Uber data, there is variation in both hours (intensive margin) and employment (extensive margin), while most of the literature estimates a firm-level labor supply elasticity by focusing on the extensive margin. To allow for a more straightforward comparison, the authors calculate an extensive margin firm-level labor supply elasticity, while accounting for additional driver entry that would occur if Uber were to increase its wages for all drivers. Calculated in this way, the firm-level labor supply elasticity is about 2. Based on the latter estimate, the authors report a markdown of 30 %.

3.2.2 Other forms of identification

[Bassier et al. \(2022\)](#) use Oregon, USA, UI microdata from 2000–2017. They estimate the separation elasticity, i.e. the change in the probability that a worker separates from a firm in response to a change in the wage. They use fixed effects and matching for identification. They match workers with relevant controls by focusing on pairs of workers who began at the same origin firm (step 1) and transitioned to a new firm (step 2). The separation elasticity at the new step 2 firm is estimated by comparing the separation probability of workers who are in a step 2 firm with a higher wage vs. a step 2 firm with a lower wage. They estimate the firm's wage using firm fixed effect derived from an AKM regression ([Abowd et al., 1999](#)), which allows them to focus on the part of the wage that is due to the firm rather than to worker characteristics. They use a stacked event-study design centered on the transition to a new firm (step 2), and where the outcome of interest is workers' separation probability to a final firm or to unemployment. They regress the separation outcome on the individual wage difference between the new (step 2) and old (step 1) firm instrumented by the difference in the AKM firm fixed effects, and controlling for a rich set of variables capturing workers' history (including a fixed effect for the origin firm). The separation elasticity estimate from the preferred matched event study approach is – 2.1. The preferred labor supply elasticity is 4.2. They find markdowns of around 20 %.

[Goolsbee and Syverson \(2023\)](#) estimate the labor supply elasticity for faculty at US four-year non-profit colleges and universities. They estimate the inverse labor supply elasticity to the firm by regressing faculty wages on faculty employment. They instrument employment with a labor demand shifter, here the lagged number of student applications at the institution. They find a 5.1 firm-level labor supply elasticity for tenure-track faculty and a highly elastic labor supply for non-tenure track faculty (28.6). Among tenure-track faculty, the labor supply elasticity declines with rank, with a labor supply

TABLE 6 Elasticity of labor supply and markdowns across studies using mergers and concentration, calibration and simulation, and the production function approach.

Study	Data set	Identification	Firm-level elasticity	Market-level elasticity	Estimated markdown ^a
Arnold (2020)	US Longitudinal Business Database and Longitudinal Employer Household Dynamics (LEHD), 1999–2009	Mergers and concentration: Response of wages and employment to quasi-exogenous mergers and acquisitions	0.87 ^b	4–5 %	
Berger et al. (2022a)	US Census Longitudinal Business Database, 1977–2011.	Calibration and simulation: Calibrate differentiated jobs oligopsony model using reduced form estimates of labor supply elasticity based on variation in state-level corporate tax rates	2.6	39 % ^c	
Yeh et al. (2022)	US, The Census of Manufactures (CM) and the Annual Survey of Manufactures (ASM), 1976–2014.	Production function approach: Estimate the production function to recover the marginal revenue product of labor.	1.88	53 %	
Mertens (2022)	Administrative data on German manufacturing firms, 1995–2014.	Production function approach: Estimate the production function to recover the marginal revenue product of labor.		31 %–42 % ^d	
Brooks et al. (2021)	Annual Survey of Chinese Industrial Enterprises, 1999–2007; and India's Annual Survey of Industries, 1999–2011	Production function approach: Estimate markdown using production function approach. Regress markdown on market share to recover inverse market-level elasticity. ^e	2.2 ^f , 2.1 ^g	16 % ^f , 18 % ^g	

^aAll estimated markdowns are reported as percentages calculated as $(MRPL-w)/w$, where MRPL is marginal revenue product of labor and w is the wage.

^bMarket defined as four-digit industry and commuting zone.

^cThey report a markdown of $w/MRPL=0.72$, so this corresponds to 39 % markdown in our definition $(MRPL-w)/w$.

^dThey report the markdown as $MRPL/w$, and we converted it to $MRPL/w - 1 = (MRPL - w)/w$.

^eMarket defined as 4-digit industry province-level for China and state-level for India.

^fChina

^gIndia

elasticity of 7.7 for assistant professors, 3.0 for associate professors and 1.8 for full professors. This is evidence of monopsony power for tenure-track faculty, while the market for non-tenure track faculty is fairly competitive.

[Amadio and de Roux \(2023\)](#) add to the literature by estimating markdowns for a developing country, Colombia, focusing on the estimation of the firm-level labor supply elasticity for manufacturing firms. They estimate the impact of plausibly exogenous exchange rate shocks separately on average wages and on employment at the plant level. The ratio of these coefficients gives the elasticity of the inverse labor supply curve. Using this strategy, they find a firm-level labor supply elasticity of around 2.5, which corresponds to a markdown of about 40 %. Further, they find that the firm-level inverse labor supply elasticity is larger in firms with above median market share, consistent with the predictions of the oligopsony model (recall that, per [equation \(2\)](#) above, the inverse firm-level labor supply elasticity is s_j/η , where s_j is the firm's market share and η is the labor supply elasticity to the market).

The elasticity of labor supply and markdown estimates from the articles discussed in this section are summarized in [Table 4](#).

3.3 Labor market concentration

3.3.1 Concentration, market power and welfare

In this section, we elucidate the intricate relationships between concentration, wages, and overall welfare. Higher concentration is associated with greater market power, but it can also be associated with higher wages if it results from higher productivity, leading to ambiguous welfare effects. Additionally, the effects of concentration on welfare depend on the definition of the social welfare function, and whether one takes into account income inequality and positional externalities, or the potential impacts of market power on the political system.

In oligopsony models (with or without differentiated jobs), there is a positive relationship in equilibrium between the average markdown in a labor market and its concentration level (HHI). Since markdowns are typically not directly observable, empirical work often uses wages as the outcome variable. The variation in wages drives the variation in the markdown ($MRPL/w - 1$), but only if we can hold productivity $MRPL$ fixed. Further, when attempting to explain variation in wages, it is important to remember that factors other than concentration theoretically affect the equilibrium markdown. These factors include the market-level labor supply elasticity and the degree of substitutability between jobs (see [Table 1](#)). Therefore, even if, ceteris paribus, we expect a negative relationship in equilibrium between wages and concentration, this relationship may not hold empirically when we cannot sufficiently control for the relevant variables.

[Miller et al. \(2022\)](#) illustrate the concerns that industrial organization economists have with regressing prices or wages¹⁸ on concentration.¹⁹ They demonstrate through a theoretical model that higher concentration can coexist with lower prices when there are also productivity differences across markets. This insight is pivotal in understanding that the relationship between market concentration and price levels is not always positive, and therefore one may not find empirically that concentration increases prices. Consider two firms competing in a Cournot duopoly in market A, both with equal costs and market shares. These firms also compete in market C, where one firm enjoys lower costs due to higher productivity, thus gaining a larger market share. The result is a paradoxical scenario where market C, despite higher concentration, has lower prices than market A. This example underscores that higher productivity, leading to lower average costs, can result in lower prices even if it makes the market more concentrated. However, these lower prices do not imply that *market power* is reduced, nor that concentration is a poor measure of market power. Rather, the example highlights that market power, defined as the ability to set prices above marginal cost, can increase alongside consumer welfare if firms are more productive.²⁰

Not only can higher concentration be sometimes associated with lower prices, it can also be associated with higher wages and higher worker welfare. Recasting [Miller et al. \(2022\)](#)'s example within the labor market context reveals that increased labor market concentration can result in higher wages if it is the result of an increase in a firm's labor productivity. Further, *product* market concentration can also increase worker welfare in some situations. For example, [Boar and Midrigan \(2024\)](#) develop a model where the welfare-maximizing product market regulation simultaneously increases product market concentration, markups, *and* wages. This counterintuitive outcome arises because such regulation shifts production to the most productive firms, which are also the largest firms.

While market power does not invariably diminish welfare, it has the potential to do so, particularly when stemming from anticompetitive mergers (see [Section 4.1](#)).

¹⁸ Remember that wages are the price of labor services.

¹⁹ [Miller et al. \(2022\)](#) note that variation in concentration that arises due to mergers (or entry or exit events) can be informative about the effects of mergers on competition. Thus, to the extent that other confounding factors can be accounted for, a merger is a quasi-experiment that allows for evaluating the impact of HHI on the markdown.

²⁰ This example also highlights the potential for confusion arising from the label "pro-competitive" in antitrust jargon. The confusion arises from the definition of increased competition: is it lower prices or lower markups? In antitrust parlance, "pro-competitive" refers to a merger that leads to lower prices. Yet, if we assess competition based on markups rather than prices, then there are some "pro-competitive" mergers that are in fact anti-competitive, because they lower prices but increase the markup. This does not mean that markup-reducing pro-competitive mergers could not exist. For example, when two smaller firms merge, they may lower marginal costs and capture market share from a larger competitor, decreasing the equilibrium HHI and average markups. The term "pro-competitive" is arguably more aptly applied to such markup-reducing mergers. See [Section 4.1](#) below for a more extensive discussion of mergers and their impacts on the labor market.

Additionally, market power can affect the political system, potentially undermining welfare through various channels (Cowgill et al., 2023; Khan and Vaheesan, 2017; Wu, 2018; Zingales, 2017). Further, the utilitarian welfare function may not fully capture the broader impact of market power. For instance, even if wages increase, a reduction in the labor share due to heightened market power can erode workers' relative societal status, with negative welfare implications if positional externalities matter (Frank, 2005). Finally, understanding the nuances of labor market power is crucial for evaluating the effectiveness of economic policies, including minimum wage legislation (see Section 4.3) and unionization efforts, among others (Azar and Marinescu, 2024).

In summary, market power is important for reasons that go beyond its impact on prices and wages. To the extent that labor market concentration captures market power in the labor market, it can be a useful addition to the labor economist's toolbox.

3.3.2 *Measurement of labor market concentration*

Papers that focus on the impact of HHI on labor market outcomes must not only find plausibly exogenous variation in HHI, but also confront some additional choices:

- Define the labor market: this is necessary in order to calculate the HHI, since the HHI depends on *shares*. For example, some papers define the labor market using industries, and others using occupations.
- Decide on whether one uses data on employment levels (stocks), or on job vacancy or new hires to represent flows of newly available jobs. This choice mostly affects the level of the HHI, with vacancy flow-based HHIs being typically higher than employment stock-based HHIs because firms do not all have open vacancies during a given time period (Marinescu et al., 2021). The flow measurement better reflects market conditions that job seeking workers face at a point in time, but flow and stock measures are highly correlated across markets (Marinescu et al., 2021).

For example, Azar et al. (2020) calculate concentration measures for the US based on online job vacancies from the website BurningGlass (now Lightcast). An updated version of these concentration measures based on the Lightcast data is available for researchers to download (Choi and Marinescu, 2024). The data covers 2007–2021 (except 2008–2009) and is at the quarter by commuting zone by 6-digit SOC code level.

Because there are many ways of defining markets, labor market concentration may be a noisy measure of market power. For example, if the labor market is assumed to be the whole state while workers mostly look for jobs in their commuting zone (Marinescu and Rathelot, 2018), then firms' market shares will be systematically underestimated. The empirical literature and

antitrust practitioners have developed tools to deal with this challenge. We will return to market definition when discussing merger policy in [Section 4.1](#).

Another reason why the HHI may be a noisy measure of labor market power is that, in practice, firms can be linked through a network of interlocking shareholdings (including both cross-ownership and common ownership).²¹ The literature has shown that common ownership in particular has increased in recent decades, in part due to the rise of index funds ([Matvos and Ostrovsky, 2008](#); [Harford et al., 2011](#); [Azar, 2012](#); [Posner et al., 2016](#); [Posner and Weyl, 2018](#); [Backus et al., 2021](#); [Azar and Vives, 2021](#)).

The first step in quantifying the effect of common ownership for competition is to calculate the weight that each firm puts on the profits of other firms in its objective function.²² Given these objective function weights λ_{jk} , we can construct a modified HHI (MHHI), first proposed by [Bresnahan and Salop \(1986\)](#); [O'Brien and Salop \(1999\)](#), that quantifies the level concentration including both the market shares and the information from the interlocking shareholding network. In particular, in the case of common ownership, given the labor market shares s_j , the MHHI is:

$$MHHI = \sum_{j=1}^J s_j^2 + \sum_{j=1}^J \sum_{k \neq j} s_j s_k \lambda_{jk} \quad (77)$$

The first term is the HHI in the market, and the second term is the MHHI delta, which quantifies the increase in concentration in the market from common ownership. All of this analysis presupposes that shareholders are powerful enough that managerial objectives do not enter the objective function of the firm. If managers also have some power, then under some assumptions the formula for the MHHI delta is similar, except that there is a mitigation factor τ_j that reduces the shareholders' λ 's ([Azar and Ribeiro, 2022](#)). Thus, as shareholders become less powerful (for example, if highly dispersed), the MHHI tends towards the HHI, even if there is common ownership.

To sum up, labor market concentration can be used in some contexts as a proxy for market power, especially when researchers focus on explaining variation in labor market level outcomes.

²¹ Cross-ownership refers to direct ownership, by a firm that participates in a market, of shares in a competing firm. Common ownership refers to companies having shareholders in common.

²² We here give the formula for the weight that firm j puts on the profits of firm k in its objective function (relative to its own profits). When there are I shareholders indexed by i and β_{ij} denotes shareholder i 's percent ownership stake in firm j and γ_{ij} denotes shareholder i 's percent of control in firm j , the weight is:

$$\lambda_{jk} = \frac{\sum_{i=1}^I \gamma_{ij} \beta_{ik}}{\sum_{i=1}^I \gamma_{ij} \beta_{ij}} \quad (76)$$

For example, if firm j has only one shareholder, and that shareholder has a 30 % ownership stake in firm k , then the λ_{jk} weight is 0.3.

3.3.3 Empirical evidence on labor market concentration

The empirical literature shows that higher labor market concentration is associated with lower wages (Azar et al., 2020; Jarosch et al., 2019; Berger et al., 2022a). Focusing on mergers specifically, the literature shows that mergers that greatly increase labor market concentration decrease wages (Prager and Schmitt, 2021; Arnold, 2019). Prager and Schmitt (2021) focus on US hospital mergers and Arnold (2019) considers mergers in all industries in the US. Prager and Schmitt (2021) construct a measure of hospital concentration in each commuting zone (CZ) using data from the American Hospital Association (AHA) on hospital mergers and acquisitions. They use the HHI to capture the market share of each hospital system in a CZ. They find that mergers decrease wage growth for workers with specialized skills when concentration (HHI) increases greatly: for these specialized workers, mergers reduce wage growth by more than 25 %. By contrast, they find no effects for mergers that merge hospitals across CZs rather than within CZs, suggesting that effects are not due to firm-wide policies aimed at reducing labor costs, but rather to changes in labor market competition within a CZ. Arnold (2019) finds that mergers that greatly increase labor market concentration reduce earnings by more than 2 %. If the effect were driven by increased product market power and output reductions that lead to decreased labor demand, one would expect more negative wage effects for non-tradable industries whose product market is also local. However, Arnold (2019) finds no difference in effects between industries with tradable and non-tradable goods, implying that the adverse wage effect of mergers is not due to changes in product market power. Finally, Arnold (2019) uses the Cournot oligopsony model (see Section 2.2) together with the empirical estimates to calculate an implied market-level labor supply elasticity of 0.87, and an implied markdown of about 5 %. We report these estimates in Table 6, alongside estimates from other studies.

