

ESSAYS IN HONOR OF JOON Y. PARK

Econometric Methodology in
Empirical Applications

Edited by Yoosoon Chang, Sokbae Lee
and J. Isaac Miller

ADVANCES IN
ECONOMETRICS

VOLUME 45B

ESSAYS IN HONOR OF
JOON Y. PARK

ADVANCES IN ECONOMETRICS

Series editors: Thomas B. Fomby, R. Carter Hill, Ivan Jeliazkov, Juan Carlos Escanciano, Eric Hillebrand, Daniel L. Millimet, Rodney Strachan

Recent Volumes:

- Volume 27A Missing Data Methods: Cross-sectional Methods and Applications – Edited by David M. Drukker
- Volume 27B Missing Data Methods: Time-series Methods and Applications – Edited by David M. Drukker
- Volume 28 DSGE Models in Macroeconomics: Estimation, Evaluation and New Developments – Edited by Nathan Balke, Fabio Canova, Fabio Milani and Mark Wynne
- Volume 29 Essays in Honor of Jerry Hausman – Edited by Badi H. Baltagi, Whitney Newey, Hal White and R. Carter Hill
- Volume 30 30th Anniversary Edition – Edited by Dek Terrell and Daniel Millimet
- Volume 31 Structural Econometric Models – Edited by Eugene Choo and Matthew Shum
- Volume 32 VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims – Edited by Thomas B. Fomby, Lutz Kilian and Anthony Murphy
- Volume 33 Essays in Honor of Peter C. B. Phillips – Edited by Thomas B. Fomby, Yoosoon Chang and Joon Y. Park
- Volume 34 Bayesian Model Comparison – Edited by Ivan Jeliazkov and Dale J. Poirier
- Volume 35 Dynamic Factor Models – Edited by Eric Hillebrand and Siem Jan Koopman
- Volume 36 Essays in Honor of Aman Ullah – Edited by Gloria Gonzalez-Rivera, R. Carter Hill and Tae-Hwy Lee
- Volume 37 Spatial Econometrics – Edited by Badi H. Baltagi, James P. LeSage and R. Kelley Pace
- Volume 38 Regression Discontinuity Designs: Theory and Applications – Edited by Matias D. Cattaneo and Juan Carlos Escanciano
- Volume 39 The Econometrics of Complex Survey Data: Theory and Applications – Edited by Kim P. Huynh, David T. Jacho-Chávez and Guatam Tripathi
- Volume 40A Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part A – Edited by Ivan Jeliazkov and Justin L. Tobias
- Volume 40B Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modelling Part B – Edited by Ivan Jeliazkov and Justin L. Tobias
- Volume 41 Essays in Honor of Cheng Hsiao – Edited by Tong Li, M. Hashem Pesaran and Dek Terrell
- Volume 42 The Econometrics of Networks – Edited by Áureo de Paula, Elie Tamer and Marcel-Cristian Voia
- Volume 43A Essays in Honor of M. Hashem Pesaran – Edited by Alexander Chudik, Cheng Hsiao and Allan Timmermann
- Volume 43B Essays in Honor of M. Hashem Pesaran – Edited by Alexander Chudik, Cheng Hsiao and Allan Timmermann
- Volume 44A Essays in Honour of Fabio Canova – Edited by Juan J. Dolado, Luca Gambetti and Christian Matthes
- Volume 44B Essays in Honour of Fabio Canova – Edited by Juan J. Dolado, Luca Gambetti and Christian Matthes
- Volume 45A Essays in Honor of Joon Y. Park – Edited by Yoosoon Chang, Sokbae Lee and J. Isaac Miller

ADVANCES IN ECONOMETRICS VOLUME 45B

ESSAYS IN HONOR OF JOON Y. PARK: ECONOMETRIC METHODOLOGY IN EMPIRICAL APPLICATIONS

EDITED BY

YOOSOON CHANG

Indiana University, USA

SOKBAE LEE

Columbia University, USA

And

J. ISAAC MILLER

University of Missouri, USA



United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Howard House, Wagon Lane, Bingley BD16 1WA, UK

First edition 2023

Editorial matter and selection © 2023 Yoosoon Chang, Sokbae Lee and J. Isaac Miller.
Individual chapters © 2023 The authors.

Published under exclusive licence by Emerald Publishing Limited.

Reprints and permissions service

Contact: permissions@emeraldinsight.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters' suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-83753-213-1 (Print)

ISBN: 978-1-83753-212-4 (Online)

ISBN: 978-1-83753-214-8 (Epub)

ISSN: 0731-9053 (Series)



ISOQAR certified
Management System,
awarded to Emerald
for adherence to
Environmental
standard
ISO 14001:2004.

Certificate Number 1985
ISO 14001



INVESTOR IN PEOPLE

CONTENTS

List of Contributors	vii
Introduction	ix

PART I MACROECONOMETRICS

Chapter 1 Aggregate Output Measurements: A Common Trend Approach <i>Martin Almuzara, Gabriele Fiorentini and Enrique Sentana</i>	3
Chapter 2 Markov Switching Rationality <i>Florens Odendahl, Barbara Rossi and Tatevik Sekhposyan</i>	35
Chapter 3 The Econometrics of Oil Market VAR Models <i>Lutz Kilian and Xiaoqing Zhou</i>	65

PART II FINANCIAL ECONOMETRICS

Chapter 4 Quantile Impulse Response Analysis with Applications in Macroeconomics and Finance <i>Whayoung Jung and Ji Hyung Lee</i>	99
Chapter 5 Risk Neutral Density Estimation with a Functional Linear Model <i>Marine Carrasco and Idriss Tsafack</i>	133
Chapter 6 Estimating Diffusion Models of Interest Rates at the Zero Lower Bound: From the Great Depression to the Great Recession and Beyond <i>Lealand Morin</i>	159
Chapter 7 A Market Crash or Tail Risk? Heavy Tails and Asymmetry of Returns in the Chinese Stock Market <i>Zeyu Xing and Rustam Ibragimov</i>	181

PART III
PANDEMIC, CLIMATE, AND DISASTER

Chapter 8 Predicting Crashes in Oil Prices During the COVID-19 Pandemic with Mixed Causal-Noncausal Models <i>Alain Hecq and Elisa Voisin</i>	209
Chapter 9 Depth-Weighted Forecast Combination: Application to COVID-19 Cases <i>Yoonseok Lee and Donggyu Sul</i>	235
Chapter 10 Identification of Beliefs in the Presence of Disaster Risk and Misspecification <i>Saraswata Chaudhuri, Eric Renault and Oscar Wahlstrom</i>	261
Chapter 11 A New Model for Agricultural Land-Use Modeling and Prediction in England Using Spatially High-Resolution Data <i>Namhyun Kim, Patrick Wongsa-art and Ian J. Bateman</i>	291
Chapter 12 Local Climate Sensitivity: What Can Time Series of Distributions Reveal About Spatial Heterogeneity of Climate Change? <i>J. Isaac Miller</i>	319

PART IV
MICROECONOMETRICS AND PANEL DATA

Chapter 13 Maximum Likelihood Estimation of Dynamic Panel Data Models with Interactive Effects: Quasi-Differencing Over Time or Across Individuals? <i>Cheng Hsiao and Qiankun Zhou</i>	353
Chapter 14 Informational Content of Factor Structures in Simultaneous Binary Response Models <i>Shakeeb Khan, Arnaud Maurel and Yichong Zhang</i>	385

PART V
RETROSPECTIVE

Chapter 15 Forty Years of <i>Advances in Econometrics</i> <i>Asli Ogunc and Randall C. Campbell</i>	413
---	------------

LIST OF CONTRIBUTORS

<i>Martín Almuzara</i>	Federal Reserve Bank of New York, USA
<i>Ian J. Bateman</i>	University of Exeter, UK
<i>Randall C. Campbell</i>	Mississippi State University, USA
<i>Marine Carrasco</i>	University of Montreal, Canada
<i>Saraswata Chaudhuri</i>	McGill University, Canada
<i>Gabriele Fiorentini</i>	Universitá di Firenze and RCEA, Italy
<i>Alain Hecq</i>	Maastricht University, Netherlands
<i>Cheng Hsiao</i>	University of Southern California, USA
<i>Rustam Ibragimov</i>	Imperial College London, UK; and St. Petersburg State University, Russia
<i>J. Isaac Miller</i>	University of Missouri, USA
<i>Whayoung Jung</i>	Korea Capital Market Institute, South Korea
<i>Shakeeb Khan</i>	Boston College, USA
<i>Lutz Kilian</i>	Federal Reserve Bank of Dallas, USA; and CEPR, UK
<i>Namhyun Kim</i>	University of Exeter, UK
<i>Ji Hyung Lee</i>	University of Illinois Urbana-Champaign, USA
<i>Yoonseok Lee</i>	Syracuse University, USA
<i>Arnaud Maurel</i>	Duke University and NBER, USA; and IZA, Germany
<i>Lealand Morin</i>	University of Central Florida, USA
<i>Florens Odendahl</i>	Banco de España, Madrid, Spain
<i>Asli Ogunc</i>	Texas A&M University-Commerce, USA
<i>Eric Renault</i>	University of Warwick, UK
<i>Barbara Rossi</i>	University Pompeu Fabra, Spain
<i>Tatevik Sekhposyan</i>	Texas A&M University, USA
<i>Enrique Sentana</i>	CEMFI, Spain; and CEPR, UK
<i>Donggyu Sul</i>	University of Texas at Dallas, USA
<i>Idriss Tsafack</i>	University of Montreal, Canada
<i>Elisa Voisin</i>	Maastricht University, Netherlands

<i>Oscar Wahlstrom</i>	Brown University, USA
<i>Patrick Wongsart</i>	Cardiff University, UK
<i>Zeyu Xing</i>	Imperial College London, UK; and UBP Investment Management, China
<i>Yichong Zhang</i>	Singapore Management University, Singapore
<i>Qiankun Zhou</i>	Louisiana State University, USA
<i>Xiaoqing Zhou</i>	Federal Reserve Bank of Dallas, USA

INTRODUCTION

Volume 45 of *Advances in Econometrics* honors Professor Joon Y. Park, who has made numerous and substantive contributions to the field of econometrics over a career spanning four decades since the 1980s and counting. Volume 45 consists of 28 chapters and is in fact split between two volumes with the first focusing on econometric theory and the second focusing on econometric applications. These papers have been contributed by Joon's friends, colleagues, coauthors, former students, and even his dissertation advisor, Professor Peter C. B. Phillips, and the volume is edited by his wife and most frequent collaborator, Professor Yoosoon Chang, and two of his former students.

In the typical fashion of *Advances in Econometrics*, the papers were to be submitted in early 2021 after a conference in Joon's honor in April 2020, which would have nearly coincided with his 65th birthday. Of course, the COVID-19 pandemic forced much of the world into lockdown in April 2020, so plans changed. Papers were still submitted in 2021, but the conference was delayed and, as of this writing, is scheduled for September 29–30, 2023, in Bloomington, Indiana, which Joon and Yoosoon have called home for nearly 15 years.

We introduce the 15 chapters of the second volume. The first 14 are grouped into 4 sections that are related – some closely and some very loosely – to Professor Park's work and especially to his more recent work. We conclude the volume with a retrospective article summarizing four decades of this series, *Advances in Econometrics*.

The first two decades of Joon's published record is dominated by contributions to theoretical time series that relate to empirical macroeconomics, broadly defined, more closely than to any other field outside of econometrics. However, much of his work then and up to the present has been motivated by a sincere interest in how the tools he pioneered could be used in empirical applications. These methodologies have influenced empirical work of all sorts.

In macroeconomics, he has published influential work on regime switching (Chang et al., 2017) and common stochastic trends (Chang et al., 2010), for example. Intensively studying high-frequency time series in the early 2000s, his contributions to financial econometrics include Ait-Sahalia and Park (2012), Choi, Jeong, and Park (2014), and Kim and Park (2017). His recent record contains several well-cited papers with methodologies motivated by understanding and forecasting energy consumption for the Republic of Korea and the world more generally, including Chang et al. (2014) and Chang et al. (2021). Related to energy consumption is climate change, and Park continues to make contributions to the econometric analysis of climate change, starting with Chang et al. (2020).

Following the themes mentioned above, the chapters in this volume are grouped as follows: (I) macroeconomics, (II) financial econometrics, (III) pandemic, climate, and disaster, and (IV) microeconomics and panel data.