Merger effects in the labor market can help identify models that best fit the facts. In oligopsony models, mergers decrease both wages and employment. However, in search and matching models, mergers could decrease wages due to a worsening of outside options (e.g., Jarosch et al., 2019), but employment may not decrease (Rudanko, 2023). Prager and Schmitt (2021) finds that hospital mergers do not reduce employment even when they decrease wages, suggesting that a search and bargaining model may be more appropriate in their specific setting. By contrast, Arnold (2019) uses a sample of all US industries and finds a negative employment impact of mergers that greatly increase labor market concentration, which is compatible with the oligopsony model.

Azar et al. (2022) measure the level of concentration from common ownership in US labor markets and find that concentration has increased substantially over time. Using entries of firms into the S&P 500 index as a source of variation in labor market common ownership, they find that common ownership reduces wages, but increases employment. They argue that this

cannot be rationalized by a full-employment oligopsony model with an increasing labor supply, but could be explained by a search model of the labor market in which firms increase hiring efforts when wages are lower ([Rudanko, 2023](#); [Gottfries and Jarosch, 2023](#)).

3.4 Reduced-form approach based on workers' outside options

[Caldwell and Danieli \(2023\)](#) introduce the Outside Options Index (OOI), a measure of the jobs available to an individual worker based on the characteristics of the job and the worker. This index of outside options is derived using a matching framework, and resembles the measure derived in [Jarosch et al. \(2019\)](#) (see [Section 2.4.2](#)). While the OOI is similar to the HHI, it can be calculated for each individual worker instead of being calculated at the market level. To calculate the OOI, [Caldwell and Danieli \(2023\)](#) use a matched employer-employee dataset from Germany covering the years 1993–2014. They find that willingness to commute or move is an important factor in explaining variation in outside options; this result is consistent with the importance of geography as an element of job differentiation. To identify the impact of outside options on wages and employment, they compare the experiences of individuals involved in the same mass layoff who had different levels of the OOI (as calculated using pre-layoff characteristics). Relative to peers in the same layoff, a worker with one unit higher OOI (slightly more than one standard deviation) has 10 % higher earnings (as a share of pre-layoff earnings) one year after the separation, and is roughly 1.5 % more likely to be employed.

3.5 Calibration and simulation

[Berger et al. \(2022a\)](#) use US Census Longitudinal Business Database micro-data covering 1977–2011 to estimate reduced-form labor supply elasticities exploiting changes in state corporate taxes. The reduced-form labor supply elasticity is around two. They use these estimated labor supply elasticities to calibrate their oligopsony model with job differentiation (see [Section 2.3.2](#)) and firm differences in productivity. Using the calibrated model, they find an aggregate markdown (which they define as $w/MRPL$) of 0.72 (in our definition of the markdown as $(MRPL - w)/w$, the markdown is 39 %), corresponding to a labor supply elasticity of 2.6. Comparing steady states at an aggregate Frisch elasticity of labor supply of 0.50, they measure a welfare loss of 7.6 % and an output loss of 20.9 % relative to a perfectly competitive labor market. They find that roughly 60 % of welfare losses are driven by employment misallocation, as more productive firms keep their employment inefficiently low in order to exercise more monopsony power. 30 % of welfare losses are due to pure markdowns, and the remainder is due to the interaction of markdowns with misallocation. The estimates from this study are reported in summary [Table 6](#), alongside estimates from other approaches.

[Azar and Vives \(2021\)](#) calibrate a model of general equilibrium oligopoly with simultaneous product and labor market power, and with common ownership. In this model, the equilibrium markdown is a function of the labor market MHHI and the market-level elasticity of labor supply.²³ They find that the increase in common ownership between 1985 and 2017 implies a decline in the labor share that is similar in magnitude to the one observed over that period in the US.²⁴

[Jarosch et al. \(2019\)](#) use Austrian labor market data that covers the universe of private sector employment in Austria from 1997–2015. To get at the mechanism in their model (see [Section 2.4.2](#)), they measure “re-encounter” rates over a four-year period: taking transitions of workers who left firm a for firm b and then leave firm b as well, what share of these transitions are returns to firm a after leaving firm b ? Calibrating their model using these re-encounter rates, they find that eliminating size-based market power increases the labor share by about 1.3 to 2 % points. Since the labor share is around 2/3, this increase amounts to a 2.3 to 3 % increase in wages, depending on the year.

3.6 Structural estimation

[Azar et al. \(2022\)](#) estimate a nested logit model using job application data from Careerbuilder.com in the US in April-June 2012. They find direct evidence of substantial job differentiation. Geographic distance is an important component in job differentiation: 18 % of the variance in the total utility of a job is explained by the geographic distance (ZIP code to ZIP code) between the job and the job seeker. Without the use of instruments for wages, job applications appear very inelastic with respect to wages. Their preferred instrument for the wage is the predicted wage posted by the firm in markets other than the focal market. Using this instrument results in more elastic firm level application supply curves, with a firm level labor supply elasticity of 4.8. This implies a markdown of about 21 %. They define the labor market as an SOC-6 occupation by commuting zone, and find that the implied market level labor supply elasticity is 0.5. Considering variation in this elasticity across markets, they conclude that the market level elasticities are low enough to typically make an SOC-6 occupation by commuting zone a plausible definition of a relevant labor market for the purpose of antitrust analysis (also see below [Section 4.1](#) on the hypothetical monopsonist test).

[Lamadon et al. \(2022\)](#) use US matched employer-employee panel data together with business tax filings and worker-level filings for 2001–2015. They

²³ In this sentence, we use markdown as we have defined it in this chapter. In the paper, they define the “markdown” differently, in physical units, i.e., relative to the marginal physical product of labor instead of the marginal revenue product of labor.

²⁴ However, the decline in the labor share in the model would be lower if one added corporate governance frictions that reduced the MHHI, as discussed in [Section 3.3.2](#).

estimate a differentiated job model, and specifically a nested model with monopsonistic competition (see [Section 2.3.5](#)). The model incorporates two-sided heterogeneity, and heterogeneous worker preferences for amenities. To recover the firm level labor supply elasticity, they use instruments. They find a firm-level elasticity of 6.5, and a market-level elasticity of 4.6. Overall, their estimates imply that the marginal revenue product of labor is 15 % higher than the wage. Regarding job differentiation, they find that average workers would be willing to pay 13 % of their earnings to stay in their current jobs. Additionally, they show that more productive firms offer more amenities – consistent with [Sockin \(2021\)](#) –, and that better workers place a higher value on more amenities.

[Roussille and Scuderi \(2023\)](#) use data on full-time, high-wage engineering jobs, with a discrete choice model to calculate a wage markdown. Their data from Hired.com has a number of institutional features that they exploit to obtain rich estimates of both labor supply and demand without requiring an instrument. They focus on workers based in San Francisco, and estimate labor supply elasticities to the firm of 3.6 to 5.7. Under their preferred monopsonistic model of competition, the average predicted markdown is \$31,640 (or 19.5 % of productivity), which translates to a 25 % markdown under our definition of markdown with the wage in the denominator. Further, relevant to the sources of the markdown, they find significant evidence for job differentiation, with heterogeneous worker preferences over amenities.

[Kroft et al. \(2022\)](#) use matched employer-employee panel data for the US construction industry spanning 2001–2015, and data on procurement auctions. They compare first-time procurement contract winners to non-recipients who had never won contracts. They estimate a firm-specific labor supply elasticity of about 4.1, and that wages are marked down 25–29 % relative to the marginal revenue product of labor (using our definition of the markdown). Further taking into account imperfect competition in the product market, they find that wages are marked down 44–49 % relative to the *value* of the marginal product of labor (for a theoretical definition of this concept, see [Section 2.3.6](#)). In other terms, product market power and labor market power both significantly contribute to reducing wages relative to perfectly competitive labor *and* product markets.

[Table 5](#) synthesizes the estimates from the articles discussing structural estimation in this section.

3.7 Production function approach

The production function approach offers a distinct method for estimating the markdown “directly” by estimating the marginal revenue product and comparing it with the wage. This contrasts with other methods that rely on the labor supply elasticity. The production function approach can accommodate the possibility that firms do not always set profit-maximizing wages based on the

labor supply elasticity, due to factors like adjustment costs and fairness considerations. Consequently, the production function approach emerges as a valuable alternative to estimating the markdown by relying primarily on estimating the marginal revenue product of labor instead of the labor supply elasticity.

Note that the production function literature tends to define markups or markdowns as price over marginal cost, whereas the theoretical industrial organization (IO) literature tends to define it as price minus marginal cost over price (or, for markdowns, marginal product minus wage over wage). Therefore, the reader should be careful when comparing the formulas for the markdown and markup in this section with the formulas in [Section 2](#) above, as the latter follow the IO convention.

To understand how this production function method works, consider first the formula for the value of the marginal product of labor L (so the VMPL, not the marginal *revenue* product, see discussion in [Section 2.3.6](#)) divided by the wage, given a production function $F(L)$:

$$\frac{VMPL}{W} = \frac{P \frac{\partial F}{\partial L}}{W} \quad (78)$$

If we multiply both the numerator and the denominator by the quantity labor share of output L/Q (and we move the price P from the numerator to the inverse of P in the denominator), then we obtain a similar expression, commonly referred to as the “labor wedge”. This expression is in terms of the elasticity of output with respect to labor θ^L in the numerator, and the labor share α^L in the denominator.²⁵ The labor wedge is:

$$\frac{\frac{\partial F}{\partial L} \frac{L}{Q}}{\frac{W}{P} \frac{L}{Q}} = \frac{\theta^L}{\alpha^L} \quad (79)$$

The labor wedge uses VMPL, not MRPL, which makes a difference when firms have market power in both product and labor markets (see [Section 2.3.6](#)). When using MRPL, the formula for ν^L , the markdown for the labor input, is:

$$\nu^L \equiv \frac{MRPL}{W} = \left(1 + \frac{P'(F(L))F(L)}{P(F(L))}\right) \frac{P \frac{\partial F}{\partial L}}{W} = \frac{1}{\mu} \frac{\theta^L}{\alpha^L}, \quad (80)$$

where $\mu \equiv P/C'$ is price over marginal cost, i.e. the product market markup as defined in the production function literature.

²⁵ See above [equation \(42\)](#) for a similar formula we derived in the case where market power arises from product and job differentiation.

Therefore, the following holds for every factor of production i :

$$\nu^i \mu = \frac{\theta^i}{\alpha^i} \quad (81)$$

If there is an input M with no markdown (i.e., $\nu^M = 1$), then we can use that input's wedge to estimate the markup:

$$\mu = \frac{\theta^M}{\alpha^M} \quad (82)$$

Having estimated the markup μ , we can plug it into [equation \(80\)](#), to obtain the markdown for monopsonistic factor L :

$$\nu^L = \frac{\theta^L/\alpha^L}{\theta^M/\alpha^M} \quad (83)$$

[Yeh et al. \(2022\)](#) use this approach with data from the Census of Manufacturers (CM) and the Annual Survey of Manufactures (ASM) for 1976–2014 to estimate markdowns in the US manufacturing sector. They first identify the markup μ as in [equation \(82\)](#) based on the wedge for materials, a non-labor flexible input that is assumed not to be subject to monopsony power. They then combine this markup estimate with the labor elasticity of output and the labor share to obtain the markdown as in [equation \(83\)](#). They find an average markdown of 53 % across establishments.²⁶ The implied average firm-level labor supply elasticity is 1.88.

[Mertens \(2022\)](#) estimates markdowns and markups for manufacturing firms in Germany over the period 1995–2014 using the production function approach. The administrative dataset he uses contains data on both quantities and prices, which allows him to identify both the markup and the markdown (see discussion below page 41 about price data). He finds that, over 1995–2014, there was a small increase in markups (defined as price over marginal cost), from 1 to 1.03 and a large increase in markdowns (defined as marginal revenue product of labor over the wage), from 1.31 to 1.42, over this period. The 1.31–1.42 markdowns correspond to 31 %–42 % in our preferred definition of the markdown. The study concludes that rising market power can explain about half of the decline in the labor share for the German manufacturing sector, and that almost all of the increase in market power is due to labor rather than product market power.

[Brooks et al. \(2021\)](#) use the production function approach to estimate markdowns for manufacturing firms in China and India. To account for econometric issues in the estimation of the markup (see below), they use different estimation approaches, which end up giving similar markdown

²⁶The markdown for the average worker is higher because larger firms have higher markdowns (see their Figure 1).

estimates for Chinese and Indian firms. They find that the employment-weighted average markdown is 16 % in China and 18 % in India.²⁷ They also estimate the market-level inverse elasticities of labor supply by regressing the estimated markdowns on the firms' market shares (as in [equation \(2\)](#) from the oligopsony model discussed in [Section 2.2](#)), and find an elasticity of about 2.2 in both China and India.

A common criticism in the production function estimation literature is that output production functions estimated with revenue data (as opposed to quantity and price data, which are generally not available) are not always well-identified. [Bond et al. \(2021\)](#) argue that the production function approach fails to provide any information on the markup, because when calculated using the revenue elasticity, the markup estimator from [De Loecker and Warzynski \(2012\)](#) is equal to one when there are no input market frictions.²⁸ However, [Yeh et al. \(2022\)](#) show that, while this criticism applies to markup estimation, it does not affect markdown estimation, because the biases in the ratio of the labor elasticity and the materials elasticity (see [equation \(83\)](#)) cancel each other out.²⁹

One of the main assumptions in the production function approach to estimating markdowns is that there is a frictionless input. [Hashemi et al. \(2022\)](#) propose a direct method for estimating markdowns that bypasses the initial markup calculation, and therefore does not rely on the assumption that there is at least one frictionless input. They propose estimating markdowns based on estimation of the *revenue* production function and the Marginal Revenue Product of Labor (MRPL). For this approach to work, the econometrician needs data on the input's price by firm. In the case of labor, these prices are wages, and they are available in many widely used datasets for production function estimation. This makes labor monopsony estimation a potentially attractive application of this method.

The estimates from the articles discussed in this section utilizing the production function approach are summarized in [Table 6](#).

²⁷ The unweighted average markdowns are 3 % in China, and 1 % in India. Interestingly, the median markdown is negative in both countries.

²⁸ This is due to the downward bias pointed out by [Klette and Griliches \(1996\)](#). [Bond et al. \(2021\)](#) go a step further and show that this downward bias effectively renders the estimate uninformative about markups. In the presence of input market frictions, the derivation from [Bond et al. \(2021\)](#) shows that the markup estimator recovers the input market frictions (which explains why the literature does not, in fact, find a markup equal to one with this method).

²⁹ Other researchers have shown that, with more assumptions (such as constant returns to scale) and more sophisticated methods, one can recover the markup using revenue data ([Flynn et al., 2019; Kirov and Traina, 2021, 2023](#)). Both [Yeh et al. \(2022\)](#) and [Brooks et al. \(2021\)](#) implement a version of the constant-returns to scale approach to markup estimation.

3.8 Summary and discussion of markdown estimates

Tables 4, 5 and 6 present the summary of firm-level elasticities, market-level elasticities and markdowns. We always report our preferred estimation of the markdown as $(MRPL - w)/w$, adding footnotes when we converted the authors' reported markdown into our preferred format. For market-level elasticities, we report in footnotes how the market is defined.

Firm level elasticities of labor supply range between about 2 and 6, with the exception of a very low elasticity of 0.096 for MTurk workers, and a very high elasticity of 28.6 for non tenure-track faculty (Table 4). Market level elasticities of labor supply are less commonly estimated, and they range between 0.5 and 5.

The markdown is reported in most studies and typically ranges between about 15 % and 50 %, so that eliminating monopsony power would all other things equal increase wages by 15 to 50 %. Equivalently, we can say that wages are equal to between 67 % ($1/(1+0.5)$) and 87 % ($1/(1+0.15)$) of workers' marginal revenue product. There are a few estimates of the markdown that fall outside this range, with a low estimate of 4–5 % based on merger effects (Table 6), and a high estimate of 10.4 on MTurk (Table 4).

The fact that the two markdown estimates using the production function approach are not widely different from other estimates is mildly encouraging. Indeed, the production function approach estimates the marginal revenue product of labor, whereas most alternative estimates rely on the labor supply elasticity, alongside the assumption that firms optimize wages to maximize profits in light of the labor supply elasticity. These two kinds of approaches therefore rely on different core assumptions, and it would thus be useful for future research to confirm that the two approaches yield similar results.

Are the estimated markdowns using micro data too large to be reconciled with macro-level statistics such as the labor share, profit share, etc.? First, one needs to be careful when using the micro estimates in a macro context, as some estimates in the literature give more weight to large firms. This can occur either because of data availability or because of estimation constraints when firm fixed effects are used. If large firms have higher markdowns, as found by Yeh et al. (2022); Brooks et al. (2021), then micro-level markdowns overestimate the average markdown in the economy.

That said, the extent to which micro-level evidence can be reconciled with the aggregate statistics is still an open question. Consider, for example, the labor share, which, according to the BLS, has ranged from between 56 % and 64 % since 2000. As a back-of-the-envelope calculation, if the aggregate production function is Cobb-Douglas $Y = AK^\alpha L^{1-\alpha}$ with an $\alpha = 0.3$, the elasticity of labor supply to the firm is 5 (implying a markdown of 20 %), and the product market elasticity is 7 (implying a markup of 14 %), then the labor

share in the economy is 50%.³⁰ This is lower than the 56 % labor share, but arguably not completely out of the ballpark. Moreover, this is a back-of-the-envelope calculation that has several degrees of freedom and relies to some extent on assumptions about parameters for which there is no consensus in the literature. For example, if the product market elasticity for the aggregate economy were 20 instead of 7, then the labor share would be 55.4 %, much closer to the BLS numbers. Further, if the actual aggregate labor supply were higher than 5 (let's say because of underrepresentation of small firms in the existing estimates), the implied labor share would be even higher. We conclude that the question of whether the estimated markdowns are in line with macro aggregates remains unresolved in the literature, indicating that this issue is still very much open for further investigation.

While the empirical results clearly support the existence of monopsony power, one can also glean some insights about the empirical relevance of the different types of monopsony models. Specifically, what do the empirical estimates of the markdown tell us about which mechanisms of monopsony (see Table 3) are empirically relevant? There is empirical support for the oligopsony model and for the job differentiation model, while the empirical support for search is still developing. In the oligopsony model, a finite number of firms engage in strategic interactions and face a finite labor supply elasticity to the market. Evidence about concentration and mergers supports the strategic interaction aspect, while the relatively low estimated labor supply elasticity to the market supports the finite labor supply aspect. In the job differentiation model, jobs are imperfect substitutes. Evidence from nested logit models (Azar et al., 2022; Lamadon et al., 2022) and calibrated oligopsony models with job differentiation (Berger et al., 2022a) support the idea that jobs are imperfect substitutes. This differentiation also helps explain other empirical facts, such as the existence of wage dispersion. The empirical evidence for search as a mechanism behind monopsony power is less straightforward and more limited in the current state of the literature. Jarosch et al. (2019) calibrate a model with search, but this approach is not highly informative about whether search is important in explaining the markdown. Some literature finds that large firms

³⁰ Specifically, we calculate the labor share as

$$LS = (1 - \alpha) \times \frac{1 - \frac{1}{\sigma}}{1 + \frac{1}{\theta}} \quad (84)$$

where α is the exponent of capital in the Cobb-Douglas, θ is the elasticity of labor supply to the firm, and σ is the product market elasticity. To derive this equation, we start from the first-order condition of the firm, which is similar to equation (38), but using the marginal product of labor in the Cobb-Douglas production function, and assuming a representative firm so that its price is equal to the price level, and its wage is equal to the wage level in the economy in equilibrium.

Operating on that equation implies $\frac{W \times L}{P \times Y} = (1 - \alpha) \times \frac{1 - \frac{1}{\sigma}}{1 + \frac{1}{\theta}}$.

have higher markdowns (Yeh et al., 2022; Brooks et al., 2021), which is incompatible with simple search models but can be rationalized with more complex search models, e.g. with on the job search (see Section 2.5.1).

The literature does not currently allow us to rule out any of the models, but rather shows some support for all of them. This does not mean that every analysis must include oligopsony, job differentiation, and search. Instead, researchers should focus on the mechanism most relevant to their application, while considering whether other mechanisms might confound their results.

Overall, the empirical literature yields increasingly reliable estimates of monopsony power, with markdowns typically in the range of 15 % to 50 %, implying that completely eliminating monopsony power would increase wages by 15 to 50 %. However, it is important to recognize that real-world policy interventions would diminish, rather than eliminate, monopsony power, setting these figures as the theoretical maximum for wage increases achievable solely through monopsony power reduction. Nonetheless, the estimated markdowns indicate substantial scope for meticulously crafted policies to bolster wages by curtailing monopsony power.

4 Policy and monopsony power

In this section, we focus on three policies that have been carefully studied in connection with monopsony power: merger control in antitrust policy, non-competition agreements, and the minimum wage. We discussed these policies and other policies related to monopsony power in a policy-focused literature review (Azar and Marinescu, 2024).

4.1 Merger control

Antitrust authorities have the power to sue to block anticompetitive mergers. Merger review is fundamentally a prediction exercise, where antitrust authorities assess the risk that a merger will generate anticompetitive effects.

A merger of competing employers can substantially increase labor market power. Indeed, a merger increases the market share of the merging parties. For this reason, a merger is predicted to increase the markdown either in an oligopsony model (potentially with job differentiation, as in Berger et al. (2022a) or Azar et al. (2022)), or in a search and matching model that uses the HHI as a measure of market power (Jarosch et al., 2019). While an increase in productivity from the merger (commonly referred to as “efficiencies”) can theoretically offset the increase in market power (see Section 3.3.1), evidence shows that mergers that substantially increase labor market concentration reduce wages (Prager and Schmitt, 2021; Arnold, 2019). This suggests that efficiencies are, on average, not large enough to lead to an increase in wages or a decrease in the markdown.

The remainder of this section focuses on American antitrust policy, but similar approaches are used in merger enforcement by competition authorities

around the world. Until recently, US antitrust enforcement focused on competition in the product market, and thus mergers affecting the labor market received little to no attention ([Marinescu and Posner, 2019; Marinescu and Hovenkamp, 2019](#)). In a significant development, the US Department of Justice Antitrust Division successfully sued to block a publisher merger that would have reduced the compensation for labor. The merger between two major book publishers, Penguin-Random House and Simon & Schuster, was blocked in October 2022. The Antitrust Division predicted that authors would see lower pay as a result of the merger, because the merger would lead to reduced competition in bidding for authors' manuscripts. This was the first merger blocked in the US on the grounds of labor monopsony.

The US antitrust authorities – the Federal Trade Commission (FTC) and the Department of Justice Antitrust Division – periodically publish merger guidelines. The guidelines describe how agencies determine which mergers they will sue to block. They are designed to help the public, business community, practitioners, and courts understand the factors and frameworks the Agencies consider when investigating mergers ([Federal Trade Commission, 2022](#)).

While the merger guidelines mentioned the issue of buyer power as early as 1982 (in footnote 5), there was at first no explicit discussion of the power of employers as buyers of labor services. The 2010 guidelines expanded the discussion of monopsony with a whole section on powerful buyers. Finally, the 2023 Merger Guidelines ([U.S. Department of Justice & the Federal Trade Commission, 2023](#)) included Guideline 10, which is dedicated to the analysis of mergers that may substantially lessen competition for workers, creators, suppliers or other providers.

There are several tools that can be used to assess the effects of mergers on labor market power. In guidelines 1 through 6, the 2023 merger guidelines identify six distinct frameworks that can be used to assess whether a merger may substantially lessen competition.

Guideline 1 focuses on concentration, which is a key indicator that can be used by the antitrust authorities in all merger reviews, whether involving the labor market or other markets. A merger that significantly increases concentration in a highly concentrated market will attract the attention of US antitrust authorities (see Guideline 1 of the 2023 Merger Guidelines ([U.S. Department of Justice & the Federal Trade Commission, 2023](#))). Concentration is a useful indicator for a merger's likely effects on competition: a merger reduces direct competition between the merging firms (also see Guideline 2), and typically facilitates coordination (i.e. tacit or explicit collusion; covered in Guideline 3) among the remaining firms ([Baker and Farrell, 2020](#)).

For the product market, there is wide-ranging empirical evidence that mergers that increase concentration increase prices on average (e.g. [Kwoka, 2014; Bhattacharya et al., 2023](#)). For the labor market, we have mentioned parallel results above: mergers that increase labor market concentration tend to decrease wages ([Prager and Schmitt, 2021; Arnold, 2019](#)).

From a legal standpoint, a merger that significantly increases the HHI in a highly concentrated market leads to a “structural presumption”. The structural presumption shifts the burden of proof to the merging parties: unless sufficient contradictory evidence or rebuttal is provided by the merging firms, the antitrust authorities *presume* that such a merger is anticompetitive.³¹ Specifically, markets with an HHI greater than 1800 are highly concentrated, and a merger in a highly concentrated market that increases the HHI by more than 100 points is presumed to substantially lessen competition or tend to create a monopoly ([U.S. Department of Justice & the Federal Trade Commission, 2023](#)).³²

In order to calculate the HHI, it is necessary to define a market where *market shares* can be calculated. There is no one single market that is “correct”. Too broad a market would make sure that the market captures all relevant substitutes, but would include many options that are not reasonable substitutes. Too narrow a market would only include the very closest substitutes but would miss relevant competition from slightly more distant substitutes. Recognizing this continuum, the antitrust authorities do not define “the market” but “a relevant market” where the loss of competition stemming from a proposed merger can be well understood.

The 2023 Merger Guidelines offer tools to define a relevant *labor* antitrust market in Section 4.3.D.8. Broadly, a labor market is defined so that jobs are reasonably substitutable from the point of view of workers: “[d]epending on the occupation, alternative job opportunities might include the same occupation with alternative employers, or alternative occupations. Geographic market definition may involve considering workers’ willingness or ability to commute, including the availability of public transportation.” One tool to define a labor market based on this substitutability principle is the hypothetical monopsonist test: the test asks whether it would be profitable for a hypothetical monopsonist to undertake at least a small but significant (a 5 % threshold is often considered) and non-transitory decrease in wage or other worsening of terms or working conditions. When considering a change in wages or other terms, current market conditions are often taken as the benchmark, but other more competitive benchmarks can be used.³³

³¹ Rebuttal evidence can be provided by showing that the acquired firm is failing, that there will be firm entry or product repositioning, or that the merger brings efficiencies, i.e. higher productivity. In all cases, the merging parties must demonstrate that, based on the rebuttal evidence, the merger is not threatening a substantial reduction in competition ([U.S. Department of Justice & the Federal Trade Commission, 2023](#)).

³² The HHI thresholds have been revised downward in the 2023 guidelines relative to the 2010 guidelines, going back to the thresholds that were used before 2010, and that were established by the 1982 merger guidelines.

³³ When a market is already close to a monopsony, a hypothetical monopsonist may not find it profitable to impose a *further* 5 % decrease in wages because the dominant firm has already used most of the market power of a monopsonist. This means that, paradoxically, monopsony labor markets would fail the test. This is why the agencies can decide to use more competitive benchmarks than the current market conditions. This paradoxical situation is referred to in antitrust jargon as the *cellophane fallacy*.

When the market level labor supply is low enough, workers are unable to sufficiently substitute out of the market to defeat the monopsonist's strategy. In this case, the monopsonist finds it profitable to impose lower wages or worse terms, and the candidate labor market passes the test, and can be thought of as a relevant antitrust market.³⁴

While we have focused so far on the role of concentration and market definition in merger review, there are five additional core frameworks that agencies use to analyze mergers (Guidelines 2–6). Here, we briefly highlight how a merger affecting the labor market can be analyzed under Guidelines 2 or 3. Guideline 2 describes how a merger leads to the loss of head-to-head competition between the merging parties. In the case of the labor market, such “head-to-head” competition can be measured, for example, by the workers’ propensity to move from one of the merging firms to the other in response to changes in wage or working conditions. Guideline 3 describes the risk of coordination: after a merger, there are fewer employers in a labor market, which can make it easier for the remaining employers to coordinate, e.g. by agreeing on wages, or agreeing not to solicit each other’s employees.

To sum up, antitrust authorities have paid increasing attention to mergers that may reduce competition for labor. The US Department of Justice won its first case in this area in 2022, and the 2023 Merger Guidelines explain that anticompetitive effects in the labor market alone are sufficient for the antitrust authorities to sue to block a merger. All along, research in economics has informed this policy agenda.

4.2 Non-competition agreements

Non-compete clauses are legal agreements that prevent employees from joining competing firms within a specified time frame after their employment ends. Theoretically, non-competes can serve as a response to the employer’s holdup problem when investing in human capital (Hart and Moore, 1990). Employers may hesitate to provide extensive training or share sensitive information if employees can subsequently transfer this knowledge to a rival firm (Barron et al., 1999; Acemoglu and Pischke, 1999). Enforceable non-compete agreements mitigate this risk by ensuring that employees cannot immediately benefit a competitor with the training and insights they have gained (Rubin and Shedd, 1981; Posner et al., 2004). However, while non-competes increase incentives to provide general training, they can reduce incentives to provide firm-specific training (Meccheri, 2009). This is because under non-competes, employers are not worried about losing their workers to competitors and therefore can provide them with general training and skills applicable to all companies rather than just firm-specific training.

³⁴ Note that there can be more than one candidate market that passes this test.