PART I: MACROECONOMETRICS

We open the volume with a contribution by Martín Almuzara, Gabriele Fiorentini, and Enrique Sentana, entitled “Aggregate Output Measurements: A Common Trend Approach,” which relates the study of common trends, an area in which Professor Park has contributed significantly, to macroeconomic aggregates. The authors analyze a model for different measurements of a single persistent latent time series with mean-reverting measurement errors, thereby allowing for a common trend among these measurements. They find that over-differencing drives potentially large biases in estimation and reduces the precision of smoothed estimates of the latent variable. They obtain an improved aggregate output measure using US quarterly data.

Forecast rationality is a key principle of macroeconomics. While existing forecast rationality tests either focus on constant deviations from forecast rationality over the full sample or are constructed to detect smooth deviations based on nonparametric techniques, in “Markov Switching Rationality,” Florens Odendahl, Barbara Rossi, and Tatevik Sekhposyan propose novel parametric tests for detecting Markov switching deviations from forecast rationality. They find that forecasters tend to systematically overpredict interest rates during periods of monetary easing, while the forecasts are unbiased otherwise. Their findings emphasize the special role played by monetary policy in shaping interest rate expectations above and beyond macroeconomic fundamentals.

Energy is a factor of production in the macroeconomic production function of any country, so few commodities are as important as oil in understanding macroeconomic fluctuations. Lutz Kilian and Xiaoqing Zhou survey the extensive literature on oil market VARs in their contribution, “The Econometrics of Oil Market VAR Models.” As this literature has expanded at a rapid pace, it has become increasingly difficult for most economists to track the differences between alternative oil market models and the basis for divergent conclusions reached in the literature. This survey provides a useful guide, with a particular focus on the econometric foundations of the analysis of oil market models.

PART II: FINANCIAL ECONOMETRICS

In their contribution entitled “Quantile Impulse Response Analysis with Applications in Macroeconomics and Finance,” Whayoung Jung and Ji Hyung Lee study the dynamic responses of the conditional quantiles and their applications in macroeconomics and finance. This chapter builds a multi-equation autoregressive conditional quantile model and proposes a new construction of quantile impulse response functions (QIRFs). The new QIRF toolset the authors provide adds nicely to the burgeoning research efforts that have been devoted to measuring distributional effects of economic shocks. Using the QIRFs, the authors find that the left tail of economic activity is most responsive to monetary and financial market shocks, and they use this result to evaluate the impact of economic shocks on the 5% quantile of economic activity, a measure of growth-at-risk, during the global financial crisis.

In “Risk Neutral Density Estimation with a Functional Linear Model,” Marine Carrasco and Idriss Tsafack propose a nonparametric estimator of the risk neutral density based on cross-sectional European option prices. They show that the risk neutral density can be viewed as the solution of an ill-posed integral equation and estimate it using an iterative method called Landweber-Fridman. They establish the consistency and asymptotic normality of their estimator and provide an application to S&P 500 options.

Lealand Morin, in his contribution entitled “Estimating Diffusion Models of Interest Rates at the Zero Lower Bound: From the Great Depression to the Great Recession and Beyond,” proposes a new way to properly estimate a class of parametric diffusion models that can be used to represent the interest rate over a long time span possibly including several episodes where the interest rate may stay near or at the zero lower bound. This approach makes it easier to learn about the interest rate dynamics from major historic zero lower bound episodes in the United States, most notably the Great Depression and Great Recession. This enhanced understanding may help us predict future responses of key macroeconomic variables to the interest rate that has recently gone through a new episode of zero lower bound, from the outset of the COVID-19 pandemic to monetary policy tightening implemented to moderate inflation in early 2022.

Rapid stock market growth without real economic back-up has led to the 2015 Chinese stock market crash. In “A Market Crash or Tail Risk? Heavy Tails and Asymmetry of Returns in the Chinese Stock Market,” Zeyu Xing and Rustam Ibragimov analyze structural breaks in heavy-tailedness and asymmetry properties of returns in Chinese A-share markets due to the crash using robust methods for inference on the tail index. Their empirical results show that the main determinants of heavy-tailedness in Chinese financial markets are liquidity and company size.

PART III: PANDEMIC, CLIMATE, AND DISASTER

Continuing with the theme of crashes, the next set of articles relates to disasters past, present, and future, with an emphasis on the COVID-19 pandemic and climate change. Alain Hecq and Elisa Voisin contribute “Predicting Crashes in Oil Prices During the COVID-19 Pandemic with Mixed Causal-Noncausal Models,” which sheds light on how data transformations can substantially impact predictions made by mixed causal–noncausal models that rely on specifications in which time series depend not only on their lags but also on their leads. The authors investigate oil prices and estimate probabilities of crashes before and during the first wave of the COVID-19 pandemic in 2020, comparing various mechanical detrending methods with a detrending performed using the level of strategic petroleum reserves.

Yoonseok Lee and Donggyu Sul also investigate the recent pandemic in their contribution, “Depth-Weighted Forecast Combination: Application to COVID-19 Cases.” They develop a novel forecast combination approach based on the order statistics of individual predictability from panel data forecasts. Defining the notion of forecast depth based on normalized forecast errors during the training

period, they derive the limiting distribution of the depth-weighted forecast combination. Using this novel forecast combination, they predict the national level of new COVID-19 cases in the United States and find that the proposed method yields more accurate and robust predictions compared with other popular forecast combinations, including the ensemble forecast from the Centers for Disease Control and Prevention.

While the recent COVID-19 pandemic provides a tangible example of an economic disaster, economic disasters may have disparate causes. Saraswata Chaudhuri, Eric Renault, and Oscar Wahlstrom examine economic disasters more broadly in their contribution, entitled “Identification of Beliefs in the Presence of Disaster Risk and Misspecification.” They reconsider the equity premium puzzle and related asset market puzzles in light of the effect of rare disasters on asset prices. Low-probability economic disasters can restore the validity of model-implied moment conditions only if the amplitude of disasters may be arbitrarily large in due proportion. Yet they prove that there is no such thing as a population empirical likelihood-based model-implied probability distribution in the presence of unbounded disasters.

The next two chapters do not consider disasters explicitly, but there is a broad consensus within the scientific community on the potential for climate change to induce economic disasters. Land use is an issue that is inextricably tied to both the effects of climate change, by way of changes in arable land, for example, and to the causes of climate change, by way of changes in albedo, or reflection of solar energy. In their contribution “A New Model for Agricultural Land-Use Modeling and Prediction in England Using Spatially High-Resolution Data,” Namhyun Kim, Patrick Wongsa-art, and Ian J. Bateman contribute to a better understanding of farmers’ responses to behavioral drivers of land-use decisions by establishing an alternative analytical procedure that overcomes various drawbacks suffered by methods currently used in existing studies. Specifically, high-resolution spatial data ameliorates the idiosyncratic effects of the physical environment, and their model is equipped to deal with censoring, spatial dependence, and heterogeneity in the data and errors.

Also on the topic of spatial heterogeneity, J. Isaac Miller contributes “Local Climate Sensitivity: What Can Time Series of Distributions Reveal About Spatial Heterogeneity of Climate Change?” He introduces an easily implemented semiparametric statistical approach based on a physical energy balance climate model to estimate net heat transport and allow for spatial heterogeneity in the response of temperature to climate forcings. He finds that areas dominated by ocean tend to import energy and are relatively more sensitive to climate forcings, but that these areas warm more slowly than areas dominated by land.

PART IV: MICROECONOMETRICS AND PANEL DATA

In “Maximum Likelihood Estimation of Dynamic Panel Data Models with Interactive Effects: Quasi-Differencing Over Time or Across Individuals?” Cheng Hsiao and Qiankun Zhou consider the quasi-maximum likelihood

estimation (MLE) of dynamic panel models with quasi-differencing to remove interactive effects. They show that the quasi-difference MLE over time is inconsistent when T is large, whether N is fixed or large, and is consistent and asymptotically unbiased when the difference is across individuals when N is large, whether T is fixed or large.

Factor structures have been employed in a variety of settings in cross sectional and panel data models. In “Informational Content of Factor Structures in Simultaneous Binary Response Models,” Shakeeb Khan, Arnaud Maurel, and Yichong Zhang investigate the informational content of factor structures in discrete triangular systems. Their main finding is that imposing a factor structure yields point identification of parameters of interest, such as the coefficient associated with the endogenous regressor in the outcome equation, under weaker assumptions than are usually required in these models.

PART V: RETROSPECTIVE

Advances in Econometrics is a series of research volumes first published in 1982. Professor Park’s contribution of the variable addition test for cointegration to *Advances in Econometrics* in Park (1990) is both one of his most highly cited works and one of the most highly cited in the series. So, it seems appropriate to conclude this volume in his honor with a retrospective piece on the series. Asli Ogunc and Randall C. Campbell, with “Forty Years of *Advances in Econometrics*,” present an update to the history of the *Advances in Econometrics* series published in 2012. They describe key events in the history of the series and provide information about key authors, contributors, and other historical data on the series.

One of the joys of compiling this volume, “*Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications*,” has been to see the many in ways in which Joon’s research has influenced the methodologies used in empirical applications throughout the discipline. We hope you enjoy reading these chapters as much as we have.

ADDITIONAL REFERENCES

- Ait-Sahalia, Y., & Park, J. Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *Journal of Econometrics*, 169, 279–292.
- Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., & Park, J. Y. (2021). Forecasting regional long-run energy demand: a functional coefficient panel approach. *Energy Economics*, 96, 105117.
- Chang, Y., Choi, Y., & Park, J. Y. (2017). A new approach to model regime switching,. *Journal of Econometrics*, 196, 127–143.
- Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics*, 214, 274–294.
- Chang, Y., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2014). Time-varying long-run income and output elasticities of electricity demand with an application to Korea. *Energy Economics*, 46, 334–347.

- Chang, Y., Miller, J. I., & Park, J. Y. (2009). Extracting a common stochastic trend: Theory with some applications. *Journal of Econometrics*, 150, 231–247.
- Choi, H., Jeong, M., & Park, J. Y. (2014). An asymptotic analysis of likelihood-based diffusion model selection using high frequency data. *Journal of Econometrics*, 178, 539–557.
- Kim, J., & Park, J. Y. (2017). Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics*, 196, 37–54.
- Park, J. Y. (1990). Testing for unit roots and cointegration by variable addition. In G. F. Rhodes & T. B. Fomby (Eds.), *Advances in econometrics* (pp. 107–133). JAI Press.

PART I

MACROECONOMETRICS

This page intentionally left blank

CHAPTER 1

AGGREGATE OUTPUT MEASUREMENTS: A COMMON TREND APPROACH

Martín Almuzara^a, Gabriele Fiorentini^b
and Enrique Sentana^c

^a*Federal Reserve Bank of New York, New York, United States*

^b*Università di Firenze and RCEA, Florence, Italy*

^c*CEMFI and CEPR, Madrid, Spain*

ABSTRACT

The authors analyze a model for N different measurements of a persistent latent time series when measurement errors are mean-reverting, which implies a common trend among measurements. The authors study the consequences of overdifferencing, finding potentially large biases in maximum likelihood estimators (MLE) of the dynamics parameters and reductions in the precision of smoothed estimates of the latent variable, especially for multiperiod objects such as quinquennial growth rates. The authors also develop an R^2 measure of common trend observability that determines the severity of misspecification. Finally, the authors apply their framework to US quarterly data on GDE and GDI, obtaining an improved aggregate output measure.

Keywords: Cointegration; GDE; GDI; overdifferencing; signal extraction; statistical discrepancy

JEL Classification: C32; E01

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 3–33

Copyright © 2023 by Martín Almuzara, Gabriele Fiorentini and Enrique Sentana

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:[10.1108/S0731-90532023000045B001](https://doi.org/10.1108/S0731-90532023000045B001)

1. INTRODUCTION

Aggregate measurements, particularly those of output, are a key input to research economists and policy makers. Assessing the state of business cycles, making predictions of future economic activity, and detecting long-run trends in national income are some of their most popular uses. These measurements are typically regarded as noisy estimates of the quantities of interest, but accounting for the role of measurement error in applications is a difficult task. An important exception arises when more than one measurement of the same quantity is available. This makes it possible to combine the different measurements to produce a better estimate, ideally assigning higher weights to more precise ones.