Although non-competes can safeguard the employer's investment, they also influence the dynamics of labor market competition. These agreements can suppress wages and limit worker mobility (Krueger and Posner, 2018; Marx, 2018). When viewed through the search model framework, non-competes effectively narrow the pool of potential employers, thereby increasing market concentration and reducing the bargaining power of employees (Jarosch et al., 2019). Viewed through an oligopsony model, a firm with a non-compete becomes a monopsonist for its current employees. However, the impact of non-compete agreements varies depending on whether an individual is currently bound by such an agreement or is a prospective employee without existing restrictions. The former may face limited labor market options, while the latter encounters a more competitive environment. Nonetheless, wage equality constraints within firms – as in the search model by Rudanko (2023) (or in the oligopsony model) – can lead to adverse wage outcomes for all workers, regardless of their non-compete status.

Empirical research consistently shows that non-compete agreements tend to suppress wages. A significant focus of the empirical literature has been on the enforceability of non-competes. The term "enforceability" in this context means the likelihood of a court upholding a non-compete's restrictions on an employee's post-employment activities. The degree to which non-competes are enforceable varies by state law and individual case specifics. For example, California has outlawed non-competes, while other states have set certain restrictions or conditions for their application. Research by Starr (2019) indicates that stronger enforcement of non-competes led to increases in employer-provided training but also reduced wages by 4 %.

Non-compete agreements are surprisingly widespread across various job sectors in the United States. Starr et al. (2021) reveal that 18.1 % of U.S. workers are subject to non-competes, with these agreements being more common in high-skill occupations but also widespread in low-skill occupations. Only a small percentage of employees negotiate the terms of non-competes. The same research found that stricter enforcement is associated with 1.7 % to 2.5 % lower wages. However, non-competes offered before a job offer³⁵ tend to result in better wage outcomes, suggesting that workers may receive compensation for agreeing to these terms.

Non-compete agreements can disproportionately affect wages for certain demographic groups, and their impact can extend beyond individual workers to the labor market as a whole. Johnson et al. (2020) explored the impact of changes in state-level enforceability of non-competes, showing that such agreements disproportionately lower wages for women and non-white workers.

³⁵ 61 % of non-competes are presented before job offer and 29 % after, with the remainder being for promotions/raises, and respondents who couldn't remember the timing.

The study suggests that enforceability curtails workers' leverage to secure higher pay during favorable labor market conditions. Moreover, [Johnson et al. \(2020\)](#) indicate that strict non-compete enforcement in one state adversely affects earnings even for workers in other states within the same job market (for local labor markets that straddle state borders). This negative externality challenges the notion that workers consent to non-competes because they are mutually beneficial contracts.

Banning non-compete agreements can boost worker mobility and wages. [Balasubramanian et al. \(2022\)](#) investigated the consequences of Hawaii's non-compete ban on tech employees, finding a 4 % wage increase and an 11 % boost in job-switching activity. Overall, tech workers in Hawaii saw a 4.6 % cumulative earnings gain compared to those in states with moderate non-compete enforceability levels.

Non-compete agreements, while designed to protect business interests, can inadvertently stifle industry growth and innovation. Economic theory suggests that, in sectors where the high benefits of investments are less likely to be passed on to consumers, non-competes may prevent the emergence of spin-offs, to the detriment of consumers ([Lipsitz and Tremblay, 2021](#)). A policy analysis ([Shi, 2023](#)) extending the [Postel-Vinay and Robin \(2002\)](#) search and matching framework shows that limiting the duration of non-competes for top executives to roughly one month strikes an optimal balance. The calibration show this one-month limit is optimal, as it minimizes the negative impact on companies unable to recruit talent, while still encouraging investment by a worker's current employer. The research suggests that a quasi-ban on non-competes could be optimal, even for employees like executives, who are well-informed and who may indeed have access to knowledge that companies would seek to protect from competitors.

The empirical literature shows that non-competes can depress wages, with low-wage workers being disproportionately affected. These agreements shift the power balance in favor of employers, making it more challenging for employees to negotiate better pay. For low-wage positions, the justification for protecting training investments and proprietary information is weaker. Non-competes can further restrict career advancement and wage increases. Employers, particularly those hiring specialized or high-level staff, can often safeguard their interests post-employment through less restrictive means, such as confidentiality agreements and intellectual property laws. Given the availability of alternative protective measures, it appears that the typical employer places little importance on the enforceability of non-compete clauses ([Hiraiwa et al., 2023](#)).

In a significant policy development, the US Federal Trade Commission (FTC) issued a rule banning non competition agreements (with few exceptions); the rule is set to take effect in September 2024. According to the FTC, non-compete clauses constitute an unfair competitive practice under Section 5 of the FTC Act. In preparation for the rule, the FTC conducted a

comprehensive review of the empirical evidence on the effects of non-competes ([Federal Trade Commission, 2023](#)). This move indicates the FTC's efforts to address concerns related to employer monopsony power, and to promote more competitive labor markets.

4.3 Minimum wage

Economic theory predicts that competitive labor markets typically see a decrease in employment when a minimum wage is implemented. However, as explained above, employers with monopsony power pay workers less than the marginal revenue product of labor. Therefore, employers could afford to pay workers somewhat more. This implies that the effect of the minimum wage on employment is ambiguous ([Robinson, 1933; Manning, 2013,2021](#)): it depends on the degree of competition in the labor market, and on the level of the minimum wage relative to the marginal revenue product of labor.

To examine the employment effect of the minimum wage, we can use different models: an oligopsony framework (with or without job differentiation), or a random search model with posting and on the job search ([Burdett and Mortensen, 1998](#)). As explained in the theory [Section 2](#) above, under oligopsony, both wages and employment levels are suppressed compared to a perfectly competitive market. Firms aiming to maximize profits will set wages at a point where the additional cost of hiring one more employee (the wage of the new employee, plus the increase in the wage bill for all incumbents) equals the marginal revenue product. Introducing a minimum wage changes the optimal choice for the firm. When the minimum wage is set above the monopsonistic equilibrium but does not exceed the competitive minimum wage, the marginal cost of hiring an additional employee is simply the minimum wage. Firms can thus hire more workers without having to raise pay for existing employees, as they must already earn at least the minimum wage. This reduced marginal cost allows the monopsonist to hire more workers at the minimum wage, up to the point where the last worker's marginal revenue product equals the minimum wage ([Bhaskar and Manning, 2002](#)). To summarize, when the minimum wage increases while staying below the competitive level, employment increases in a monopsonistic labor market.

[Azar et al. \(2023\)](#) show that the negative impact of a local minimum wage on employment is mitigated – or can even turn positive – in more concentrated labor markets. This outcome can be predicted by a simple oligopsony model with Cournot competition, where the only difference between labor markets is the number of employers. Markets with fewer employers have a higher HHI and lower wages. When the minimum wage is set relative to the equilibrium wage prevailing before the policy change, the employment impact is more positive in markets with higher labor market concentration. This is because in more concentrated markets, markups are higher, leaving more scope to raise

wages without cutting employment. Empirical evidence from local minimum wage hikes in the US align with these predictions, suggesting that monopsony power may help account for the varying effects of minimum wage on employment found in the literature.

Minimum wage effects can also be modeled in a differentiated jobs framework, as in [Dustmann et al. \(2022\)](#). [Dustmann et al. \(2022\)](#) study the introduction of the national minimum wage in Germany in 2015. They assess the policy's impact using administrative data and quasi-exogenous variation in exposure across different German labor markets. They use high-wage workers as a control group for low-wage workers who were affected by the minimum wage. They find that the introduction of the minimum wage increased wages but did not decrease employment. The policy instead reallocated workers toward higher paying firms that are also larger and more desirable as measured by the poaching rank (i.e. how likely the firm is to be chosen over other firms among workers who have a job-to-job transition). The reallocation effect explains 17% of the wage increase induced by the introduction of the minimum wage. The differentiated jobs model – which is similar to the model in [Section 2.3.3–](#) explains the reallocation effect in a monopsony framework with heterogeneous firms: lower wage and less productive firms are unable to pay workers the new minimum wage and lose employment, while higher wage firms are induced to pay more, and therefore gain employment. The model predicts ambiguous wage and employment effects on the highest wage firms. Despite these ambiguous theoretical effects for the most productive firms, average worker welfare increases in the model as long as the minimum wage employment effect is not negative, which is the case that is empirically relevant in [Dustmann et al. \(2022\)](#).

Macroeconomic models suggest that minimum wages can increase social welfare through both efficiency gains and redistribution. [Berger et al. \(2022b\)](#) calibrate a general equilibrium oligopsony model. The calibration indicates that the welfare-maximizing minimum wage is \$15 per hour, with a range from \$0 to \$31 depending on the social welfare function weights. However, focusing solely on efficiency and ignoring benefits from redistribution, the optimal minimum wage drops to approximately \$8 per hour, and generates only marginal gains in efficiency.

In Japan, just as in the US, the extent of monopsony power across labor markets has been quantified using a production function approach. [Okudaira et al. \(2019\)](#) find a negative impact of minimum wages on employment in the more competitive labor markets where minimum wages approach the marginal product of labor. However, in markets with significant monopsony power, minimum wage increases do not affect employment levels.

A study examining Walmart's wage structure found that workers place substantial value on non-wage amenities. This preference is evident in workers' willingness to pay for "dignity at work," as measured using survey experiments ([Dube et al., 2022](#)). Following Walmart's corporate minimum

wage hike in 2014, there was no observed reduction in job amenities, suggesting that higher minimum wages may enhance the overall job value without necessitating a cutback in other benefits. This contradicts the traditional compensating differentials theory, which would predict a decrease in non-wage amenities in response to wage increases. The findings align more closely with a monopsonistic labor market perspective based on job differentiation, where wages and benefits are not highly substitutable from the employees' viewpoint.³⁶

The minimum wage has historically been a powerful tool for narrowing the earnings gap between black and white workers, particularly during the pivotal period of the late 1960s and early 1970s in the United States. However, if, as predicted by the competitive theory of the labor market, the minimum wage had disproportionately reduced black employment, the minimum wage's effect on overall racial inequality in the labor market would have been dampened. [Derenoncourt and Montialoux \(2021\)](#) examine the impact of the expansion of federal minimum wage laws (Fair Labor Standards Act) in 1966 to include industries with high shares of black workers, such as agriculture, restaurants, and personal services. They find that the racial earnings gap, when adjusted for measurable factors, was virtually eliminated within these industries post-reform, thus falling by 25 log points. Moreover, the research indicates that this policy did not adversely affect the employment of black workers, nor did it have a negative impact on overall employment levels. The findings are therefore consistent with the existence of monopsony power in the labor market, which allowed employers to increase wages without cutting employment.

This evidence underscores the potential of the minimum wage to serve as an effective mechanism for boosting earnings for the lowest-paid workers. The absence of pervasive negative employment consequences suggests that employers' monopsony power may play a role in these outcomes. In a purely competitive market, raising the minimum wage is predicted to lower employment. However, in labor markets with substantial monopsony power, an increase in the minimum wage may simultaneously boost employment and wages.

5 Conclusion

This chapter provides a comprehensive theoretical and empirical examination of monopsony power in labor markets. Our chapter summarizes recent theoretical contributions from three classes of models: oligopsony, job differentiation, and search and matching. The markdown is our preferred summary statistic for monopsony power: we define it as $(MRPL - w)/w$, the percent wage increase that would occur if monopsony power were eliminated, so that wages equal the marginal revenue product of labor ($w = MRPL$). We derive the

³⁶ However, Walmart's self-imposed wage increase may stem from a strategic business decision to boost overall compensation. Such a decision would not necessarily lead to a reduction in amenities, and firms' reaction to a minimum wage imposed by regulation may be different.

markdown for each theoretical framework and show that the average markdown often depends on labor market concentration, with higher concentration typically leading to higher markdowns in equilibrium. Beyond concentration, we also identify other determinants of market power, such as job differentiation and the labor supply elasticity, which are crucial for a comprehensive understanding of monopsony power.

The empirical literature has used different approaches to quantifying labor market power, using labor market concentration, the labor supply elasticity, and the production function approach. A first strand of the literature estimates the labor supply elasticity using exogenous or quasi-exogenous variation, and often proceeds to infer the markdown as the inverse of the labor supply elasticity. A second strand estimates the relationship between labor market concentration and wages, and must contend with productivity as a key potential confound; this literature typically finds that labor market concentration lowers wages. A third strand uses calibration and simulation to infer the markdown. And a fourth strand estimates the markdown using structural estimation or a production function approach. Across the overwhelming majority of studies, markdown estimates range from 15 % to 50 %. Most labor supply elasticity estimates to the firm fall between 2 and 6, and estimates of the labor supply elasticity to the market fall between 0.5 and 5, though market-level elasticities naturally depend on the definition of the market.

Given these substantial markdowns, we conclude that wages are not exclusively determined by marginal productivity. Therefore, variations in the markdown should also be considered as a source of wage variation, and researchers should be careful not to interpret the wage as a fully reliable measure of the marginal product of labor.

We then investigate policy implications of monopsony power, discussing merger control, non-competition agreements and the minimum wage. The literature suggests that mergers can lead to anticompetitive effects in labor markets similar to those in product markets. Monopsony power also helps reconcile the disparate estimates of the employment effects of the minimum wage: when there is enough monopsony power, a minimum wage increase may not reduce employment, and could in fact increase it.

While the literature has made strides in understanding monopsony power, there are certain limitations that future research could helpfully address. We outline three main limitations here. First, the treatment of non-wage amenities in the presence of monopsony power could be improved. The literature has generally assumed amenities are fixed and costless to firms; it would be interesting to examine how key conclusions are affected when detailed data on amenities is used, and when the endogeneity of firms' amenity choices is more fully addressed. Second, the empirical literature estimating the markdown could better integrate the estimation of labor and product market power, seeking a robust estimation of key parameters. Third, future research could help us more systematically understand which sources of monopsony power –

oligopsony, job differentiation, and search and matching frictions – are most quantitatively important in different contexts. This exploration would be helpful to determine which policy interventions are most likely to succeed in addressing the adverse effects of monopsony power.

Overall, the literature has progressed in a few ways since Manning's chapter in this same Handbook (Manning, 2011). First, it has developed new models of monopsony power, drawing both on the IO literature, and on labor economics' search and matching models. Second, the estimates of the markdown are getting more reliable. And third, the proliferation of new theories and empirical estimates has been accompanied by new insights on policy, showing that mergers and non-competition agreements may negatively affect wages by reducing competition for labor, and that moderate minimum wages may benefit workers by reducing the markdown and increasing employment. In essence, the recent surge in monopsony research has catalyzed a critical reevaluation of our understanding of labor markets, propelling the once-marginalized notion of imperfect competition to the forefront of economic analysis and policy.

References

- Abowd, J.M., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. *Econometrica* 67 (2), 251–333. <http://www.jstor.org/stable/2999586>.
- Acemoglu, D., Pischke, J.-S., 1999. The structure of wages and investment in general training. *Journal of Political Economy* 107 (3), 539–572. <https://doi.org/10.1086/250071>
- Amodio, F., de Roux, N., 2023. Measuring labor market power in developing countries: evidence from colombian plants. *Journal of Labor Economics*. <https://doi.org/10.1086/725248>. (Publisher: The University of Chicago Press).
- Anderson, S.P., De Palma, A., Thisse, J.-F., 1988. The CES and the logit: two related models of heterogeneity. *Regional Science and Urban Economics* 18 (1), 155–164.
- Arnold, D., 2019. Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes. SSRN Electronic Journal. <https://www.ssrn.com/abstract=3476369>.
- Azar, J., 2012. A new look at oligopoly: implicit collusion through portfolio diversification (Ph.D. thesis). Princeton University.
- Azar, J., Berry, S., Marinescu, I.E., 2019. Estimating Labor Market Power. SSRN Scholarly Paper ID 3456277, Social Science Research Network, Rochester, NY. <https://papers.ssrn.com/abstract=3456277>.
- Azar, J., Marinescu, I., Steinbaum, M., Taska, B., 2020. Concentration in US labor markets: evidence from online vacancy data. *Labour Economics* 66, 101886. <https://www.sciencedirect.com/science/article/pii/S0927537120300907>.
- Azar, J., Qiu, Y., Sojourner, A., 2021. Common Ownership Reduces Wages and Employment. SSRN Scholarly Paper ID 3954399, Social Science Research Network, Rochester, NY. <https://papers.ssrn.com/abstract=3954399>.
- Azar, J., Qiu, Y., Sojourner, A., 2022. Common ownership in labor markets. Available at SSRN 4158482.
- Azar, J., Ribeiro, R.M., 2022. Estimating oligopoly with shareholder voting models.
- Azar, J., Vives, X., 2021. General equilibrium oligopoly and ownership structure. *Econometrica* 89 (3), 999–1048.

- Azar, J., Huet-Vaughn, E., Marinescu, I., Taska, B., vonWachter, T., 2023. Minimum wage employment effects and labor market concentration. *The Review of Economic Studies*. <https://doi.org/10.1093/restud/rdad091>
- Azar, J., Marinescu, I., 2024. forthcoming. Monopsony Power in the Labor Market: From Theory to Policy. *Annual Review of Economics* 16.
- Azar, J.A., Berry, S.T., Marinescu, I., 2022. Estimating Labor Market Power. Working Paper. <https://www.nber.org/papers/w30365>.
- Backus, M., Conlon, C., Sinkinson, M., 2021. Common ownership in America: 1980–2017. *American Economic JournalMicroeconomics* 13 (3), 273–308.
- Baker, J., Farrell, J., 2020. Oligopoly Coordination, Economic Analysis, and the Prophylactic Role of Horizontal Merger Enforcement. *American University Washington College of Law Research Paper No. 2020–23* https://digitalcommons.wcl.american.edu/facsch_lawrev/1916.
- Balasubramanian, N., Chang, J.W., Sakakibara, M., Sivadasan, J., Starr, E., 2022. Locked in? The enforceability of covenants not to compete and the careers of high-tech workers. *Journal of Human Resources* 57 (S), S349–S396. <https://jhr.uwpress.org/content/57/S/S349> (Publisher: University of Wisconsin Press Section: Articles).
- Barron, J.M., Berger, M.C., Black, D.A., 1999. Do workers pay for on-the-job training? *The Journal of Human Resources* 34 (2), 235. <https://www.jstor.org/stable/146344?origin=crossref>.
- Bassier, I., Dube, A., Naidu, S., 2022. Monopsony in movers: the elasticity of labor supply to firm wage policies. *Journal of Human Resources* 57 (S), S50–S86. <https://doi.org/10.3368/jhr.monopsony.0319-10111R1>
- Berger, D., Herkenhoff, K., Mongey, S., 2019. Labor market power. Tech. Rep. w25719. (National Bureau of Economic Research, Cambridge, MA). <http://www.nber.org/papers/w25719.pdf>.
- Berger, D., Herkenhoff, K., Mongey, S., 2022a. Labor market power. *American Economic Review* 112 (4), 1147–1193. <https://doi.org/10.1257/aer.20191521>
- Berger, D.W., Herkenhoff, K.F., Mongey, S., 2022b. Minimum Wages, Efficiency and Welfare. Working Paper. <https://www.nber.org/papers/w29662>.
- Bertrand, J., 1883. Book review of *theorie mathematique de la richesse sociale* and of *recherches sur les principes mathématiques de la théorie des richesses*. *Journal de Savants* 67, 499–508.
- Bhaskar, V., Manning, A., To, T., 2002. Oligopsony and monopsonistic competition in labor markets. *Journal of Economic Perspectives* 16 (2), 155–174. <https://doi.org/10.1257/0895330027300>
- Bhattacharya, V., Illanes, G., Stillerman, D., 2023. Merger Effects and Antitrust Enforcement: Evidence from US Consumer Packaged Goods. <https://www.nber.org/papers/w31123>.
- Boal, W.M., Ransom, M.R., 1997. Monopsony in the labor market. *Journal of Economic Literature* 35 (1), 86–112. <http://www.jstor.org/stable/2729694>.
- Boar, C., Midrigan, V., 2024. Markups and Inequality. *Review of Economic Studies*.
- Bond, S., Hashemi, A., Kaplan, G., Zoch, P., 2021. Some unpleasant markup arithmetic: production function elasticities and their estimation from production data. *Journal of Monetary Economics* 121, 1–14.
- Bresnahan, T.F., Salop, S.C., 1986. Quantifying the competitive effects of production joint ventures. *International Journal of Industrial Organization* 4 (2), 155–175.
- Brooks, W.J., Kaboski, J.P., Li, Y.A., Qian, W., 2021. Exploitation of labor? Classical monopsony power and labor's share. *Journal of Development Economics* 150, 102627. <https://www.sciencedirect.com/science/article/pii/S0304387821000043>.
- Burdett, K., Mortensen, D.T., 1998. Wage differentials, employer size, and unemployment. *International Economic Review* 39 (2), 257–273. <https://www.jstor.org/stable/2527292>

- (Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University]).
- Caldwell, S., Danieli, O., 2023. Outside Options in the Labor Market. *Review of Economic Studies*.
- Caldwell, S., Oehlsen, E., 2022. Gender Differences in Labor Supply: Experimental Evidence from the Gig Economy.
- Card, D., 2022. Who set your wage? *American Economic Review* 112 (4), 1075–1090. <https://doi.org/10.1257/aer.112.4.1075>
- Card, D., Cardoso, A.R., Heining, J., Kline, P., 2018. Firms and labor market inequality: evidence and some theory. *Journal of Labor Economics* 36 (S1), S13–S70. <https://doi.org/10.1086/694153>. (Publisher: The University of Chicago Press).
- Choi, H., Marinescu, I., 2024. Data for labor market concentration using Lightcast (formerly Burning Glass Technologies). *Data in Brief* 55, 110647. <https://linkinghub.elsevier.com/retrieve/pii/S2352340924006140>.
- Cowgill, B., Prat, A., Valletti, T., 2023. Political Power and Market Power. <http://arxiv.org/abs/2106.13612>. ArXiv:2106.13612 [econ, q-fin].
- Dal, B., Finan, E.F., Rossi, M.A., 2013. Strengthening state capabilities: the role of financial incentives in the call to public service. *The Quarterly Journal of Economics* 128 (3), 1169–1218. <https://academic.oup.com/qje/article/128/3/1169/1849634>.
- Davis, S.J., Faberman, R.J., Haltiwanger, J.C., 2012. Recruiting intensity during and after the great recession: national and industry evidence. *American Economic Review* 102 (3), 584–588. <https://www.aeaweb.org/articles?id=10.1257>.
- Deb, S., Eeckhout, J., Patel, A., Warren, L., 2022. What drives wage stagnation: monopsony or monopoly? *Journal of the European Economic Association* 20 (6), 2181–2225. <https://doi.org/10.1093/jeea/jvac060>
- Derenoncourt, E., Montialoux, C., 2021. Minimum wages and racial inequality. *The Quarterly Journal of Economics* 136 (1), 169–228. <https://doi.org/10.1093/qje/qjaa031>
- Dube, A., Jacobs, J., Naidu, S., Suri, S., 2020. Monopsony in online labor markets. *American Economic Review: Insights* 2 (1), 33–46. <https://doi.org/10.1257/aeri.20180150>
- Dube, A., Naidu, S., Reich, A.D., 2022. Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers. Working Paper 30441, National Bureau of Economic Research. <http://www.nber.org/papers/w30441>.
- Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., vom Berge, P., 2022. Reallocation effects of the minimum wage. *The Quarterly Journal of Economics* 137 (1), 267–328. <https://doi.org/10.1093/qje/qjab028>
- Federal Trade Commission, 2022. Federal Trade Commission and Justice Department Seek to Strengthen Enforcement Against Illegal Mergers. <https://www.ftc.gov/news-events/news/press-releases/2022/01/federal-trade-commission-justice-department-seek-strengthen-enforcement-against-illegal-mergers>.
- Federal Trade Commission, 2023. Non-Compete Clause Rule NPRM. Federal Register Notice 88 Fed. Reg. 3482, Federal Trade Commission. <https://www.federalregister.gov/documents/2023/01/19/2023-00414/non-compete-clause-rule>.
- Flynn, Z., Traina, J., Gandhi, A., 2019. Measuring markups with production data. Available at SSRN 3358472.
- Frank, R.H., 2005. Positional Externalities Cause Large and Preventable Welfare Losses. *The American Economic Review* 95(2): 137–141. <https://www.jstor.org/stable/4132805>. Publisher: American Economic Association.

- Gabszewicz, J.J., Vial, J.-P., 1972. Oligopoly "a la Cournot" in a general equilibrium analysis. *Journal of Economic Theory* 4 (3), 381–400.
- Gabszewicz, J.J., Vial, J.-P., 2023. Monopsony power in higher education: a tale of two tracks. *Journal of Labor Economics* 41 (S1), S257–S290. <https://www.journals.uchicago.edu/doi/abs/10.1086/726720> (Publisher: The University of Chicago Press.).
- Gottfries, A. , Jarosch, G., 2023. Dynamic monopsony with large firms and noncompetes.Tech. rep., National Bureau of Economic Research.
- Harford, J., Jenter, D., Li, K., 2011. Institutional cross-holdings and their effect on acquisition decisions. *Journal of Financial Economics* 99 (1), 27–39.
- Hart, O., Moore, J., 1990. Property rights and the nature of the firm. *Journal of Political Economy* 98 (6), 1119–1158.
- Hashemi, A., Kirov, I., Traina, J., 2022. The production approach to markup estimation often measures input distortions. *Economics Letters* 217, 110673.
- Hiraiwa, T., Lipsitz, M., Starr, E., 2023. Do firms value court enforceability of noncompete agreements? A revealed preference approach. <https://papers.ssrn.com/abstract=4364674>.
- Jarosch, G., Nimczik, J.S., Sorkin, I., 2019. Granular Search, Market Structure, and Wages. Tech. Rep. w26239. (National Bureau of Economic Research). <https://www.nber.org/papers/w26239>.
- Johnson, M.S., Lavetti, K., Lipsitz, M., 2020. The Labor Market Effects of Legal Restrictions on Worker Mobility. <https://papers.ssrn.com/abstract=3455381>.
- Khan, L.M., Vaheesan, S., 2017. Market power and inequality: the antitrust counterrevolution and its discontents. *Policy Review* 11.
- Kirov, I., Traina, J., 2021. Measuring markups with revenue data.Available at SSRN 3912966.
- Kirov, I., Traina, J., 2023. Labor market power and technological change in US manufacturing. Tech. rep., Working Paper.
- Klette, T.J., Griliches, Z., 1996. The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of applied econometrics* 11 (4), 343–361.
- Kroft, K., Luo, Y., Mogstad, M., Setzler, B., 2022. Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry.
- Krueger, A.B., Posner, E.A., 2018. A Proposal for Protecting Low-Income Workers from Monopsony and Collusion. Policy Proposal 2018–05, The Hamilton Project. https://www.emfink.net/EmploymentLaw/assets/materials/Confidentiality_Competition/Krueger-Posner.pdf.
- Kwoka, J., 2014. Mergers, Merger Control, and Remedies: A Retrospective Analysis of U.S. Policy. MIT Press, (Google-Books-ID: qcYQBgAAQBAJ).
- Lamadon, T., Mogstad, M., Setzler, B., 2019. Imperfect Competition, Compensating Differentials and Rent Sharing in the U.S. Labor Market. Working Paper 25954, National Bureau of Economic Research. <https://www.nber.org/papers/w25954>. Series: Working Paper Series.
- Lamadon, T., Mogstad, M., Setzler, B., 2022. Imperfect competition, compensating differentials, and rent sharing in the US labor market. *American Economic Review* 112 (1), 169–212. <https://doi.org/10.1257/aer.20190790>
- Lavetti, K., 2023. Compensating wage differentials in labor markets: empirical challenges and applications. *Journal of Economic Perspectives* 37 (3), 189–212. <https://doi.org/10.1257/jep.37.3.189>
- Lipsitz, M., Tremblay, M.J., 2021. Noncompete agreements and the welfare of consumers. Available at SSRN 3975864.
- De Loecker, J., Warzynski, F., 2012. Markups and firm-level export status. *American economic review* 102 (6), 2437–2471.