In the United States, the Bureau of Economic Analysis (BEA) reports both the expenditure-based Gross Domestic Expenditure (GDE) measure of output and its income-based Gross Domestic Income (GDI) counterpart. If the sources and methods of the statistical office were perfect, then the two would be identical. In practice, however, they differ (see [Landefeld et al., 2008](#), for a review). The frequent, and at times noticeable, discrepancy between them (officially known as *statistical discrepancy*) has been recently the subject of active debate in academic and policy circles,¹ and various proposals for improved measures of economic activity have been discussed (see, e.g., [Aruoba et al., 2016](#); [Greenaway-McGrey, 2011](#); [Nalewaik, 2010, 2011](#)).² The *GDPplus* measure of [Aruoba et al. \(2016\)](#), for example, is currently released on a monthly schedule by the Federal Reserve Bank of Philadelphia.

In this chapter, we propose improved output measures under the assumption that alternative measurements in levels do not systematically diverge from each other over the long run. While economic activity, like several other macro aggregates, arguably displays a strong stochastic trend, one would expect statistical discrepancies to mean-revert. In that case, measurements in levels would share a common trend. Somewhat surprisingly, though, the standard practice is to rely on models that do not impose this common trend, working instead with the growth rates of measurements. To cite a few references, Smith et al. (1998), Nalewaik (2010, 2011), Greenaway-McGrey (2011), and Aruoba et al. (2016) all apply signal extraction techniques to a model of the first differences of log GDP. Similarly, the literature on GDP data revisions also works with growth rates, e.g., [Aruoba \(2008\)](#) and [Jacobs and van Norden \(2011\)](#) and Jacobs et al (2020).

In this respect, our main goal is to explore the implications of neglecting a common trend in levels for both parameter estimators and smoothed estimates of latent variables. Specifically, we follow [Smith et al. \(1998\)](#) in analyzing a model in which N different measurements y_t of an unobserved quantity x_t are available, so that

$$y_t = x_t \mathbf{1}_{N \times 1} + v_t,$$

with v_t denoting measurement errors in levels and $\mathbf{1}_{N \times 1}$ a vector of N ones. In contrast to the literature, though, we model x_t as $I(1)$ – i.e., Δx_t is stationary and strictly invertible – but v_t as $I(0)$. The discrepancies between measurements $y_{it} - y_{jt} = v_{it} - v_{jt}$ are thus cointegrating relationships, reflecting that mean

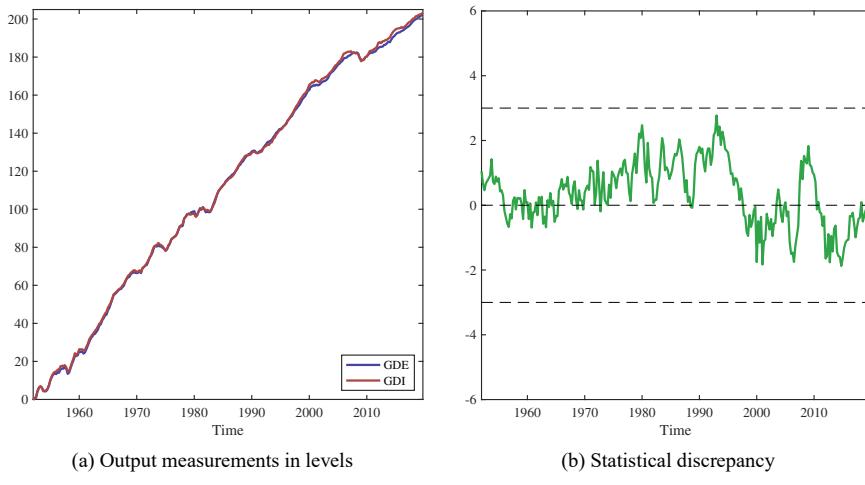


Fig. 1. GDE and GDI. *Notes:* We use November 2020's release of BEA national accounts estimates spanning 1952Q1–2019Q4; (a) $100 \times \log$ s of GDE and GDI subtracting their 1952Q1 values, i.e., percentage (log) growth in GDE and GDI accumulated since 1952Q1; (b) Differences between $100 \times \log$ of GDE and $100 \times \log$ of GDI.

reversion keeps alternative measurements from diverging. As a result, the measurement errors in first differences, Δv_i , are overdifferenced.

Fig. 1 shows US data counterparts to y_t and $y_{it} - y_{jt}$:

The parameters that describe the dynamics of x_t are typically of interest in themselves, as they inform important dimensions of business cycles and enter signal-extraction calculations. For that reason, we begin by studying the effects of ignoring cointegration among the elements of y_t on estimation procedures. We focus on Gaussian MLE in a simple parametric setup in which the model for x_t is correctly specified but that of y_t is not because of the neglected common trend. Our main finding is that even if x_t and y_t are stochastically independent, estimators of the autocorrelation parameters of x_t will be affected by misspecification in the dynamics of y_t , displaying potentially large biases and increased asymptotic variances. At the same time, we show that if the statistical model assumes Gaussian autoregressive dynamics for both Δx_t and Δy_t , then the estimators of their unconditional means and variances will be asymptotically unaffected. Consequently, the impact of misspecification will be confined to the autocorrelation structure of Δx_t .

Moreover, we prove that the extent to which inferences will be impaired is governed by (i) the severity of overdifferencing in measurement errors, and (ii) the overall signal-to-noise ratio. The more severely overdifferenced the elements of Δv_t are (i.e., the further away from unit root processes those measurement errors are), the stronger the dynamic misspecification resulting from the omitted common trend will be. In addition, a low degree of signal observability, which we

quantify by means of an R^2 measure of the relative contribution of x_t and v_t to the variation in observables, amplifies the role of incorrect modeling assumptions on v_t . In the limiting case of $R^2 = 1$, x_t is observable and misspecification in v_t is inconsequential.³ Our results therefore complement those in [Chang et al. \(2009\)](#), who derive the asymptotic distribution of the Gaussian MLE in a dynamic factor model with a single common trend. While [Chang et al. \(2009\)](#) study the case of unknown loadings under correct specification, we focus on the case of known loadings (equal to $1_{N \times 1}$) but subject to the dynamic misspecification induced by overdifferencing.⁴

Prediction, filtering and smoothing of x_t given data on y_t – signal extraction, for short – constitute the other main focus of our chapter. Given that the uncertainty of signal extraction calculations does not vanish in large T samples, unlike that of parameter estimators, we study their behavior at the pseudo-true parameter values, i.e., at the probability limits of ML estimators. Thus, we leverage on our estimation results to establish the suboptimality as a signal extraction technique of the Kalman-filter-based methods that neglect the common trend.

We find that the effect of ignoring the common trend is substantially different when signal extraction targets a short-run object and a long-run one. In particular, confidence sets for a long-run object such as an average of Δx_t over a relatively large time span are highly sensitive to even modest amounts of overdifferencing in Δv_t . This result is important because long-run objects are relevant to empirical questions about slowly evolving trends in macro variables. One example originates in the recurrent debate about growth deceleration in industrialized economies (e.g., [Gordon, 2016](#)). Another instance is the secular stagnation hypothesis, which implies a downward trend in interest rates (e.g., [Hansen, 1939](#); [Summers, 2015](#)). Similarly, the apparent secular decline in labor shares (e.g., [Blanchard, 1997](#); [Kaldor, 1957](#); [Karabarbounis & Neiman, 2014](#)) provides another case in point.

On the empirical side, we fit our proposed common trend model to US data on GDE and GDI. Through standard Kalman smoothing calculations, we obtain an improved measure of economic activity, which we compare to other existing measures in the literature. We then use our improved measure to assess the robustness of a variety of empirical facts on economic activity, involving both short- and long-run objects. Our main findings are the following: (1) point estimates of the serial correlation structure of economic activity appear robust to common trend assumptions, (2) the same seems to be true of point estimates of the quarterly average rate of growth in GDP, but (3) our common trend model gives rise to lower signal extraction uncertainty about economic activity than its competitors. Our third finding is conceptually important because point estimates of latent variables cannot be justified by an appeal to consistency – uncertainty about latent variables remains high regardless of the sample size, implying that such estimates must be accompanied by a measure of their precision. This is particularly important from an empirical point of view because the “putative” precision of estimates of economic activity which do not impose a common trend is so low that no sharp conclusion can be drawn about trends in growth from them. In contrast, our common-trend model provides noticeably more precise inference about such long-run objects.⁵

Of course, whether or not there is a common trend is an empirical question in its own right. The evidence that the statistical discrepancy between US GDE and GDI, although persistent, is mean-reverting is suggestive but not conclusive.⁶ Yet, the fact that, absent a common trend, the probability of observing large deviations between different measurements tends to one, lends strong support to our framework in the context of aggregate measurement problems.

The rest of the chapter is organized as follows. In Section 2, we present the basic setup. Section 3 discusses the properties of MLE while Section 4 is devoted to filtering. We report the results of our empirical analysis in Section 5. Finally, Section 6 concludes. Additional results are relegated to the Appendix and the supplemental material.

Notation. We use $\omega_{t_0:t_1}$ to denote the sequence $\{\omega_t\}_{t=t_0}^{t_1}$. If ω_t is a $d_1 \times d_2$ array for all t , and if it raises no confusion, we also use $\omega_{t_0:t_1}$ to denote the $d_1(t_1 - t_0 + 1) \times d_2$ array obtained by vertical concatenation of the terms of $\{\omega_t\}_{t=t_0}^{t_1}$. Analogously, $\psi_{1:N}$ denotes the column vector $(\psi_1, \dots, \psi_N)'$. We write $\mathbb{E}_T[\omega_t] = T^{-1} \sum_{t=1}^T \omega_t$ for the sample average of $\omega_{1:T}$, $\mathbb{E}[\omega_t]$ for its population counterpart, “ \xrightarrow{P} ” for convergence in probability and “ \Rightarrow ” for weak convergence.

2. MODEL

In our setup, the statistical office collects N (log) measurements y_t of an unobserved scalar (log) quantity x_t . Let v_t be the vector of (multiplicative) measurement errors so that, in first differences,

$$\Delta y_t = \Delta x_t 1_{N \times 1} + \Delta v_t, t = 1, \dots, T. \quad (1)$$

For a sample $\Delta y_{1:T}$, the data generating process is given by the probability distribution \mathbb{P} .

Assumption 1. \mathbb{P} satisfies the following:

- (A) The time series $\Delta x_{0:T}, v_{1,0:T}, \dots, v_{N,0:T}$ are cross-sectionally independent;
- (B) Δx_t is a Gaussian AR(1) process: For some values $\mu_0, \rho_0 \in (-1, 1), \sigma_0 > 0$,

$$\Delta x_0 \sim N(\mu_0, \sigma_0^2),$$

$$\Delta x_t | \Delta x_{0:(t-1)} \sim N\left(\mu_0 + \rho_0(\Delta x_{t-1} - \mu_0), (1 - \rho_0^2)\sigma_0^2\right), t = 1, \dots, T;$$

- (C) v_i is a Gaussian AR(1) process: For some values $\rho_i \in (-1, 1], \sigma_i > 0$,

$$v_{i0} \sim N\left(0, \frac{(1 + \rho_i)}{2}\sigma_i^2\right),$$

$$v_{it} | v_{i,0:(t-1)} \sim N\left(\rho_i v_{i,t-1}, \frac{(1 + \rho_i)}{2}\sigma_i^2\right), t = 1, \dots, T, i = 1, \dots, N.$$

Assumptions 1(A) and 1(B) are made in essentially every paper in the literature (e.g., Almuzara et al., 2019; Aruoba et al., 2016; Greenaway-McGrevy, 2011; Smith et al., 1998). Independence between Δx_t and measurement errors rules out cyclical patterns in the statistical discrepancy. Although potentially of substantive interest, introducing dependence between Δx_t and v_t or across the v_t 's complicates identification of the spectra of latent variables. Similarly, AR(1) dynamics for Δx_t is generally agreed to be a reasonable benchmark for economic activity data. Normality is unnecessary for most of our analysis, but since our focus is on the modeling of measurement errors and the role of dynamic misspecification, we adopt it for ease of exposition.