- Manning, A., 2011. Chapter 11 - Imperfect competition in the labor market. *Handbook of Labor Economics*, vol. Volume 4, Part B. Elsevier, pp. 973–1041. <http://www.sciencedirect.com/science/article/pii/S0169721811024099>.
- Manning, A., 2013. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.
- Manning, A., 2021. Monopsony in labor markets: a review. *ILR Review* 74, pp. 3–26. <https://doi.org/10.1177/0019793920922499>. (Publisher: SAGE Publications Inc).
- Marinescu, I., Hovenkamp, H., 2019. Anticompetitive mergers in labor markets. *Indiana Law Journal* 94 (3), 1031. <https://www.repository.law.indiana.edu/ilj/vol94/iss3/5>.
- Marinescu, I., Ouss, I., Pape, L.-D., 2021. Wages, hires, and labor market concentration. *Journal of Economic Behavior & Organization* 184, 506–605. <https://www.sciencedirect.com/science/article/pii/S0167268121000548>.
- Marinescu, I., Rathelot, R., 2018. Mismatch Unemployment and the geography of job search. *American Economic Journal: Macroeconomics* 10 (3), 42–70. <https://doi.org/10.1257/mac.20160312>
- Marinescu, I.E., Posner, E.A., 2019. Why has antitrust law failed workers? *SSRN Electronic Journal*. <https://www.ssrn.com/abstract=3335174>.
- Marx, M., 2018. Reforming Non-Competes to Support Workers. Policy Proposal 2018–04, The Hamilton Project. https://www.hamiltonproject.org/assets/files/reforming_noncompetes_support_workers_marx_policy_proposal.pdf.
- Matvos, G., Ostrovsky, M., 2008. Cross-ownership, returns, and voting in mergers. *Journal of Financial Economics* 89 (3), 391–403.
- Meccheri, N., 2009. A note on noncompetes, bargaining and training by firms. *Economics Letters* 102 (3), 198–200. <https://www.sciencedirect.com/science/article/pii/S0165176509000172>.
- Mertens, M., 2022. Micro-mechanisms behind declining labor shares: Rising market power and changing modes of production. *International Journal of Industrial Organization* 81, 102808.
- Miller, N., Berry, S., Morton, F.S., Baker, J., Bresnahan, T., Gaynor, M., Gilbert, R., Hay, G., Jin, G., Kobayashi, B., Lafontaine, F., Levinsohn, J., Marx, L., Mayo, J., Nevo, A., Pakes, A., Rose, N., Rubinfeld, D., Salop, S., Schwartz, M., Seim, K., Shapiro, C., Shelanski, H., Sibley, D., Sweeting, A., Wosinska, M., 2022. On the misuse of regressions of price on the HHI in merger review. *Journal of Antitrust Enforcement* 10 (2), 248–259. <https://doi.org/10.1093/jaenfo/jnac009>
- Moen, E.R., 1997. Competitive search equilibrium. *Journal of Political Economy* 105 (2), 385–411. <https://doi.org/10.1086/262077>. (Publisher: The University of Chicago Press).
- O'brien, D.P., Salop, S.C., 1999. Competitive effects of partial ownership: financial interest and corporate control. *Antitrust LJ* 67, 559.
- OIRA, OMB, 2023. Guidance on Accounting for Competition Effects When Developing and Analyzing Regulatory Actions.
- Okudaira, H., Takizawa, M., Yamanouchi, K., 2019. Minimum wage effects across heterogeneous markets. *Labour Economics* 59, 110–122.
- Posner, E., Weyl, E., 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.
- Posner, E.A., Scott Morgan, F.M., Weyl, E.G., 2016. A proposal to limit the anticompetitive power of institutional investors. *Antitrust LJ* 81, 669.
- Posner, E.A., Triantis, G.G., Triantis, A.J., 2004. Investing in Human Capital: The Efficiency Of Covenants Not To Compete. University of Chicago Law & Economics, Oline Working Paper No. 137, 1–35. <https://papers.ssrn.com/abstract=285805>.

- Postel-Vinay, F., Robin, J.-M., 2002. The distribution of earnings in an equilibrium search model with state-dependent offers and counteroffers. *International Economic Review* 43 (4), 989–1016. <https://doi.org/10.1111/1468-2354.t01-1-00045/abstract>
- Prager, E., Schmitt, M., 2021. Employer consolidation and wages: evidence from hospitals. *American Economic Review* 111 (2), 397–427. <https://doi.org/10.1257/aer.20190690>
- Robinson, J., 1933. Imperfect competition. 1966 In: Robinson, J. (Ed.), *An Essay on Marxian Economics*. Palgrave MacMillan, UK, pp. 73–81. https://doi.org/10.1007/978-1-349-15228-5_9. 1966.
- Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy* 82 (1), 34–55. <https://doi.org/10.1086/260169>
- Rosen, S., 1986. Chapter 12: The theory of equalizing differences. In: O. Ashenfelter and R. Layard, eds. *Handbook of Labor Economics*, Vol. 1. Elsevier, pp. 641–692. <https://www.sciencedirect.com/science/article/pii/S1573446386010155>
- Roussille, N., Scuderi, B., 2023. Bidding for Talent: Equilibrium Wage Dispersion on a High-Wage Online Job Board.
- Rubin, P.H., Shedd, P., 1981. Human capital and covenants not to compete. *The Journal of Legal Studies* 10 (1), 93–110. <https://doi.org/10.1086/467672>
- Rudanko, L., 2023. Firm wages in a frictional labor market. *American Economic Journal: Macroeconomics* 15 (1), 517–550. <https://doi.org/10.1257/mac.20200440>
- Shi, L., 2023. Optimal regulation of noncompete contracts. *Econometrica* 91 (2), 425–463. <https://doi.org/10.3982/ECTA18128>
- Sockin, J., 2021. Show Me the Amenity: Are Higher-Paying Firms Better All Around <https://papers.ssrn.com/abstract=3957002>.
- Sokolova, A., Sorensen, T., 2021. Monopsony in labor markets: a meta-analysis. *ILR Review* 74 (1), 27–55. <https://doi.org/10.1177/0019793920965562>
- Starr, E., 2019. Consider this: training, wages, and the enforceability of covenants not to compete. *ILR Review* 72 (4), 783–817. <https://doi.org/10.1177/0019793919826060>
- Starr, E.P., Prescott, J.J., Bishara, N.D., 2021. Noncompete agreements in the US labor force. *The Journal of Law and Economics* 64 (1), 53–84. <https://doi.org/10.1086/712206>
- U.S. Department of Justice and the Federal Trade Commission, 2023. Draft FTC-DOJ Merger Guidelines for Public Comment. Draft - For Public Comment Purposes FTC-2023-0043-0001. https://www.ftc.gov/system/files/ftc_gov/pdf/p859910draftmergerguidelines2023.pdf.
- Verboven, F., 1996. The nested logit model and representative consumer theory. *Economics Letters* 50 (1), 57–63.
- Wu, T., 2018. The Curse of Bigness: Antitrust in the New Gilded Age. *Columbia Global Reports*. <https://www.jstor.org/stable/j.ctv1fx4h9c>.
- Yeh, C., Macaluso, C., Hershbein, B., 2022. Monopsony in the US labor market. *American Economic Review* 112 (7), 2099–2138. <https://doi.org/10.1257/aer.20200025>
- Zingales, L., 2017. Towards a political theory of the firm. *The Journal of Economic Perspectives* 31 (3), 113–130. <https://www.jstor.org/stable/44321282> (Publisher: American Economic Association).

This page intentionally left blank

Index

Note: Page numbers followed by “f” indicate figures.

A

- Abadie’s κ , 48
Accounting for cycles, 171
Active labor market policies (ALMP), 436, 437, 541
advances in labor market design on online search platforms, 555–557
discussion of cost effectiveness, 560–563
focus on programs for special groups, 546–550
growing awareness of spillover or displacement effects, 558–560
internationalization of ALMP use and evaluations, 552–555
mechanisms explaining program effects, 568
meta analysis studies, 542–544
new insights in role of caseworkers, 544–546
novel identification strategies, 568–570
program design takes demand side into account, 550–552
wide range of outcome variables, 563–567
Active spending Anglican countries, 440f
Active spending EU countries, 440f
Adding covariates, 201
Administrative data, 459
AEA. *See* American Economic Association
Age, 307
AHA. *See* American Hospital Association
AKM model, 116–117, 125
causality, 140–146
edgy interpretation of firm effects, 127–135
evaluating AKM restrictions, 135–140
AKM regression, 797
AKM restrictions, 135
accounting for noise, 136–140
visualizing goodness of fit, 136
ALMP. *See* Active labor market policies
Amazon Mechanical Turk experiments, 795
Amenities, 124–125, 323
employer-provided health insurance, 323–324

- indirect evidence on, 324
on-the-job training, 324
American Economic Association (AEA), 620
American Hospital Association (AHA), 803
Analogous estimators, 49
Annual Survey of Manufactures (ASM), 808
ASM. *See* Annual Survey of Manufactures
ATE. *See* Average treatment effect
Average causal response (ACR), 28, 29, 53
Average monotonicity, 23
Average treatment effect (ATE), 13
Average treatment on the treated (ATT), 55

B

- Back-of-the-envelope calculation, 647
Baily-Chetty approach, 437, 535
Baily-Chetty formula, 506, 507
Ban the Box policies, 703
Baseline framework, 498–500
Bayes’ rule, 229
BCP. *See* Border county pairs
Becker, Gary, 680
Behavioral costs, 413, 503
of UI benefit levels, 506
of UI PBD, 507
Bertrand competition, 770, 774
Bertrand paradox, 768, 769
Bias-corrected variance estimation, 197, 217–221
Bias correction, 148–155
Binary instrument, 16–18
Binary threshold-crossing model, 27
Binary treatment, 16–18
Binomial mixtures, 209–210
Bonferroni correction, 228
Bootstrapping, 67
Border county pairs (BCP), 289–290
Border discontinuity, 289–291
Boston Public Schools (BPS) district, 213
Bounding and imputation, 151–152
Brazilian estimates, 160
Bridge, 135
Burdett-Mortensen model, 781, 789

C

- Calibration and simulation, 804–805
 California UI system, 415
 Callaway-Sant'Anna estimators, 363
 Call-back rates for workers with criminal records, 704
 Canada Emergency Wage Subsidy (CEWS), 417
 Canadian Child Benefit Expansions, 609
 Cardinal treatments, 27–29
 Career-family trade-offs, 639–644
 Cash and near-cash transfers, 590, 595, 608
 Causality, 140
 indirect contrasts and spanning trees, 142–144
 restricting selection, 144–146
 Census of Manufactures (CM), 808
 CES model, 775
 CEWS. *See* Canada Emergency Wage Subsidy
 Chebyshev's inequality, 147
 Chicago program, 695
 Child outcomes, 606
 public policies and, 606–609
 Child penalty, 641–642
 Cholesky factorization, 132
 Classical consumption-based approach, 507–508
 Classical IV model, 10, 11
 Classical linear IV model, 2
 Clean controls, 369
 Clustering approaches, 156–158
 Clustering methods, 117
 CM. *See* Census of Manufactures
 Cobb-Douglas preference, 634
 Cohabitation. *See* Marriage, divorce, and cohabitation
 Combining estimators, 203–204
 Commuting zone (CZ), 803
 Composition of jobs, 308
 Compound decision problems, 184, 195–196
 Compound decisions and shrinkage strategies, 233–235
 Conditional effect ignorability (CEI), 45
 Conditional location scale estimators (CLOSE), 208
 Conjugate gradient (CG) methods, 132
 Connected at random (CAR), 154
 Constant relative risk aversion (CRRA), 507
 Consumer demand, 339–340
 Consumers reactions, 330–332
 Continuous covariate variation, 65
 Continuous treatments, 76–78
 Correlated random effects (CRE) approach, 158
 Correlation coefficient, 146
 Cost-benefit calculation, 682
 Cournot competition, 770, 774, 818
 Cournot duopoly, 800
 Cournot oligopsony model, 769, 771
 Covariance between person and firm effects, 172–175
 Covariates, 11, 34
 controlling for, 288–289
 linearly, 34–35
 nonparametrically, 34
 estimating LATEs and ACRs in presence of, 46
 double robustness and machine learning, 49–50
 empirical illustration, 50–53
 propensity score weighting, 46–49
 level-dependence caused by, 35–36
 methods of controlling for, 51t
 monotonicity-correct first stage specifications, 37–38
 specification considerations with, 38–39
 weighting expression for linear IV under rich covariates, 36–37
 Covid-19 crisis, 421–422
 Covid-19 pandemic, 386, 388, 417–418, 441, 611, 647
 CPS. *See* Current Population Survey
 Credibility revolution, 680, 706
 Crime and labor market, 679, 681f
 descriptive statistics and stylized facts, 683–686, 684f, 685f
 earnings and prices, 696–698
 education and crime, 705, 706f, 708t–739t, 740–741
 education, crime impacts on, 743–744
 education policies, crime and, 744
 effects of a record on labor market outcomes, 700–702
 firm willingness to hire workers with criminal records, 703
 future directions, 744–750
 gangs and organized crime, 748–750
 incapacitation, 741–742
 policies to improve labor outcomes for workers with criminal records, 703–705
 schooling quantity and quality, 742–743
 unemployment, 689–694
 victimization, 744–748

- wages and income, 687–689
youth labor markets, 694–696
- Criminal records, 698
 effects on labor market outcomes, 700–702
 firm willingness to hire workers with, 703
 policies to improve labor outcomes for workers with, 703–705
- Cross-fitting, 148–155
- Cross-fit variance estimator, 150
- Current Population Survey (CPS), 323
- Cycle space, 129
- CZ. *See* Commuting zone
- D**
- DARA. *See* Decreasing absolute risk-aversion
- Data-driven machine learning techniques, 52
- Decision rules, 204–206
- Deconvolution methods, 185, 188–190, 235
- Decreasing absolute risk-aversion (DARA), 511
- Demographic prediction-based approach, 292, 294–295
- Department of Justice Antitrust Division, 813
- Diagonal weighting matrix, 131
- Diamond-Mortensen-Pissarides wage bargaining framework, 784
- Difference-in-differences (DiD) event, 362
- Differential job sorting and organization of work, 646–649
- Differentiated jobs, 769
 monopsonistic competition and oligopsony with discrete choice, 773–776
 monopsonistic competition, nesting with, 777–778
 monopsonistic competition with, 770–772
 nested logit, 776–777
 oligopsony, 772–773
 simultaneous labor and product market power, 778–781
- Directed search model, monopsony power in, 787–791
- Disability Insurance (DI), 534
- Discordance aversion, 232
- Dispersion, 118, 171
- Displacement, 123–124
- Distributional-implications, 341–348, 344–345
- Diversion, 700–701
- Divorce. *See* Marriage, divorce, and cohabitation
- Dixit-Stiglitz price index, 779
- Double/debiased machine learning (DDML), 50
- Double robustness, 49–50
- Downstream effects, 341–348
- Downstream socioeconomic outcomes, 263
- Dual wage ladder (DWL) model, 167
- Dunlop, John, 269
- Dynamic framework, 500–503
- E**
- Earned Income Tax Credit (EITC), 312, 595, 600, 660
- Earnings accelerator feature, 795
- Earnings and prices, 696–698
- EB methods. *See* Empirical Bayes methods
- ECM. *See* Error-correction model
- Economic thinkers, 268
- Economists, 119
- Edge effects, 127
- Edgy interpretation of firm effects, 127
 combination weights and smoothing, 132–135
 estimators, 131–132
 as restricted edge effects, 129–131
- Education and crime, 705, 706f, 708t–739t
 causal impacts of, 740–741
 crime and education policies, 744
 crime impacts on education, 743–744
 incapacitation, 741–742
 schooling quantity and quality, 742–743
- EITC. *See* Earned Income Tax Credit
- Elasticity of labor supply, 793
 experiments, 795–797
 identification, forms of, 797–799
- Emergency Unemployment Compensation (EUC) programs, 505
- Empirical Bayes (EB) methods, 184
 connections to machine learning, 211–213
 decision rules, 204–206
 empirical Bayes recipe, 186
 deconvolution, 188–190
 shrinkage, 190–191
 three-step EB recipe, 191–192
 gains from shrinkage, 192
 compound decision problems, 195–196
 James/Stein theorem, 193–195
 MSE improvements in normal/normal model, 192–193
 generalizations of linear shrinkage, 201
 adding covariates, 201
 combining estimators, 203–204
 multivariate EB, 201–202
- linear shrinkage application, 213–216
- non-parametric empirical Bayes, 216

- bias-corrected variance estimation, 217–221
 compound decisions and shrinkage strategies, 233–235
 firm-level labor market discrimination, 235–255
 for multiple testing, 228–231
 non-parametric priors and posteriors, 221–225
 partial identification, 225–228
 ranking problems, 231–233
 practical shrinkage issues, 196
 correlated vs. uncorrelated random effects, 199–200
 distributions of true parameters, unbiased estimates, and posterior means, 196–197
 long regression or mean residuals, 199–200
 shrinkage and regression, 197–199
 precision-dependence, 206
 noisy standard errors, 210–211
 testing and modeling precision-dependence, 208–209
 variance-stabilizing transformations, 209–210
 Empirical Bayes recipe, 186
 deconvolution, 188–190
 shrinkage, 190–191
 three-step EB recipe, 191–192
 Empirical Bayes shrinkage, 185
 Empirical evaluations, 342
 Empirical example, 152–155
 Empirical illustration, 50–53
 Empirical literature, definition of markdown in, 792–793, 794t, 796t, 798t
 Empirically measuring monopsony power
 calibration and simulation, 804–805
 definition of the markdown in the empirical literature, 792–793, 794t, 796t, 798t
 elasticity of labor supply, 793–799
 labor market concentration, 799–804
 production function approach, 806–809
 reduced-form approach based on workers' outside options, 804
 structural estimation, 805–806
 summary and discussion of markdown estimates, 810–812
 Employer-employee panel data, 806
 Employer-provided health insurance, 323–324
 Employment, 271–274
 Employment effects, 302
 estimates across all studies, 304–305
 estimates for broad groups, 305–306
 heterogeneity of, 306–313
 Endogeneity concerns, 697
 Endogeneity problem, 5
 Enforceability, 816
 Enforceable non-compete agreements, 815
 Equilibrium distribution of wages across firms, 783
 Equilibrium markup, 780
 Equivalence theorem, 9f
 Error-correction model (ECM), 697
 Estimators, 131–132
 Event study, 277–280
 vs. TWFE distributed lag estimates, 362–364
 Evidence from job finding rates and reemployment wages, 451
 basic search model rationalize evidence on job finding rates and reemployment wages, 455–459
 effects of UI on unemployment duration, 452–453
 effects of UI the hazard rate, 453–454
 reemployment wages, 454–455
 Exogenous mobility, 126
 Extended monotonicity (EM), 31
- F**
- Fair Labor Standards Act, 265, 820
 False discovery proportion (FDP), 229
 False discovery rate (FDR), 229
 Families, public policies, and labor market, 581
 child outcomes, 606–609
 conceptual framework, 582–585
 family labor supply, 596–604, 596f, 597f, 598f, 599f
 fertility, 586–591, 587f, 588f
 future research, avenues for, 611
 gender inequality, 604–606, 605f
 marriage, divorce, and cohabitation, 591–595, 592f, 593f, 594f
 norms and spillovers, 609–610
 OECD countries, public policies in, 585–586, 585f, 586f
 Family Care Act, 589
 Family decision-making models, 584
 Family labor supply
 public policies and, 598–604
 rates and trends, 596–598, 596f, 597f, 598f, 599f
 Family leave, 603

- Family-wise error rate (FWER), 228
FCSL. *See* Food services, cleaning, security, and logistics
FDP. *See* False discovery proportion
FDR. *See* False discovery rate
Federal Trade Commission (FTC), 813, 817–818
Fertility
 public policies affecting, 587–591
 rates and trends, 586–587, 587f, 588f
Firm-level exposure, 300–301
Firm-level labor market discrimination, 235
 contrasting shrinkage approaches, 245–255
 distributions of discrimination, 236–240
 multiple testing to detect discrimination, 243–245
 posterior predictions of discrimination, 240–243
Firm level labor supply elasticity of Uber, 795
Firm-level wages, 786
Firms' incentives to invest in costly hiring efforts, 790
Firm size heterogeneity, role of, 791–792
Firm wage effects
 AKM model, 125
 causality, 140–146
 edgy interpretation of firm effects, 127–135
 evaluating AKM restrictions, 135–140
 covariance between person and firm effects, 172–175
 hiring origins and state dependence, 166
 destination effects and origin effects, 169–170
 information and conduct, 170–171
 structural interpretation, 167–168
 testable restrictions, 168–169
regressing firm effects on observables, 161
 one step *vs.* two, 162–163
 revisiting firm size wage premium, 164–166
 variance estimation, 163–164
sorts of firms pay high wages, 120
 entry, reallocation, and dynamics, 122–123
 industry structure and amenities, 124–125
 productivity, worker flows, and firm size, 121–122
 sorting, outsourcing, and displacement, 123–124
variance decomposition, 146
clustering approaches, 156–158
cross-fitting and bias correction, 148–155
limited mobility bias, 147–148
variable worker and firm effects, 159–161
Firm willingness to hire workers with criminal records, 703
First-order approximation, 508
Flexibility penalties, 647
Food services, cleaning, security, and logistics (FCSL), 123
Formal asymptotic theory, 67
Forward engineering, 44, 91
 assuming away problem, 45–46
 binary treatments when monotonicity violated, 70–71
 estimating LATEs and ACRs in presence of covariates, 46–53
 marginal treatment effects, 53–70
 no selection model, 81–85
 ordered treatments, 72–78
 summary of, 86–87
 unordered treatments, 78–81
Frequency distribution, 292–294
Frisch-Waugh-Lovell algebra, 92
FTC. *See* Federal Trade Commission
Full-time equivalent (FTE), 313
Fundamental cycle, 128
Fuzzy regression-discontinuity framework, 701

G

- Gangs and organized crime, 748–750
Gender inequality, 619, 621f, 622f
 career-family trade-offs, 639–644
 differential job sorting and organization of work, 646–649
 evolving perspectives on, 635–644
 gender biology and productivity, 644–646
 gender norms, 658–664
 identity and norms in understanding, 652–664
labor supply and household specialization, relevance for, 653–655
micro-macro linkages, 664–665
monopsonistic labor markets, 649–652
motherhood, anatomy of career costs of, 644–652
preferences, traits, and constraints, 636–639
public policies and, 605–606
rates and trends, 604–605, 605f

- real world and academic developments in, 624–630, 625f, 628f, 629t–630t
- stereotypes, beliefs, and discrimination, 655–658
- women's labor supply and gender gap, 630–635, 633f
- Gender norms, 658
- cultural change and learning, 659–661
 - historical origins and persistence, 659
 - information gaps, 663–664
 - transmission channels, 662–663
- Generalized method of moments (GMM), 208
- Gram matrix, 58
- Gronau-Heckman normal selection model, 60, 61
- Gronau-Heckman selection model, 3, 45
- H**
- Handbook of Labor Economics, 581
- Harassment-related mobility, 637
- Hartz reforms, 421
- Hazard rate, 445–446
- Heckman selection correction, 60
- Herfindahl-Hirschman Index (HHI), 768, 773, 786, 791, 799, 800, 801, 812, 814
- Heterogeneity, 306–313
- Heterogeneous treatment effects, 35
- HHI. *See* Herfindahl-Hirschman Index
- Hicks-Marshall rule, 268
- High vs. low minimum wages, 310–311
- HILDA survey. *See* Household, Income, and Labor Dynamics in Australia survey
- Hosios condition, 536
- Household, Income, and Labor Dynamics in Australia (HILDA) survey, 654
- Human capital costs of youth victimization, 745
- I**
- Identity and norms in understanding gender inequalities, 652
- gender norms, 658–664
- labor supply and household specialization, relevance for, 653–655
- stereotypes, beliefs, and discrimination, 655–658
- Imperfect competition, 337–338
- Incapacitation, 741–742
- Incidence matrix, 128
- Income, 687–689
- Incorporating precision-dependence, 224–225
- Incumbent worker based approach, 295
- Indirect contrasts, 117
- Industrial organization (IO) literature, 807
- Industry structure, 124–125
- Inequality, 341–348
- Infinitesimal firms, monopsony power with, 781–784
- In-kind subsidies, 589, 594, 601, 606
- Instrumental variable (IV) methods, 2, 640
- from classical linear IV model to potential outcomes, 5–7
 - definition of weakly causal estimand, 92–94
 - derivations for marginal treatment effects, 96–99
 - deriving average causal response and alternative decomposition, 94–96
 - empirical production possibility frontier for, 86f
 - estimating average causal response with covariates, 96
 - forward engineering, 44
 - assuming away problem, 45–46
 - binary treatments when monotonicity violated, 70–71
 - estimating LATEs and ACRs in presence of covariates, 46–53
 - marginal treatment effects, 53–70
 - no selection model, 81–85
 - ordered treatments, 72–78
 - summary of, 86–87
 - unordered treatments, 78–81
- full exogeneity, 10–11
- in nutshell, 4
- potential outcomes or latent variables, 92
- recommendations for practice, 87
- assessing the likely role of UHTE, 87–88
 - forward engineer estimates of interpretable target parameters, 90–91
 - reverse engineer with caution, 88–90
- reverse engineering, 14
- binary treatment, binary instrument, no covariates, 16–18
 - covariates, 34–39
 - estimators, estimands, and weak causality, 15–16
 - multiple instruments, 26–27
 - multivalued instruments, 18–21
 - ordered, cardinal treatments, 27–29
 - summary of, 39–44
 - unordered or non-cardinal treatments, 29–34

- violations of monotonicity, 21–26
selection models, 7–10
target parameters, 11–14
testability, 14
unobserved heterogeneity in treatment effects, 4–5
- Intent to treat (ITT), 47
Internal Revenue Service (IRS), 604, 701
Interpretation, 2
Intransitive firms, 172
In-vitro fertilization (IVF), 641, 642–644
IO literature. *See* Industrial organization literature
IRS. *See* Internal Revenue Service
“isolated firm” monopsony model, 762, 767
IVF. *See* In-vitro fertilization
- J**
- James/Stein theorem, 193–195
JobKeeper program, 417
Job-level employment discrimination, 225–228
Job loss, 414
Job search model, 442
 effects of UI on job finding and reemployment wages, 449
 non-stationary environment, 450–451
 static environment, 449–450
empirical moments, 445–446
first order conditions, 444–445
search effort and reservation wages
 throughout unemployment spell, 446
 multiple types, 446–448
 single type, 446
steady state, 445
- Judge designs, 23, 25
- K**
- Kernel-based linear regression estimator, 77
Kerr, Clark, 269
Keynesian model, 421
Kirchhoff’s matrix tree theorem, 131
K-means clustering, 156
Knighthood, 442
- L**
- Labor and product market power, simultaneous, 778–781
Labor costs, 274
Labor demand, 271–274
Labor economists, 184
Labor market, 270, 746–747
 policy, 439f
- regulation, 403–404
victimization and, 746–747
- Labor market concentration
 concentration, market power and welfare, 799–801
empirical evidence on, 803–804
measurement of, 801–802
- Labor wedge, 807
Large firms, monopsony power with, 784–787
Large-scale inference, 228
LATEs. *See* Local average treatment effects
Least squares estimator, 117, 131
Leave-out connectedness, 151
Leibniz’s rule, 209
Lester, Richard, 269
Limited mobility bias, 147–148
Linear IV estimator, 51
Linear regression formulation, 58–59, 74–76
Linear shrinkage, 224
Liquidity to moral hazard ratio approach, 508–509, 513
LMP crime reporting system. *See* London Metropolitan Police’s crime reporting system
Local average treatment effects (LATEs), 2, 3, 13, 17, 273
Local false discovery rate, 234
Local instrumental variables (LIV), 64–66
Local projections DiD (LP-DiD) regression, 278
Log-spline deconvolution estimator, 221–223
London Metropolitan Police’s (LMP) crime reporting system, 697
Long-run effect, 363
Loss function, 232
- M**
- Machine learning (ML), 49–50, 211, 292
 algorithms, 185
 connections to, 211–213
Macroeconomic approach, 419
Macroeconomic models, 819
Manski-Robins and IV intersection bounds, 81–82
Marginal cost (MC), 779
Marginal-propensity-to-consume (MPC)
 approach, 509–511
Marginal rate of substitution (MRS), 510
Marginal revenue (MR), 779
Marginal Revenue Product of Labor (MRPL), 761–762, 780, 793, 799, 807, 809
Marginal treatment effect (MTE), 3, 53

- applications and uses of, 68–70
 definitions, 53–55
 derivations for, 96
 derivations of weighting expressions, 96–98
 normal selection model, 98
 saturated MTR specifications reproduce LATE, 98–99
 empirical applications of, 69t
 estimation and inference, 66–67
 identification, 60–64
 linear regression formulation, 58–59
 motivation, 55–58
 unstratified regressions and local instrumental variables, 64–66
- Marginal treatment response (MTR), 55, 74
- Marginal value of public funds (MVPFs), 413, 437, 518
 framework, 518–519
 of UI policies, 519–522
- Margins of adjustment, 314
 amenities, 323–324
 employment and refinements, 317t–318t
 firm-entry and exit, 326–328
 incidence, 320t–321t
 input prices and rent, 332
 migration and participation, 328–330
 modeling implications and open questions, 337–341
 non-compliance, 315–323
 output prices and consumers reactions, 330–332
 productivity, 334–335
 profits, 332–333
 substitution with other inputs, 324–326
 summary of evidence on margins of adjustment, 335–337
 wage retrenchment of higher-skilled workers, 330
 worker turnover and reduction in training costs, 334
- Marital breakup, 584
- Markdown estimates, 810–812
- Market concentration, 312
- Market-level elasticity, 805
- Marriage, divorce, and cohabitation
 public policies affecting, 594–595
 rates and trends, 591–593, 592f, 593f, 594f
- MATLAB, 132
- Maximum a posteriori (MAP), 212
- MC. *See* Marginal cost
- Mean squared error (MSE), 185
 improvements in normal/normal model, 192–193
- Mechanical cost, 413
- Mechanical Turk (MTurk) experiments, 795
- Merger control, 812–815
- Merger effects in labor market, 803
- Meta analysis studies, 542–544
- #MeToo movement, 637
- Mexico's Regional Development Program, 795
- MHHI. *See* Modified Herfindahl-Hirschman index
- Micro elasticity, 536
- Micro-macro linkages, 664–665
- Migration, 328–330
- Minimum wages, 261, 818–820
 bias from heterogeneous pre-existing trends, 361–364
 constructing historical QCEW restaurant data, 366
 construction of 60 state-level minimum wage events, 368–369
 construction of probability groups using demographic predictors, 370
 data sources for cross-country Kaitz indices, 364
 Brazil, 364
 China, 364–365
 India, 365–366
 OECD countries, 366
 United States, 366
 debate, 268–271
 in developing countries, 348–350
 heterogeneity analysis and robustness for impacts of increased MW, 360t
 inequality, distributional implications, and downstream effects, 341–348
 literature, 262
 margins of adjustment, 314
 amenities, 323–324
 firm-entry and exit, 326–328
 input prices and rent, 332
 migration and participation, 328–330
 modeling implications and open questions, 337–341
 non-compliance, 315–323
 output prices and consumers reactions, 330–332
 productivity, 334–335
 profits, 332–333
 substitution with other inputs, 324–326
 summary of evidence on margins of adjustment, 335–337