According to Assumption 1(B), we can regard $\Delta x_{0:T}$ as a segment from a strictly stationary process $\Delta x_{-\infty:\infty}$,

$$\Delta x_t = (1 - \rho_0)\mu_0 + \rho_0\Delta x_{t-1} + \sqrt{1 - \rho_0^2}\sigma_0\varepsilon_{0t},$$

with $\varepsilon_{0t} \stackrel{iid}{\sim} N(0,1)$. Our parameterization of the process for the signal ensures that $\mathbb{E}[\Delta x_t] = \mu_0$ and $\text{Var}(\Delta x_t) = \sigma_0^2$, which do not depend on ρ_0 , thereby separating these unconditional moments from the parameters governing the dynamics of Δx_t . Thus, we can summarize the serial dependence structure of the growth rate by its spectral density

$$f_0(\lambda) = \sigma_0^2 \frac{(1 - \rho_0^2)}{(1 - \rho_0 e^{i\lambda})(1 - \rho_0 e^{-i\lambda})} = \sigma_0^2 \left(\sum_{\ell=-\infty}^{\infty} \rho_0^{|\ell|} e^{i\ell\lambda} \right).$$

Assumption 1(C) implies Δv_{it} is overdifferenced, the severity of overdifferencing increasing as ρ_i moves away from unity. In fact, Δv_{it} is a strictly noninvertible ARMA(1,1) process, except in the limiting case $\rho_i = 1$, when Δv_{it} becomes white noise. We can view $\Delta v_{i,0:T}$ as a segment from a strictly stationary process $\Delta v_{i,-\infty:\infty}$,

$$\Delta v_{it} = \rho_i \Delta v_{i,t-1} + \sqrt{\frac{(1 + \rho_i)}{2}} \sigma_i \Delta \varepsilon_{it},$$

with $\varepsilon_{it} \stackrel{iid}{\sim} N(0,1)$. We have $\mathbb{E}[\Delta v_{it}] = 0$ and $\text{Var}(\Delta v_{it}) = \sigma_i^2$, and the spectral density of Δv_{it} is

$$f_i(\lambda) = \sigma_i^2 \frac{(1 + \rho_i)(1 - e^{i\lambda})(1 - e^{-i\lambda})}{2(1 - \rho_i e^{i\lambda})(1 - \rho_i e^{-i\lambda})},$$

which vanishes at frequency $\lambda = 0$ if $\rho_i \neq 1$ – an unequivocal symptom of overdifferencing.

When $\rho_i \neq 1$ for all i , the spectral density matrix of Δy_t at $\lambda = 0$ is $f_0(0)I_{N \times N}$. Therefore, it is singular with finite positive diagonal, implying the cointegration (of rank $N - 1$) of y_t . Thus, y_t is driven by a single common trend, x_t , while the statistical discrepancies $d_{ij,t} = y_{it} - y_{jt}$ are cointegrating relationships.⁷

Henceforth, we assume the econometrician formulates a statistical model $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ where $\theta = (\vartheta', \psi_{1:N}')'$ with $\vartheta = (\mu, \rho, \sigma)'$ and $\Theta = \Theta_x \times \Theta_v$, $\Theta_x \subset \mathbb{R} \times (-1, 1) \times \mathbb{R}_{>0}$ and $\Theta_v \subset \mathbb{R}_{>0}^N$. The distribution \mathbb{P}_θ is such that

- (a) The time series $\Delta x_{0:T}, \Delta v_{1,0:T}, \dots, \Delta v_{N,0:T}$ are cross-sectionally independent;
- (b) $\Delta x_t \sim N(\mu, \sigma^2)$ and $\Delta x_t | \Delta x_{0:(t-1)} \sim N(\mu + \rho(\Delta x_{t-1} - \mu), (1 - \rho^2)\sigma^2)$, $t = 1, \dots, T$;
- (c) $\Delta v_{it} \stackrel{iid}{\sim} N(0, \psi_i^2)$, $i = 1, \dots, N$.

From (a) and (b) it follows that the econometrician has correctly specified the model for $\Delta x_{0:T}$ conditional on $\vartheta_0 = (\mu_0, \rho_0, \sigma_0)' \in \Theta_x$, an assumption we maintain in what follows. Similarly, $\sigma_{1:N} \in \Theta_v$. In contrast, the model for the observed data $\Delta y_{1:T}$ is misspecified unless $\rho_i = 1$ for all i . In effect, (c) captures the idea that the econometrician neglects the common trend in y_t caused by the mean reversion of measurement errors because she assumes that $y_t = \sum_{\tau=1}^t \Delta v_\tau + v_0$ is a set of N independent random walks.

To ease the comparisons, the statistical model is also parameterized so that $\mathbb{E}_\theta[\Delta x_t] = \mu$ and $\text{Var}_\theta(\Delta x_t) = \sigma^2$, where the subscript θ indicates moments of the assumed distribution, so that the implied spectral density of Δx_t becomes

$$f_\vartheta(\lambda) = \sigma^2 \frac{(1 - \rho^2)}{(1 - \rho e^{i\lambda})(1 - \rho e^{-i\lambda})},$$

which coincides with f_0 at $\vartheta = \vartheta_0$. For measurement errors, $\mathbb{E}_\theta[\Delta v_{it}] = 0$ and $\text{Var}_\theta(\Delta v_{it}) = \psi_i^2$. Importantly, the assumed spectral density matrix of Δy_t at $\lambda = 0$ is $f_\vartheta(0)I_{N \times N} + \text{diag}(\psi_{1:N}^2)$, which is nonsingular.

Identification. A statistical model that makes use of Assumption 1(A) attains nonparametric identification of the spectra of latent variables. Given a spectral density matrix $f_{\Delta y}$ for the observables, equation (1) and Assumption 1(A) deliver

$$f_{\Delta y}(\lambda) = f_{\Delta x}(\lambda)I_{N \times N} + \text{diag}[f_{\Delta v}(\lambda)],$$

where $f_{\Delta x}$ is the spectral density of Δx_t , $f_{\Delta v}$ is the N -dimensional vector of spectral densities of $\Delta v_{1t}, \dots, \Delta v_{Nt}$ and $I_{N \times N}$ is a square matrix with N^2 ones. Therefore, the ij -th entry of $f_{\Delta y}(\lambda)$, for any $i \neq j$, equals $f_{\Delta x}(\lambda)$, which subtracted from the diagonal of $f_{\Delta y}(\lambda)$ yields $f_{\Delta v}(\lambda)$. In fact, Assumption 1(A) imposes overidentifying restrictions on $f_{\Delta y}$ for $N > 2$, as it implies that the off-diagonal elements of $f_{\Delta y}$ must be equal. Consequently, the joint probability distribution of the time series $\{\Delta x_t, \Delta v_t\}$ is identified under Gaussianity, provided one adds some restrictions on the unconditional means of the latent variables, which are necessary because there are $N + 1$ unconditional means but we only observe N measurements. Assumption 1, for example, imposes that the expectation of all measurement errors are zero, which is enough to identify $\mu_0 = \mathbb{E}[\Delta x_t]$ for any $N \geq 1$.

2.1. Observability of the Signal: A Key Parameter

Measures of the relative contributions of signal and noises to variation in observables are often important for understanding the quality of estimation and filtering in unobserved components models. To develop such a measure, we use the idea of minimal sufficient statistic for dynamic factor models in [Fiorentini and Sentana \(2019\)](#). With L the lag operator and $F_i(\cdot)$ the autocovariance generating function of Δv_{it} ,⁸ the Generalized Least Squares (GLS) estimator of Δx_t based on the past, present and future of Δy_t is

$$\Delta y_t^* = \frac{\sum_{i=1}^N F_i^{-1}(L) \Delta y_{it}}{\sum_{i=1}^N F_i^{-1}(L)} = \Delta x_t + \frac{\sum_{i=1}^N F_i^{-1}(L) \Delta v_{it}}{\sum_{i=1}^N F_i^{-1}(L)}.$$

[Fiorentini and Sentana \(2019\)](#) show that Δy_t^* , a one-dimensional linear filter applied to Δy_t , contains all relevant information about Δx_t in Δy_t , in the sense that the application of the Kalman filter to Δy_t^* delivers the same predictions for Δx_t as the Kalman filter applied to Δy_t . We denote the resulting error by Δv_t^* ,

whose spectral density is given by $f_*(\lambda) = \left(\sum_{i=1}^N f_i^{-1}(\lambda) \right)^{-1}$.

[Fiorentini and Sentana \(2019\)](#) also derive the frequency-domain analogue to Δy_t^* , namely

$$\sum_{t=-\infty}^{\infty} \Delta y_t^* e^{i\lambda t} = \Delta y^*(\lambda) = \Delta x(\lambda) + \frac{\sum_{i=1}^N f_i^{-1}(\lambda) \Delta v_i(\lambda)}{\sum_{i=1}^N f_i^{-1}(\lambda)} = \Delta x(\lambda) + \Delta v^*(\lambda).$$

In this context, we take

$$R^2(\lambda) = \frac{f_0(\lambda)}{f_0(\lambda) + f_*(\lambda)},$$

i.e., the fraction of the variance of $\Delta y^*(\lambda)$ explained by $\Delta x(\lambda)$, as an indicator of the degree of observability of the signal at frequency λ . This measure is useful to gauge the frequencies at which the effect of misspecification in measurement errors is more severe for inferences about Δx_t . In addition, we can obtain an overall measure of observability of the signal by simply replacing spectral densities by their integrals over $[0, 2\pi]$, which yields

$$R^2 = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_*^2},$$

since $\sigma_0^2 = \int_0^{2\pi} f_0(\lambda) d\lambda$ and $\sigma_*^2 = \int_0^{2\pi} f_*(\lambda) d\lambda$. Interestingly, this is the usual measure of observability (or signal strength) in the error-in-variable model

$$\frac{\sum_{i=1}^N \sigma_i^{-2} \Delta y_{it}}{\sum_{i=1}^N \sigma_i^{-2}} = \Delta x_t + \frac{\sum_{i=1}^N \sigma_i^{-2} \Delta v_{it}}{\sum_{i=1}^N \sigma_i^{-2}},$$

and coincides with $R^2(\lambda)$ at each frequency λ when Δv_{it} is white noise for each i . Thus, a “more observable” signal is indicated by $R^2(\lambda)$ and R^2 closer to unity.⁹ In particular, when one of the measurement error variances is zero, $R^2(\lambda) = 1$ and $R^2 = 1$.

3. ESTIMATION

Let $\hat{\theta}$ be the Gaussian MLE of θ . We give asymptotics for $\hat{\theta}$ in the following setup:

Assumption 2. As $T \rightarrow \infty$, the parameters $\mu_0, \rho_0, \sigma_0, \rho_{1:N}, \sigma_{1:N}$ are held constant.

Remark. An alternative local embedding in which parameters drift in a $1/\sqrt{T}$ -neighborhood of a fixed value can be used with little change as long as the autoregressive roots $\rho_{0:N}$ are bounded away from unity. To keep the exposition focused, though, we do not allow for local-to-unity asymptotics for the persistence of measurement errors. A setup in which $\rho_i = 1 - \varrho_i/T$ with ϱ_i held fixed would capture a situation in which the researcher is uncertain about imposing cointegration because the probability that a unit-root test on the differences $y_{it} - y_{jt}$ rejects the null remains bounded between 0 and 1 as $T \rightarrow \infty$ (see, e.g., Cavanagh, 1985; Chan & Wei, 1987; Phillips, 1987). Still, our analysis suggests that the difference between unit roots and near unit roots is very relevant for constructing inference for long-run objects (see the [supplemental material](#)) but not so for estimation.

Our main estimation result, whose proof appears in the Appendix, is as follows:

Theorem 1. Let $\{\hat{\mu}, \tilde{\sigma}, \tilde{\psi}_{1:N}\}$ be the maximum likelihood estimator from the static model (i.e., the model that assumes (a), (b) with $\rho = 0$, and (c)). Similarly, let $\{\hat{\mu}, \hat{\rho}, \hat{\sigma}, \hat{\psi}_{1:N}\}$ be the maximum likelihood estimator from the dynamic model \mathcal{P} . Then, under Assumptions 1 and 2,

$$\sqrt{T} \begin{pmatrix} \hat{\mu} - \tilde{\mu} \\ \hat{\sigma} - \tilde{\sigma} \\ \hat{\psi}_{1:N} - \tilde{\psi}_{1:N} \end{pmatrix} = o_p(1).$$

Furthermore, for some B and V ,

$$\sqrt{T}(\hat{\rho} - (\rho_0 + B)) \Rightarrow N(0, V).$$

Therefore, $\hat{\mu} \xrightarrow{p} \mu_0$, $\hat{\sigma} \xrightarrow{p} \sigma_0$ and $\hat{\psi}_i \xrightarrow{p} \sigma_i$ for all i . The estimators of the unconditional mean and variance parameters of the latent variables obtained from the

static and dynamic models are asymptotically normal at the usual rate,¹⁰ and, perhaps more surprisingly, they have the same asymptotic covariance matrix. The consequences of neglecting the common trend are, thus, confined to the auto-correlation structure of Δx_t . A univariate example in the supplemental material provides further intuition.

Theorem 1 has many implications. First, one can estimate the model parameters without loss of asymptotic precision in two steps: maximizing the static model log-likelihood for $\{\mu, \sigma, \psi_{1:N}\}$ initially, and then the dynamic log-likelihood for ρ after plugging in $\{\tilde{\mu}, \tilde{\sigma}, \tilde{\psi}_{1:N}\}$.¹¹ Second, the unconditional R^2 measure of signal observability is consistently estimated even if the model is misspecified, unlike its frequency-domain counterpart. Third, the estimator of ρ_0 will typically be inconsistent, and at least when normality holds, it will display higher asymptotic variance than the estimator from the model that correctly imposes the common trend in levels.

We can implicitly characterize the inconsistency term B by means of the spectral condition

$$\int_0^{2\pi} \cos(\lambda) \left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2} \right)^2 (f_{\bar{\vartheta}}(\lambda) - f_0(\lambda) + \sigma_*^2 - \tilde{f}(\lambda)) d\lambda = 0, \quad (2)$$

where $\bar{\vartheta} = (\mu_0, \rho_0 + B, \sigma_0)$, σ_*^2 is defined in Section 2, and $\tilde{f}(\lambda) = \sum_{i=1}^N \sigma_i^{-4} f_i(\lambda) / [\sum_{i=1}^N \sigma_i^{-2}]^2$ is the spectrum of $\sum_{i=1}^N \sigma_i^{-2} \Delta v_{it} / \sum_{i=1}^N \sigma_i^{-2}$, i.e., the true error in the GLS minimal sufficient statistic for Δx_t computed under the misspecified model.

In the Appendix, we derive a time-domain counterpart to (2), namely:

$$\text{Cov}(\mathbb{E}_{\bar{\theta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]) = \text{Cov}_{\bar{\theta}}(\mathbb{E}_{\bar{\theta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]). \quad (3)$$

This means that B adjusts to match two types of covariances between the smoothed values of Δx_{t-1} and Δx_t obtained with the misspecified model: the covariance taken under the data generating process \mathbb{P} and the covariance computed using the misspecified model $\mathbb{P}_{\bar{\theta}}$ evaluated at the pseudo-true value $\bar{\theta}$.

When either $\rho_i = 1$ for all i or $\sigma_i = 0$ for at least one i , we have that $\tilde{f}(\lambda) = \sigma_*^2$ for all λ . As a consequence, one can set $f_{\bar{\vartheta}} = f_0$, which implies a consistent estimator of ρ_0 with $B = 0$. By continuity, the inconsistency term B will be small when the extent of misspecification is small (ρ_i 's all close to unity) or when the observability of the signal is high (R^2 close to unity). In contrast, noticeable biases may arise when one moves away from those limiting cases, as we illustrate in the next section.

AR(p) dynamics. Our approach can be easily extended to a model in which Δx_t is an AR(p) process with unconditional mean μ_0 , unconditional variance σ_0^2 , and AR coefficients $\varrho_{0,1:p}$ and the estimated model correctly specifies the dynamics of Δx_t . Under Assumptions 1(A) and 1(C), the asymptotic equivalence between $\hat{\mu}, \hat{\sigma}, \hat{\psi}_{1:N}$ and $\tilde{\mu}, \tilde{\sigma}, \tilde{\psi}_{1:N}$ of Theorem 1 remains, and so does asymptotic normality

of $\sqrt{T}(\hat{\varrho}_{l:p} - \bar{\varrho}_{l:p})$ where $\hat{\varrho}_{l:p}$ is the MLE of $\varrho_{l:p}$ and $\bar{\varrho}_{l:p}$ is the pseudo-true value. These are characterized by a set of spectral conditions analogous to (2), namely,

$$\int_0^{2\pi} \cos(\ell\lambda) \left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2} \right)^2 \left(f_{\bar{\vartheta}}(\lambda) - f_0(\lambda) + \sigma_*^2 - \tilde{f}(\lambda) \right) d\lambda = 0, \ell = 1, \dots, p,$$

and a set of time-domain conditions analogous to (3),

$$\text{Cov}(\mathbb{E}_{\bar{\vartheta}}[\Delta x_{t-\ell} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\vartheta}}[\Delta x_t | \Delta y_{-\infty:\infty}]) = \text{Cov}_{\bar{\vartheta}}(\mathbb{E}_{\bar{\vartheta}}[\Delta x_{t-\ell} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\vartheta}}[\Delta x_t | \Delta y_{-\infty:\infty}])$$

for $\ell = 1, \dots, p$. Some numerical experiments (available upon request) suggest that in this case the roots $\bar{\phi}_{l,p}$ that satisfy $\prod_{\ell=1}^p (1 - \bar{\phi}_{\ell} z) = (1 - \bar{\varrho}_1 z - \dots - \bar{\varrho}_p z^p)$ are subject to downward bias relative to the true roots $\phi_{0,l:p}$ that satisfy $\prod_{\ell=1}^p (1 - \phi_{0,\ell} z) = (1 - \varrho_{0,1} z - \dots - \varrho_{0,p} z^p)$.

3.1. Numerical and Simulation Evidence

We complement our foregoing discussion of estimation with some insights from numerical and simulation calculations. To begin with, we compute expression (2) by numerical quadrature to obtain the inconsistency in the estimation of ρ_0 as a function of the observability of the signal and the severity of overdifferencing. We set $\mu_0 = 3$, $\rho_0 = 0.5$ and $\sigma_0 = 3.25$.¹² We also take $N = 2$ and let R^2 (with $\sigma_1 = \sigma_2$) and $\rho_1 = \rho_2$ vary over the interval $(0, 1)$.

We display the results for this exchangeable design in Fig. 2. They clearly confirm our intuition about the roles of ρ_i and R^2 in determining B , with the inconsistency growing quickly as R^2 decreases below 0.5 even for moderate amounts of overdifferencing. Importantly, we always find that $B \leq 0$ under the form of misspecification we analyze in this chapter. The rationale is as follows. Equation (2) shows that $\rho_0 + B$ is set to match a weighted average of the difference between $f_{\bar{\vartheta}}$ and $f_0 + \sigma_*^2 - \tilde{f}$, which is depressed at lower frequencies compared to the true spectrum f_0 by the effect of overdifferencing. To see this, note that since $f_{\bar{\vartheta}}$ is an AR(1) spectrum, lower values of $f_{\bar{\vartheta}}$ at low frequencies with σ fixed at σ_0 require decreasing ρ . In particular, at frequency $\lambda = 0$, we have $f_{\bar{\vartheta}}(0) = (1 + \rho)/(1 - \rho)$ which decreases with ρ and, more generally,

$$\frac{\partial f_{\bar{\vartheta}}(\lambda)}{\partial \rho} = -2f_{\bar{\vartheta}}(\lambda) \left[\frac{\rho}{1 - \rho^2} + \frac{\rho - \cos(\lambda)}{1 + \rho^2 - 2\rho \cos(\lambda)} \right],$$

which is negative for low values of λ . Hence, $\text{plim}_{T \rightarrow \infty} \hat{\rho} = \rho_0 + B < \rho_0$.¹³

We next present simulation evidence on the finite-sample properties of the following three estimators of θ : (i) maximum likelihood for the model in first differences (i.e., $\hat{\theta}$), (ii) the two-step procedure suggested by Theorem 1, and (iii) maximum likelihood for the model in levels. The results are summarized in Tables 1–3. They show that the approximation in Theorem 1 works very well in realistic sample sizes and setups. The correlation between $\hat{\theta}$ and the two-step estimator is virtually one, as one would expect from their asymptotic equivalence,

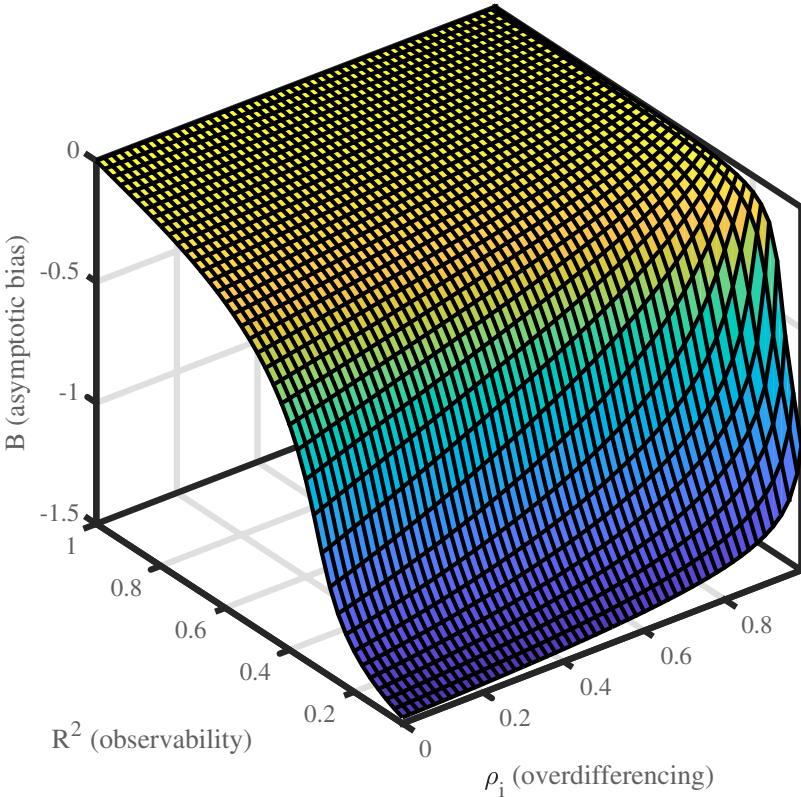


Fig. 2. Asymptotic Bias. *Notes:* Numerical computation of asymptotic bias B in the estimation of ρ_0 for different extents of overdifferencing $\rho_1 = \rho_2$ and signal observability R^2 . The true value is $\rho_0 = 0.5$. The integral in (2) is approximated by numerical quadrature with a fine grid on the interval $[0, 2\pi]$.

and the inconsistency in $\hat{\rho}$ is close to the values for B obtained from equation (2). Not surprisingly, the model in levels outperforms its competitors, although not by much for unconditional moments. The results for a second symmetric design in which $\rho_1 = \rho_2 = 0$ and for an asymmetric design, which we present in the supplemental material, display the same patterns.

Remark. The behavior of B as ρ_i approaches unity for fixed R^2 can be obtained from Fig. 2. Our calculations suggest that $\sqrt{T} | B | = o(1)$ when $\rho_i = 1 - \varrho_i / T$ for all i , and that $\sqrt{T} | B | = O(1)$ would require the alternative embedding $\rho_i = 1 - \varrho_i / \sqrt{T}$ instead. Such an embedding would allow us to pretest the existence of a bias in the estimation of ρ_0 . Although we do not formally prove these statements, they convey a sense of the relevance of estimation biases in applications. Note that if Δx_t were observable, the standard error of $\hat{\rho}$ for a sample of seventy years of quarterly data ($T = 280$) would be roughly $\sqrt{(1 - \rho_0^2) / T} \approx 0.05$. If, for

Table 1. Monte Carlo Simulation for $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.30$.

	True	Differences	Two-step	Levels
μ_0				
Mean	3	3.001	3.001	3.002
Stderr		0.344	0.344	0.345
Corr			1	0.998
ρ_0				
Mean	0.5	0.289	0.292	0.481
Stderr		0.204	0.195	0.178
Corr			0.942	0.451
σ_0				
Mean	3.25	3.233	3.194	3.204
Stderr		0.558	0.588	0.598
Corr			0.96	0.825
ρ_i				
Mean	0.85			0.831
Stderr				0.072
σ_i				
Mean	7.021	6.995	7.01	6.996
Stderr		0.385	0.391	0.384

Notes: Number of samples is $n_{MC} = 2,000$, sample size is $T = 280$, and parameter values are given under column “True.” Rows “mean” and “stderr” show mean and standard deviation across simulations of each estimator; “corr” shows the correlation with MLE in differences of the other two estimators. The bias in $\hat{\rho}$ is to be compared with the theoretical inconsistency $B \approx -0.23$ computed from equation (2) as indicated in the text.

Table 2. Monte Carlo Simulation for $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.50$.