- wage retrenchment of higher-skilled workers, 330
worker turnover and reduction in training costs, 334
rationale for minimum wage policies, 264–267
wage and employment effects of, 271
effect on total hours, 313–314
exploiting local variation in level of minimum wages, 274–288
exploiting minimum wage exemptions, 301–302
exploiting nation-wide variation in level of minimum wages, 295–301
methods to estimate overall effect of policy, 291–295
other considerations for comparability of treatment and control groups, 288–291
review of evidence on employment effects, 302–313
wages, employment and labor demand, 271–274
- Minneapolis, 703
Minnesota, 703
Mobility network, 127, 128f
Modified Herfindahl-Hirschman index (MHHI), 802, 805
Monopsonistic competition and oligopsony with discrete choice, 773–776
Monopsonistic competition, nesting with, 777–778
Monopsonistic competition with representative household and differentiated jobs, 770–772
Monopsonistic labor markets, 649–652
Monopsony power, 764–767
determinants of, 792, 793t
differentiated jobs, 769–781
in directed search model, 787–791
firm size heterogeneity, role of, 791–792
with infinitesimal firms, 781–784
with large firms, 784–787
monopsonistic competition and oligopsony with discrete choice, 773–776
monopsonistic competition with representative household and differentiated jobs, 770–772
nested logit, 776–777
nesting with monopsonistic competition, 777–778
oligopsony, 767–769
oligopsony with representative household and differentiated jobs, 772–773
positive employment effect of, 790–791
search frictions, 781–791
simultaneous labor and product market power, 778–781
- Monotonicity
condition, 3, 9
correct first stage specifications, 37–38
violations of, 21–26
Monte Carlo exercises, 158
Moore-Penrose inverse, 131
Mortensen, Dale, 442
Motherhood, anatomy of career costs of differential job sorting and organization of work, 646–649
gender biology and productivity, 644–646
monopsonistic labor markets, 649–652
Mount-joy's argument, 80
MR. *See* Marginal revenue
MRPL. *See* Marginal Revenue Product of Labor
MTE. *See* Marginal treatment effect
MTurk experiments. *See* Mechanical Turk experiments
Mueller-Smith and Schnepel study, 701
Multi-Fiber Agreement in 2005, 651
Multiple instruments, 26–27
Multiple testing, 228–231
Multivalued instruments, 18–21
Multivariate deconvolution, 211
Multivariate EB, 201–202
MVPFs. *See* Marginal value of public funds
- N**
- National Bureau of Statistics (NBS), 364
National Industrial Recovery Act (NRA), 265
1997 National Longitudinal Survey of Youth (NLSY97), 698–699
National Longitudinal Survey of Youth (NLSY) cohort data, 323, 688
Near-cash transfer, 590, 595
Neoclassical approach, 270
Nested logit, 776–777
Net public cost, 416
New York City program, 695
Neyman orthogonal, 50
NLSY97. *See* 1997 National Longitudinal Survey of Youth
NLSY cohort data. *See* National Longitudinal Survey of Youth cohort data
Noisy standard errors, 210–211

- Non-cardinal treatments, 29–34
 Non-compete agreements, 816–817
 Non-competition agreements, 815–818
 Non-compliance, 315–323
 Non-parametric empirical Bayes methods, 185,
 216
 bias-corrected variance estimation,
 217–221
 compound decisions and shrinkage strate-
 gies, 233–235
 firm-level labor market discrimination,
 235–255
 for multiple testing, 228–231
 non-parametric priors and posteriors,
 221–225
 partial identification, 225–228
 ranking problems, 231–233
 Non-parametric maximum likelihood estimator
 (NPMLE), 185, 221–222
 Non-parametric posteriors, 223–224
 Non-parametric priors, 221–225
 Nonseparable relationship, 6
 Non-stationary environment, 450–451
 Normalized selection equation, 54
 Normal selection model, 98
 Normative approach, 406–407
 Norms and spillovers, 609
 public policies and, 609–610
 North American Industry Classification System
 (NAICS), 366
 NPMLE. *See* Non-parametric maximum like-
 lihood estimator
 Nutshell, 4
- O**
 Oaxaca-Blinder decomposition, 626
 Observed heterogeneity in treatment effects
 (OHTE), 88
 OECD countries, public policies in, 585–586,
 585f, 586f
 Oligopoly, 762, 767–769, 773–776
 and job differentiation models, 763
 with representative household and differ-
 entiated jobs, 772–773
 One-step maximum likelihood approach, 207
 On-the-job training, 324
 OOI. *See* Outside Options Index
 “Opportunity cost” of crime, 687
 Optimal decision rule, 205
 Ordered treatments, 72
 continuous treatments, 76–78
 linear regression formulation, 74–76
- selection models that do not allow for het-
 erogeneity, 78
 threshold-crossing with multiple treat-
 ments, 72–73
 Organized crime, gangs and, 748–750
 Output prices, 330–332
 Outside Options Index (OOI), 804
 Outsourcing, 123–124
 Own-wage employment elasticity (OWE), 271,
 302
- P**
 Panel Study of Income Dynamics (PSID), 508
 Parameterizations, 8
 Parametric identification, 157
 Partial identification, 63, 225–228
 Partially linear IV (PLIV), 52
 Participation, 328–330, 329–330
 Passive spending Anglican countries, 440f
 Passive spending EU countries, 440f
 Paycheck Protection Program, 417
 Person effects, 160
 Pigou’s rate of exploitation, 768
 Pluralistic ignorance, 663
 Policy and monopsony power, 812
 merger control, 812–815
 minimum wage, 818–820
 non-competition agreements, 815–818
 Policy-makers, 272
 Policy-relevant treatment effects (PRTEs), 12,
 68
 Polytree, 133
 Positive approach, 407–412
 Positive employment effect of monopsony
 power, 790–791
 Positive equilibrium employment effect, 790
 Posteriors, 221–225
 Post-Great Recession period, 395
 Post-periods, 286, 369
 Precision-dependence, 206
 noisy standard errors, 210–211
 testing and modeling precision-depen-
 dence, 208–209
 variance-stabilizing transformations,
 209–210
 Pre period, 286
 Price effects, 345–348
 Probability integral transform, 54
 Pro-competitive mergers, 800
 Production function approach, 806–809
 Productivity, 334, 338
 and employment, 335

- quantity-based productivity, 335
- reallocation, 335
- revenue-based productivity, 334–335
- Profit-maximizing wages, 806
- Profits, 332–333
- Project STAR experiment, 161
- Propensity score weighting, 46–49
- PRTEs. *See* Policy-relevant treatment effects
- Public policies
 - affecting fertility, 587–591
 - affecting marriage, divorce, and cohabitation, 594–595
 - and child outcomes, 606–609
 - and family labor supply, 598–604
 - and gender inequality, 605–606
 - and norms and spillovers, 609–610
 - in OECD countries, 585–586, 585f, 586f
- Q**
- Quarterly Census of Employment and Wages (QCEW), 366
- Quebec reform, 602
- R**
- Race, 307–308
- Random coefficients model, 71
- Ranking problems, 231–233
- Rates and trends
 - family labor supply, 596–598, 596f, 597f, 598f, 599f
 - fertility, 586–587, 587f, 588f
 - gender inequality, 604–605, 605f
 - marriage, divorce, and cohabitation, 591–593, 592f, 593f, 594f
- RDP. *See* Regional Development Program
- Reallocation, 335
- Reduced-form approach based on workers' outside options, 804
- Reemployment wages, 445–446, 454–455
- Refining search model, 473
 - biased beliefs, 491–492
 - channels that determine UI response, 475–476
 - directed search *vs.* reservation wages, 476–478
 - duration dependence in job finding rates, 481–484
 - duration dependence in reemployment wages, 478–481
 - employer collusion/storable offers, 493–494
 - learning/information, 494–495
 - locus of control, 492
- present bias *vs.* exponential discounting, 484–488
- reference dependence, 488–490
- Regional bite, 298–300
- Regional Development Program (RDP), 795
- Regressing firm effects on observables, 161
 - one step *vs.* two, 162–163
 - revisiting firm size wage premium, 164–166
 - variance estimation, 163–164
- Representative household and differentiated jobs
 - monopsonistic competition with, 770–772
 - oligopsony with, 772–773
- Representative household, oligopsony with, 772–773
- Reproducibility, 172
- Research community, 262
- Reservation/target wages, 466
 - evidence on consumption during unemployment, 469–472
 - other types of evidence, 472–473
 - search platform and process data, 468–469
 - survey data, 467–468
- Reservation wage property, 444
- Restricted edge effects, 129–131
- Revealed-preference approach, 512
- Reverse engineering, 3, 14
 - binary treatment, binary instrument, no covariates, 16–18
 - covariates, 34–39
 - estimators, estimands, and weak causality, 15–16
 - linear IV estimands, 40f
 - multiple instruments, 26–27
 - multivalued instruments, 18–21
 - ordered, cardinal treatments, 27–29
 - summary of, 39–44
 - unordered or non-cardinal treatments, 29–34
 - violations of monotonicity, 21–26
- Reynolds, Lloyd, 269
- Ridge regression, 212
- Roadmap, minimum wage policies, 274
- Rosen framework, 769
- S**
- Sample temperature, 91
- Same-sex marriages, 595
- San Francisco minimum wage, 313
- Sarsons' study, 657
- Schooling quantity and quality, 742–743

- School value-added in Boston, 213–216
SciPy, 132
Search effort, 459
 search platform and process data, 463–466
 survey data, 459–463
Search frictions, 781
 monopsony power in directed search model, 787–791
 monopsony power with infinitesimal firms, 781–784
 monopsony power with large firms, 784–787
Seattle minimum wage, 313
Secondary earner, labor supply of, 631, 633f
 modeling gender norms and beliefs, 634–635
Selection model, 81
 empirical illustration, 82–84
 Manski-Robins and IV intersection bounds, 81–82
 role of, 84–85
Selection models, 7–10
Sequential auction models, 118
Severance pay (SP) system, 438
Sexist paternalism, 656
Short regression coefficient, 199
Short-run effect, 363
Short-time work (STW), 386
 design of STW schemes, 389–393
 effects of STW at macroeconomic level, 419
 in Covid-19 crisis, 421–422
 before great recession of 2008–2009, 419–420
 in great recession of 2008–2009, 420–421
 role of timing of STW regulation and eligibility criteria, 422–423
 effects on firms, 423
 employment and hours of work, 423–425
 firm productivity, profitability and firm survival, 426–427
 job reallocation and productivity, 427–428
 effects on workers trajectories, 428–429
 net public cost of, 416
 costs and benefits for public expenditure, 416–417
 effectiveness at targeting vulnerable jobs, 417–418
 and other workforce retention measures, 389
social willingness to pay for, 414
 impact on short-time non-users, 415–416
 impact on short-time users, 414–415
spread of STW since 1920s, 388–389
take-up, 393
 design and administrative capacity, 404–406
 by firms across and within sectors, 397t
 by firms across and within size, 398t
 by firms by firm size, 399t
 and labor market regulation, 403–404
 in large recessions, 401
 outside of large recessions, 401–402
 by type of workers and firms, 393–401
theoretical models of, 406
 normative approach, 406–407
 positive approach, 407–412
Short vs. medium run estimates, 308–310
Shrinkage, 190–191
Simulation, calibration and, 804–805
Slope coefficient, 197
Social Security Administration (SSA), 604
Sorting, 123–124
S&P 500 index, 803
Spanning trees, 128, 142–144
Spillovers, 342, 558–560
 empirical evidence on, 537–539
 norms and, 609–610
SSA. *See* Social Security Administration
Stable unit treatment value assumption (SUTVA), 291, 296
Standard Industrial Classification (SIC), 366
Static environment, 449–450
STEM-related activities, 658
Stereotypes, 658
Structural estimation, 805–806
Structural interpretation, 167–168
STW. *See* Short-time work
Substitution, 324
 capital-labor substitution, 325
 labor-labor substitution, 325–326
Summer youth employment programs (SYEP), 694–695
Sweden's compulsory schooling reform, 740
Swedish administrative data, 158
Swedish tax reforms, 609, 654
SYEP. *See* Summer youth employment programs
Synthetic control approach, 280

T

- Target parameters, 11–14
 Taxing earnings, 584
 Teachers, gender-biased, 658
 Testability, 14
 Testable restrictions, 168–169
 Three-step EB recipe, 191–192
 Threshold-crossing model, 7
 Time-varying covariates, 289
 TWFE-log(MW), 275–277
 Two-stage least squares (2SLS) estimator, 2, 10
 Two-way fixed effects (TWFE) models, 220
 Type II errors, 205

U

- Uber's firm level labor supply elasticity, 795
 Uncorrelated random effects model, 200
 Unemployment, 689–694
 Unemployment insurance (UI), 412, 415, 436, 439–441
 design of UI policy, 496
 marginal value of public funds, 518–522
 micro and macro effects of UI programs, 534–539
 other UI design questions, 530–534
 quantification of behavioral costs, 503–507
 quantification of social value of UI changes, 507–518
 structure of unemployment insurance policies, 497
 UI effects on job separation, 525–530
 UI effects on wages, 523–524
 welfare effects of unemployment insurance, 497–503
 micro-foundations of job search among unemployed, 441
 basic job search model, 442–451
 brief history of job search theory, 441–442
 evidence from job finding rates and re-employment wages, 451–459
 new empirical moments, 459–473
 refining search model, 473–495
 origin of unemployment insurance and active labor market policy, 437–439

Unobserved heterogeneity in treatment effects (UHTE), 2–5

- Unordered treatments, 29–34, 78–81
 Unstratified regressions, 64–66
 Upward-sloping market-level labor supply, 790
 US Census Longitudinal Business Database microdata, 804
 US Federal Trade Commission, 817–818

V

- Value of the marginal product of labor (VMPL), 780, 807
 Variable worker and firm effects, 159–161
 Variance decomposition, 146
 clustering approaches, 156–158
 cross-fitting and bias correction, 148–155
 limited mobility bias, 147–148
 variable worker and firm effects, 159–161
 Variance estimation, 163–164
 Variance-stabilizing transformations (VSTs), 209–210
 Variance-weighted ATT (VWATT), 278
 Veneto Workers History (VHW) data, 135
 Victimization, 744–748
 VMPL. *See* Value of the marginal product of labor
 Vytlačil's equivalence theorem, 9, 54

W

- Wage Dynamics Network survey (WDN3), 394, 395f, 396f
 Wages, 271–274
 and income, 687–689
 minimum, 818–820
 Wald estimand, 17
 Walmart, 819–820
 Weak causality, 15, 16
 Weighting estimators, 48
 Welfare effects of unemployment insurance, 497
 baseline framework, 498–500
 dynamic framework, 500–503
 WFH. *See* Work from home
 WFTC. *See* Working Families' Tax Credit
 Willingness to pay (WTP), 648
 Women's labor supply and gender gap, 630

secondary earner, labor supply of, 631–635, 633f
Worker-level exposure, 301
Work from home (WFH), 645–646, 647, 648
Working Families' Tax Credit (WFTC), 600
World Bank, 262
WTP. *See* Willingness to pay

X

X-efficiency, 338

Y

Youth labor markets, 694–696
Youth victimization, human capital costs of, 745

Handbooks in Economics

Handbook of Labor Economics Volume 5

Christian Dustmann and Thomas Lemieux



North-Holland

An imprint of Elsevier
elsevier.com/books-and-journals

ISBN 978-0-443-29764-9



9 780443 297649