	True	Differences	Two-step	Levels
μ_0				
Mean	3	3.002	3.002	3.002
Stderr		0.341	0.341	0.341
Corr			1	0.999
ρ_0				
Mean	0.5	0.41	0.408	0.488
Stderr		0.111	0.11	0.105
Corr			0.986	0.876
σ_0				
Mean	3.25	3.229	3.22	3.221
Stderr		0.322	0.321	0.326
Corr			0.979	0.954
ρ_i				
Mean	0.85			0.822
Stderr				0.089
σ_i				
Mean	4.596	4.583	4.589	4.581
Stderr		0.266	0.272	0.268

Notes: Number of samples is $n_{MC} = 2,000$, sample size is $T = 280$, and parameter values are given under column “True.” Rows “mean” and “stderr” show mean and standard deviation across simulations of each estimator; “corr” shows the correlation with MLE in differences of the other two estimators. The bias in $\hat{\rho}$ is to be compared with the theoretical inconsistency $B \approx -0.08$ computed from equation (2) as indicated in the text.

Table 3. Monte Carlo Simulation for $\rho_1 = \rho_2 = 0.85$ and $R^2 = 0.85$.

	True	Differences	Two-step	Levels
μ_0				
Mean	3	3.001	3.002	3.001
Stderr		0.342	0.343	0.343
Corr			0.999	0.999
ρ_0				
Mean	0.5	0.478	0.475	0.49
Stderr		0.062	0.062	0.065
Corr			0.998	0.934
σ_0				
Mean	3.25	3.231	3.23	3.238
Stderr		0.196	0.196	0.205
Corr			0.999	0.934
ρ_i				
Mean	0.85			0.781
Stderr				0.24
σ_i				
Mean	1.931	1.925	1.926	1.914
Stderr		0.148	0.165	0.252

Notes: Number of samples is $n_{MC} = 2,000$, sample size is $T = 280$, and parameter values are given under column “True.” Rows “mean” and “stderr” show mean and standard deviation across simulations of each estimator; “corr” shows the correlation with MLE in differences of the other two estimators. The bias in $\hat{\rho}$ is to be compared with the theoretical inconsistency $B \approx -0.01$ computed from equation (2) as indicated in the text.

example, the data were generated from the common-trend model with parameters $(\rho_i, R^2) = (0.35, 0.85)$, $(\rho_i, R^2) = (0.92, 0.50)$ or $(\rho_i, R^2) = (0.98, 0.30)$, then the estimation of the model in differences would yield a bias of size comparable to the standard error. These values seem plausible for a large number of applications. In fact, when R^2 is 0.5 or below, values of ρ_i which are only slightly below unity can already cause severe downward bias in the estimation of ρ_0 .

4. SIGNAL EXTRACTION

In general, neglecting the common trend should negatively impact filtered and smoothed estimates of the latent variables. We can identify two channels through which this happens: one important for short-run calculations, and the other for long-run calculations. We begin with the short-run channel, i.e., the downward bias in $\hat{\rho}$.

Consider the filtered estimate of Δx_t ,

$$\Delta \hat{x}_t = \mathbb{E}_{\hat{\theta}}[\Delta x_t | \Delta y_{1:T}].$$

As is well-known, the filtering error $\Delta \hat{x}_t - \Delta x_t$ is $O_p(1)$ for large T . This is in contrast to the estimation error $\hat{\theta} - \bar{\theta}$, with $\bar{\theta} = (\mu_0, \rho_0 + B, \sigma_0, \sigma_{1:N})'$, which is $o_p(1)$. Therefore, we can obtain a good approximation to the behavior of $\Delta \hat{x}_t - \Delta x_t$ if we simply abstract from estimation uncertainty and focus on filtered estimates at pseudo-true values,

$$\Delta \bar{x}_t = \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{1:T}] = \mu_0 + \sum_{\tau=1}^T \bar{\phi}_{\tau,T}(\Delta y_\tau - \mu_0 \mathbf{1}_{N \times 1}),$$

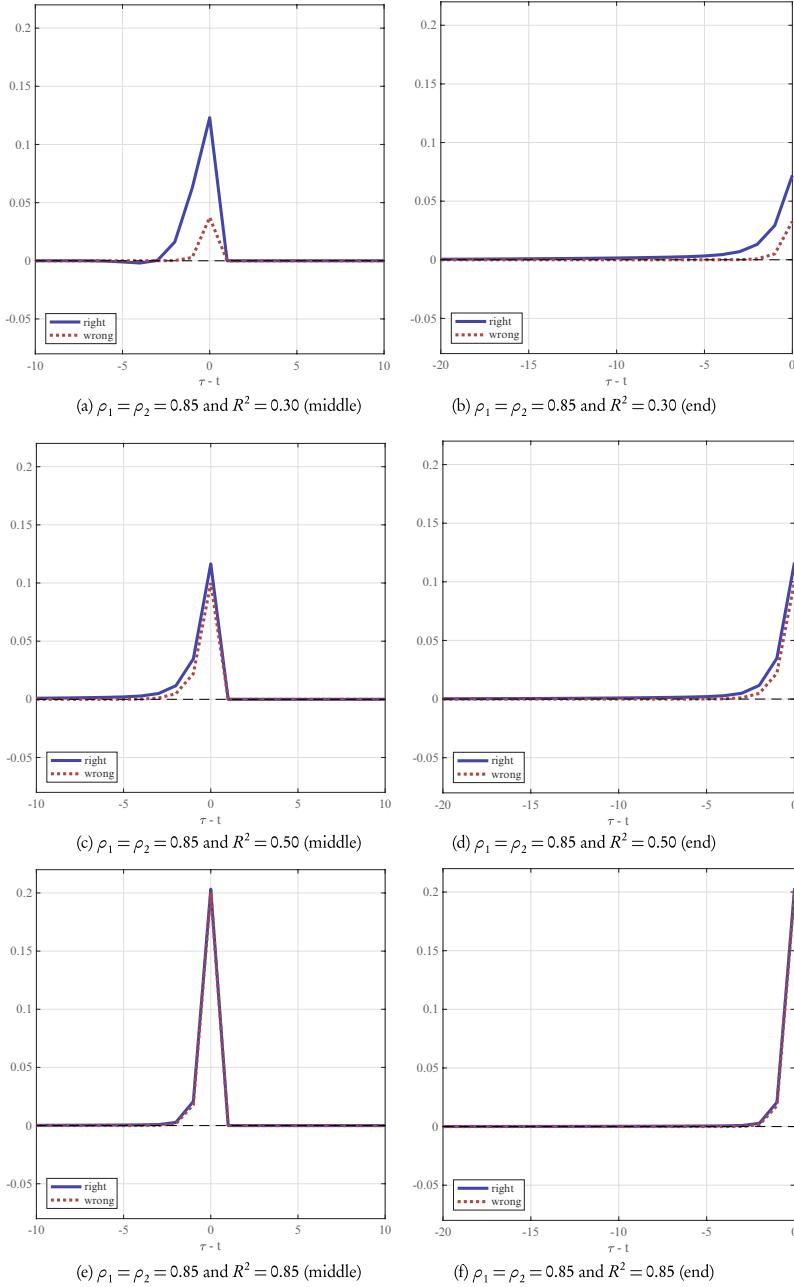


Fig. 3. Weights of Kalman Smoother. Notes: Horizontal axis is $\tau - t$; vertical axis is first entry of $\bar{\phi}_{\tau,T}$ (red) and $\phi_{\tau,T}^*$ (blue). Panels (a), (c) and (e) display weights for $t \approx T/2$ (Middle), and panels (b), (d) and (f) for $t = T$ (end). The filters are computed using $\mu_0 = 3$, $\rho_0 = 0.50$, $\sigma_0 = 3.25$. Wrong filter uses $\rho_0 + B$ as AR root with B computed from (2).

where the conditional expectation is affine because of the normality assumptions in 1(B) and 1(C).

On the other hand, the ideal filter from a mean-square error perspective is the conditional mean under the correctly specified model \mathbb{P} ,¹⁴

$$\Delta x_t^* = \mathbb{E}[\Delta x_t | \Delta y_{1:T}] = \mu_0 + \sum_{\tau=1}^T \phi_{\tau,T}^*(\Delta y_\tau - \mu_0 \mathbf{1}_{N \times 1}).$$

The discrepancy between the weights $\bar{\phi}_{1:T,T}$ and $\phi_{1:T,T}^*$ is of interest because we can decompose $\Delta \bar{x}_t - \Delta x_t$ into two orthogonal components: (i) the optimal filtering error $\Delta x_t^* - \Delta x_t$, whose variance cannot be reduced any further in the class of measurable functions of $\Delta y_{1:T}$ with bounded second moments, and (ii) the difference between the optimal and suboptimal filters $\Delta \bar{x}_t - \Delta x_t^*$.

To illustrate the consequences for signal extraction of neglecting the common trend in levels, Fig. 3 provides a comparison of the weights for our baseline calibration when overdifferencing is not so severe ($\rho_1 = \rho_2 = 0.85$) and the degree of observability varies from low ($R^2 = 0.30$) to high ($R^2 = 0.85$). We do so for the two leading signal extraction exercises encountered in practice: the computation of $\Delta \bar{x}_t$ and Δx_t^* for values of t in the middle of the sample, and for $t = T$ (i.e., “nowcasting”).

In both cases, it is clear that the filters from misspecified models tend to assign lower weights to nearby observations relative to what is optimal, the difference being larger the lower R^2 is. For the most part, this is explained by the fact that the suboptimal filters assume the signal to be less persistent than it actually is, as B is negative and grows in absolute value as R^2 decreases. Intuitively, the negative value of B resulting from neglecting the common trend leads the econometrician to underestimate the information content of current data.

Naturally, when overdifferencing is more severe, so is its impact on signal extraction. To support this claim, the supplemental material shows an analogous weight comparison in a design with $\rho_1 = \rho_2 = 0$. For a given R^2 , more severe overdifferencing means a larger downward bias in the estimation of the persistence of the signal, which implies even more depressed weights for informative nearby observations.

4.1. Long-Run Objects

By a long-run object we mean a weighted average $X = \sum_{t=1}^T \omega_t \Delta x_t$, where the weights $\omega_{1:T}$ satisfy $\|\omega_{1:T}\| = \sqrt{\sum_{t=1}^T \omega_t^2} = O(1/\sqrt{T})$. They are suitable to quantify trends in aggregate quantities and therefore regularly show up in empirical studies of growth. Compared to smoothed estimates of short-run objects, neglecting the common trend in the level measurements affects inferences about long-run objects through a different channel, namely, by inflating measures of their uncertainty, such as standard errors or confidence intervals. This can be appreciated in the comparison offered in Fig. 6 in the empirical analysis. We provide a theoretical discussion of this phenomenon in the supplemental material.

4.2. Simulation Evidence (Continued)

We compare the finite-sample behavior of the filters discussed above using the same simulation designs as in Section 3.1 (see Tables 1–3). In each simulated sample, we first obtain maximum likelihood estimates of both the misspecified and correctly specified models, and then we compute the corresponding smoothed estimates of Δx_t for $t \approx T/2$ and $t = T$. We present the results for those designs that set $\rho_1 = \rho_2 = 0.85$ in Tables 4 and 5. The supplemental material reports additional results setting $\rho_1 = \rho_2 = 0$ and $\rho_1 = 0, \rho_2 = 0.95$.

As a general rule, $T = 280$ seems large enough for $\Delta \bar{x}_t$ to provide a good approximation to $\Delta \hat{x}_t$. The same is true for Δx_t^* and the filter from the correctly specified model evaluated at the maximum likelihood estimates, which we call $\Delta \hat{x}_t^*$. The main differences in precision appear between $\Delta \bar{x}_t$ and Δx_t^* rather than between the filters evaluated at the ML estimates and their limiting values.

Table 4. Monte Carlo Simulation for $\rho_1 = \rho_2 = 0.85$ and $t \approx T/2$.

	$\Delta \hat{x}_t$	$\Delta \bar{x}_t$	$\Delta \hat{x}_t^*$	Δx_t^*
$R^2 = 0.30$				
RMSE	3.15	3.18	3.13	3.12
Increase	0.04	0.04	0.02	
$R^2 = 0.50$				
RMSE	3.05	3.05	3.04	3.03
Increase	0.02	0.01	0.02	
$R^2 = 0.85$				
RMSE	2.88	2.88	2.87	2.88
Increase	0.01	0	0.01	

Notes: Number of samples is $n_{MC} = 2,000$ and sample size is $T = 280$. Columns “ $\Delta \hat{x}_t$ ” and “ $\Delta \bar{x}_t$ ” refer to the wrong filter at the ML estimates and pseudo true values, respectively. Columns “ $\Delta \hat{x}_t^*$ ” and “ Δx_t^* ” refer to the right filter at the ML estimates and true values, respectively. Root MSE and increase in MSE as a fraction of the MSE of Δx_t^* are indicated for each filter and R^2 .

Table 5. Monte Carlo Simulation for $\rho_1 = \rho_2 = 0.85$ and $t = T$.

	$\Delta \hat{x}_t$	$\Delta \bar{x}_t$	$\Delta \hat{x}_t^*$	Δx_t^*
$R^2 = 0.30$				
RMSE	3.12	3.14	3.1	3.08
Increase	0.04	0.04	0.02	
$R^2 = 0.50$				
RMSE	3.02	3.01	3.01	3
Increase	0.02	0.01	0.02	
$R^2 = 0.85$				
RMSE	2.87	2.87	2.87	2.87
Increase	0.01	0	0.01	

Notes: Number of samples is $n_{MC} = 2,000$ and sample size is $T = 280$. Columns “ $\Delta \hat{x}_t$ ” and “ $\Delta \bar{x}_t$ ” refer to the wrong filter at the ML estimates and pseudo true values, respectively. Columns “ $\Delta \hat{x}_t^*$ ” and “ Δx_t^* ” refer to the right filter at the ML estimates and true values, respectively. Root MSE and increase in MSE as a fraction of the MSE of Δx_t^* are indicated for each filter and R^2 .

The effect of neglecting the common trend when measurement errors are highly persistent seems modest in our simulations, with an increase of at most 7% in root MSE relative to the optimal filter in low- R^2 designs. However, more severe overdifferencing combined with a low R^2 leads to a substantial reduction in the precision of filters, as the supplemental material illustrates. Therefore, researchers should be particularly concerned about their modeling assumptions on measurement error when the R^2 measure we propose in the chapter is 0.5 or less, something we already saw in the estimation results.

It is interesting to note that the nowcasting estimate $\Delta\hat{x}_T$ is less affected by misspecification than the smoothed estimate for an observation in the middle of the sample, as one would expect given that the sample is relatively less informative (and therefore receives smaller weights) when filtering Δx_T .

5. AN IMPROVED AGGREGATE OUTPUT MEASURE

In this section, we apply our framework to the US quarterly GDE and GDI data displayed in Fig. 1 with the objective of constructing a new improved measure of economic activity. Specifically, we use the November 2020's release of BEA national accounts estimates for the period 1952Q1-2019Q4 and define $y_{1t} = 400 \times \ln(\text{GDE}_t)$ and $y_{2t} = 400 \times \ln(\text{GDI}_t)$ so that their first differences Δy_{1t} and Δy_{2t} indicate the annualized (geometric) growth rates. The statistical discrepancy, which we compute as $d_{12,t} = (y_{1t} - y_{2t}) / 4 = 100 \times \ln(\text{GDE}_t / \text{GDI}_t)$, is then roughly the percentage by which GDE exceeds GDI in levels.¹⁵ Remarkably, the levels of GDE and GDI have remained within 3% of each other for about 70 years, lending strong support to our claim that the two measurements are cointegrated.

Table 6 reports maximum likelihood estimates for both the parameters of the common trend model \mathbb{P} and the statistical model \mathcal{P} we discussed in Section 3. As expected from Theorem 1, there is no significant disagreement between different estimates of the unconditional moment parameters $\{\mu_0, \sigma_0, \sigma_1, \sigma_2\}$. The estimates of our R^2 measure of common trend observability are high at about 0.92, with a small confidence interval around them. Estimates for ρ_0 , in turn, are all near 0.5, with a seemingly small downward bias in the estimators from the models that neglect the common trend. These patterns are in line with the theoretical and simulation analysis in Section 3.¹⁶ The estimates of the autoregressive coefficient ρ_2 implies that the time series of GDI's measurement error in levels, v_{2t} , seems stationary but highly persistent. In contrast, we cannot reject that the GDE's measurement error in levels, v_{1t} , is white noise. This difference in the persistence of measurement errors may be the result of the fact that GDE and GDI are computed from different sources and with different methods, therefore relying on inputs which themselves differ in their dynamics. Next, we discuss some of the implications of these results.

Our first consideration is about the serial dependence of the statistical discrepancy, $d_{12,t}$. Fig. 4, in particular, shows that our assumption of AR(1) measurement errors in levels does a good job at replicating the autocorrelations of this variable. Although the statistical discrepancy is highly persistent because the

Table 6. Estimates of Model Parameters for US Data.

	Differences	Two-step	Levels
μ_0			
Estimate	2.994	2.997	2.989
Stderr	0.338	0.204	0.338
ρ_0			
Estimate	0.488	0.485	0.499
Stderr	0.057	0.048	0.057
σ_0			
Estimate	3.237	3.227	3.223
Stderr	0.186	0.151	0.184
ρ_1			
Estimate			-0.097
Stderr			0.272
σ_1			
Estimate	1.49	1.387	1.314
Stderr	0.115	0.149	0.130
ρ_2			
Estimate			0.941
Stderr			0.021
σ_2			
Estimate	1.113	1.239	1.338
Stderr	0.137	0.162	0.113

Notes: The sample period is 1952Q1–2019Q4 ($T = 271$). Rows “estimate” and “stderr” show point estimate and standard error for each estimator. A subindex 1 in ρ_i and σ_i refers to GDE while a subindex 2 refers to GDI. The point estimate for signal observability R^2 is 0.922 and a 95% confidence interval for R^2 is [0.901, 0.943].

GDI’s measurement error in level dominates it, it is also evident that the serial dependence steadily declines, being already fairly low after 12 quarters. We also note that, if anything, the autocorrelations of the statistical discrepancy tend to decrease faster in the data than in our model, although the difference is small relative to the sampling uncertainty.

A related observation is that the model in first differences leads to an implausibly high probability of long-run divergence between GDE and GDI in levels. To get a sense for it, $d_{12,T} | d_{12,0} \sim N(d_{12,0}, 0.9 \times T)$ with the estimates of the dynamically misspecified model at hand. A quick calculation indicates that the probability that today we would observe a divergence between GDE and GDI higher than 3% is 0.99 if the two aggregate output measurements were not cointegrated.

The second consideration refers to the impact of neglecting the common trend in levels on inferences about parameters and latent variables. Because Δx_t is highly observable, our theoretical results in Sections 3 and 4 lead us to expect no significant divergence between the models in differences and in levels with regards to maximum likelihood estimates and smoothed estimates of what we have called short-run objects. We have already confirmed the similarity of the estimates in Table 6. In turn, Fig. 5 confirms our results for the smoothed estimates of Δx_t , as one can hardly

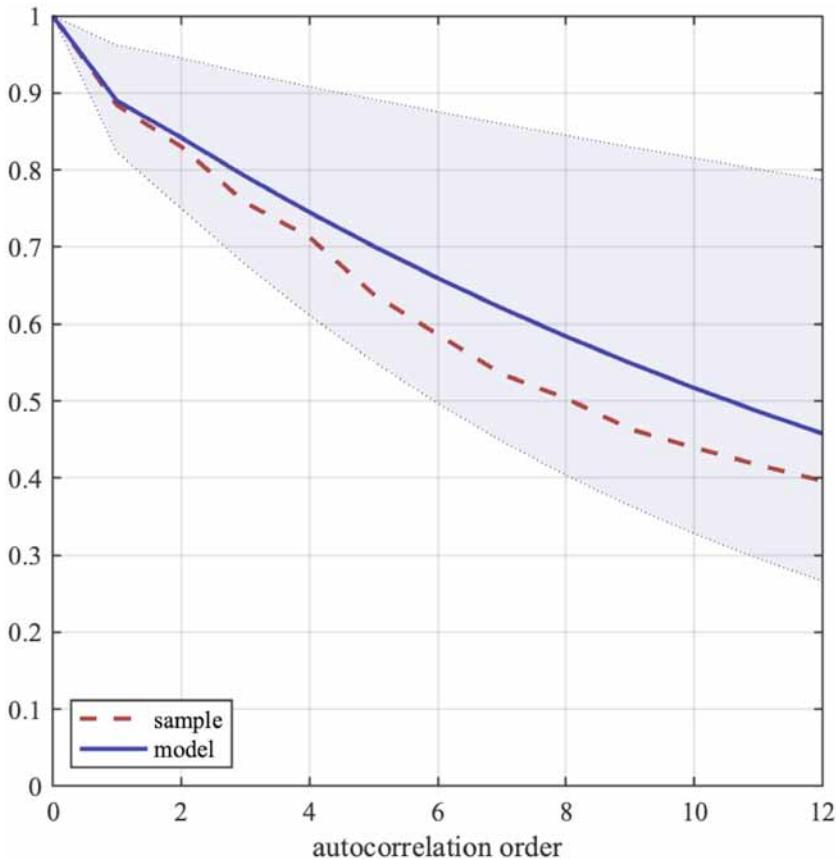


Fig. 4. Autocorrelations of the Statistical Discrepancy. Notes: The solid blue line contains the autocorrelations implied by model \mathbb{P} at point estimates. Shaded area is a pointwise 95% confidence interval for each lag.

distinguish one model from the other in terms of the conditional mean and variance of Δx_t , given $\Delta y_{1:T}$.¹⁷ Still, pointwise confidence intervals are shorter for the model that imposes the common trend: their average length is 3.5% (in annualized growth) for the model in differences against 3% for the model in levels.

However, the fact that GDE's measurement error is essentially white noise does affect inferences about long-run objects. Fig. 6 illustrates this feature with the 8-year moving averages of Δx_t . We take overlapping 8-year intervals for the purposes of averaging out the typical business-cycle periodicity.¹⁸ As expected from our results in the supplemental material, there is substantially less uncertainty for the model that exploits the common trend. Specifically, the average length of the confidence intervals is 1% for the model in differences and 0.2% for the model in levels. This reduced uncertainty is particularly important for assessing changes in aggregate trends, as such changes are typically small.

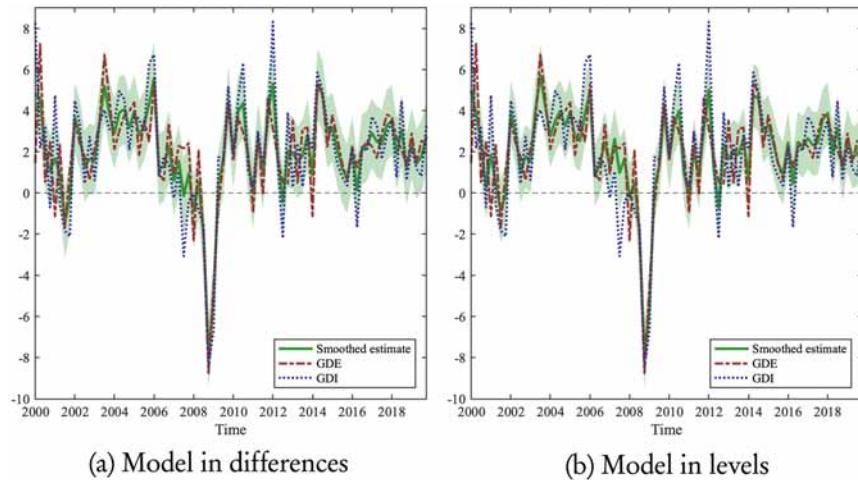


Fig. 5. Smoothed Estimates of Δx_t (Short-Run Object). Notes: The solid green line represents the smoothed estimates and the shaded area represents 95% confidence intervals (pointwise for each t).

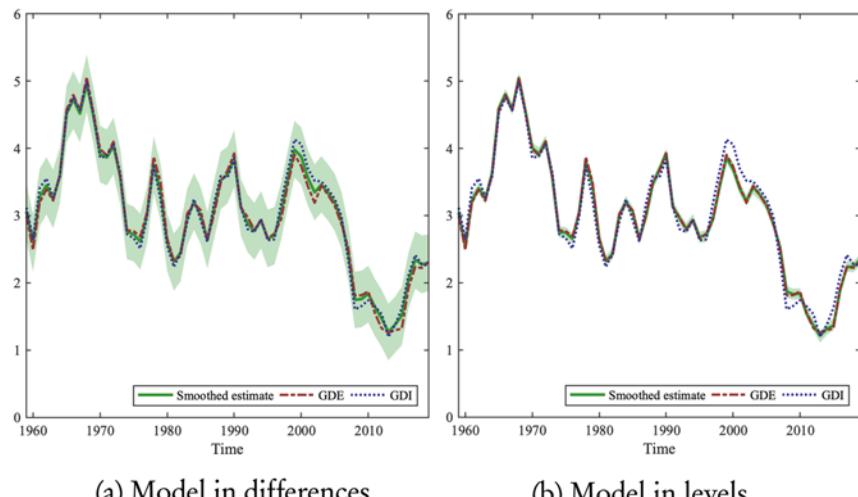


Fig. 6. Smoothed Estimates of $h^{-1} \sum_{\ell=1}^h \Delta x_{t-\ell+1}$ ($h = 32$, Long-Run Object).
 Notes: Solid green line represents the smoothed estimates and shaded area represents 99% confidence intervals (pointwise for each t).

Finally, note that our trend estimates track far more closely 8-year moving averages of GDE growth than those of GDI, which is again explained by the low persistence of $v_{l,t}$. Therefore, we may conclude that empirical patterns about economic activity previously obtained from low-frequency averages of GDE are robust to the presence of measurement error in view of the small degree of filter uncertainty implied by our model.

6. CONCLUSION

From a practical point of view, the first lesson we can extract from our study regarding aggregate output measurement is that the need to account for a common trend in levels hinges critically on how important measurement errors are in driving observable variation. For quarterly or annual data, measurement errors might be small; for monthly and, particularly, for high-frequency data the opposite should be expected. Although no direct measure of economic activity exists at the monthly frequency, nowcasting exercises at high frequencies typically contain larger amounts of measurement errors as they feed on noisier, more preliminary, input data. The nature of what is being measured matters too as different economic concepts have different associated degrees of noisiness. For example, it is not the same to look at the quarterly growth rate of GDP than to look at its quinquennial counterpart. As a practical prescription, we recommend the estimation of the R^2 measure of trend observability we develop in the chapter. We also strongly urge researchers to always impose a common trend, especially when the R^2 turns out low – an R^2 below 0.5 should be cause of concern.

Moreover, our econometric analysis yields several insights which are of theoretical interest. First, we prove that estimators of unconditional first and second moments under the misspecified model are asymptotically equivalent to static model estimators. Second, we show that the form of misspecification studied in this chapter causes a downward bias in the estimated persistence of the signal. And third, we highlight that the misspecified model will tend to overstate uncertainty of smoothed estimates of latent variables, and dramatically so for long-run objects. Although we derive these results in a simplified parametric model, our methods of analysis allow easy extension to more general setups. In particular, our analysis may be adapted to dynamic factor models with nontrivial cross-sectional dimensions – models in which the usual data preprocessing may likely lead to overdifferencing.

On the empirical side, we construct a new improved measure of US aggregate output from GDE and GDI data. Unlike existing signal-extraction measures, ours allows GDE and GDI's measurement errors in levels to mean-revert, a property that fits well with the data. Still, given that signal observability is high in this application, our estimates of the parameters of the dynamics of output growth are not affected much by ignoring the stationarity of the measurement errors. Nevertheless, our common trend approach delivers noticeable reductions in the implied uncertainty of smoothed estimates of true output growth. Specifically, measured in terms of root mean square errors, the reductions are around 15% for

short-run objects and 80% for long-run ones. One important practical issue that we have neglected in this chapter is the regular updating of the GDE and GDI measures by the BEA. In [Almuzara et al. \(2022\)](#), we are currently exploring this relevant research avenue within the common trends framework of this chapter.

NOTES

1. See [Grimm \(2007\)](#) for a detailed methodological insight.
2. [Stone et al. \(1942\)](#) is the first known reference to the signal-extraction framework of our chapter. Early literature is surveyed in [Weale \(1992\)](#). See also [Smith et al. \(1998\)](#).
3. In unreported simulation experiments, we explore the possibility that biases in parameter estimators may be reduced by means of a flexible model of the serial dependence structure of measurement errors in first differences. Specifically, we model Δv_t as a set of independent univariate AR(p) models with p large. Our analysis suggests that bias reduction is thus possible, but at the expense of significant precision loss. Large- p , large- T double-asymptotics in this context appear to be an interesting (but challenging) avenue for future research.
4. Another difference with [Chang et al. \(2009\)](#) is that their baseline analysis assumes a random walk common trend while ours assumes it to be ARI(1,1).
5. Interestingly, [Chang et al. \(2009\)](#) show that there is cointegration between the true series x_t and the smoothed estimates that exploit the correctly specified model, which reinforces the case for imposing the common trend in order to obtain a filter capable of closely tracking the level of the signal.
6. This is most probably related to the low power attributed to cointegration tests.
7. There are $N(N - 1)/2$ statistical discrepancies but only $N - 1$ of them are linearly independent. For example, all the discrepancies with respect to a fixed measurement j form a basis of the cointegration space. An error-correction representation can be derived along the lines of [Chang et al. \(2009\)](#).
8. That is, $F_j(e^{\lambda}) = f_j(\lambda)$.
9. As an alternative, one could use the signal-noise ratios $q(\lambda) = f_0(\lambda) / f_*(\lambda)$ and $q = \sigma_0^2 / \sigma_*^2$. Nevertheless, R^2 -type measures are easier to interpret because they are bounded between 0 and 1.
10. If the loadings on the common trend x_p , which we assume are $1_{N \times 1}$, had to be estimated instead, the results in [Chang et al. \(2009\)](#) imply that a linear combination of them would converge to a nonstandard distribution at the rate T along a direction determined by the cointegration space.
11. In fact, our proof suggests that the asymptotic equivalence between static and dynamic MLEs would survive in the presence of forms of dynamic misspecification other than the one we consider in this chapter, and for more general dynamic models when the latent variables follow autoregressive processes but not when they have moving average components.
12. Since they represent an affine transformation of the data, the parameters μ_0 and σ_0 (given R^2) are irrelevant for both B and the finite-sample behavior of the ML estimators. Nevertheless, we choose the values of these parameters to match estimates from US quarterly data on economic activity for the period 1952Q1–2019Q4, so that our simulated data resembles the actual dataset in our empirical application. Other sample periods usually lead to different estimates of μ_0 and σ_0 but leave ρ_0 and measures of overdifferencing and signal observability practically unchanged.
13. Another way to look at the fact that $B \leq 0$ is that over-differencing makes f_i close to zero for frequencies λ which are near zero (f_i is real and continuous), and so

$$\tilde{f}(\lambda) = \frac{\sum_{i=1}^N \sigma_i^{-4} f_i(\lambda)}{\left[\sum_{i=1}^N \sigma_i^{-2} \right]^2} < \frac{\sum_{i=1}^N \sigma_i^{-4}}{\left[\sum_{i=1}^N \sigma_i^{-2} \right]^2} = \sigma_*^2$$

for small λ . As a result, $f_0(\lambda) + [\tilde{f}(\lambda) - \sigma_*^2] < f_0(\lambda)$ for small λ . A pseudo-true value $\rho_0 + B$ lower than ρ_0 is needed to match $f_{\bar{\theta}}(\cdot)$ with $f_0(\cdot) + [\tilde{f}(\cdot) - \sigma_*^2]$. We thank a referee for providing this alternative argument.

14. The data in levels enable the use of y_0 in $\mathbb{E}[\Delta x_t | \Delta y_{1:T}, y_0]$, which dominates $\Delta x_t^* = \mathbb{E}[\Delta x_t | \Delta y_{1:T}]$ unless $\rho_0 = 0$.

15. We begin the sample in 1952Q1 to coincide with the Treasury-Fed accord. As is well-known, this accord established in its modern terms the separation between monetary and fiscal policies, inaugurating a period of more stable behavior of economic aggregates in comparison to the immediate aftermath of World War II. In turn, we end our sample at 2019Q4 to avoid the use of the yet provisional (and highly variable) data from 2020. Thus, all the data in our sample has been subject to at least one annual revision by the BEA.

16. When we restrict our sample to the one used by [Aruoba et al. \(2016\)](#), we obtain estimates of $\{\mu_0, \rho_0, \sigma_0, \sigma_1, \sigma_2\}$ comparable to theirs. For the subsample 1960Q1–2011Q4 that they use, the variance of the signal is slightly lower, and so is the R^2 measure of common trend observability.

17. In fact, the smoothed series obtained by assuming that GDE and GDI are not cointegrated lies within the credible sets obtained under the assumption that the model that imposes cointegration is correct. Therefore, the difference between the two smoothed series, which is largest in the period preceding the fall of Lehman Brothers but does not show any business cycle variation, is not significant.

18. In this respect, we follow [Müller and Watson \(2008\)](#) but similar patterns arise when we use 5- or 10-year intervals instead.

ACKNOWLEDGMENTS

We are grateful to Dante Amengual, Borağan Aruoba, Peter Boswijk, Frank Diebold, Ulrich Müller, Tommaso Proietti, Richard Smith, and Mark Watson for useful discussions, as well as audiences at Princeton, ESCoE (London 2018), ESEM (Köln 2018), Computational and Financial Econometrics (Pisa 2018) and SAEe (Madrid 2018). Yoosoon Chang and two anonymous referees have also provided very useful feedback. Of course, the usual caveat applies. The first author acknowledges financial support from the Santander Research chair at CEMFI, the second one is grateful to MIUR through the PRIN project “High-dimensional time series for structural macroeconomic analysis in times of pandemic”, while the third one received funding from the Spanish Ministry of Science and Innovation through grant PID2021-128963NB-I00. The views expressed in this chapter are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

REFERENCES

- Almuzara, M., Amengual, D., Fiorentini, G., & Sentana, E. (2022). *GDP solera: The ideal vintage mix* [CEMFI Working Paper 2204].
- Almuzara, M., Amengual, D., & Sentana, E. (2019). Normality tests for latent variables. *Quantitative Economics*, 10, 981–1017.
- Aruoba, S. B. (2008). Data revisions are not well behaved. *Journal of Money, Credit and Banking*, 40, 319–340.
- Aruoba, S. B., Diebold, F. X., Nalewaik, J., Schorfheide, F., & Song, D. (2016). Improving GDP measurement: A measurement-error perspective. *Journal of Econometrics*, 191, 384–397.
- Blanchard, O. J. (1997). The medium run. *Brookings Papers on Economic Activity*, 28, 89–158.

- Cavanagh, C. L. (1985). *Roots local to unity* [manuscript, Harvard University].
- Chan, N. H., & Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, 15, 1050–1063.
- Chang, Y., Miller, J. I., & Park, J. Y. (2009). Extracting a common trend: Theory with some applications. *Journal of Econometrics*, 150, 231–247.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed.). Oxford University Press.
- Fiorentini, G., Galesi, A., & Sentana, E. (2018). A spectral EM algorithm for dynamic factor models. *Journal of Econometrics*, 205, 249–279.
- Fiorentini, G., & Sentana, E. (2019). Dynamic specification tests for dynamic factor models. *Journal of Applied Econometrics*, 34, 325–346.
- Gordon, R. J. (2016). Perspectives on the rise and fall of American growth. *American Economic Review: Papers and Proceedings*, 106, 72–76.
- Greenaway-McGrevey, R. (2011). *Is GDP or GDI a better measure of output? A statistical approach* [Working Paper 2011-08]. Bureau of Economic Analysis.
- Grimm, B. T. (2007). *The statistical discrepancy* [Working Paper 2007-01]. Bureau of Economic Analysis.
- Hansen, A. (1939). Economic progress and declining population growth. *American Economic Review*, 29, 1–15.
- Jacobs, J. P. A. M., & van Norden, S. (2011). Modeling data revisions: Measurement error and dynamics of true values. *Journal of Econometrics*, 161, 101–109.
- Jacobs, J. P. A. M., S. Sarferaz, J. Sturm, and S. van Norden (2020). Can GDP measurement be further improved? Data revision and reconciliation. *Journal of Business and Economic Statistics*, 40, 423–431.
- Jørgensen, B., & Labouriau, R. (2012). *Exponential families and theoretical inference* (second ed.), Springer, Rio de Janeiro.
- Kaldor, N. (1957). A model of economic growth. *Economic Journal*, 67, 591–624.
- Karabarbounis, L., & Neiman, B. (2014). The global decline of the labor share. *Quarterly Journal of Economics*, 129, 61–103.
- Landefeld, J. S., Seskin, E. P., & Fraumeni, B. M. (2008). Taking the pulse of the economy: Measuring GDP. *Journal of Economic Perspectives*, 22, 193–216.
- Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44, 226–233.
- Müller, U. K., & Watson, M. W. (2008). Testing models of low-frequency variability. *Econometrica*, 76, 979–1016.
- Nalewaik, J. (2010). The income- and expenditure-side measures of output growth. *Brookings Papers on Economic Activity*, 1, 71–106.
- Nalewaik, J. (2011). The income- and expenditure-side measures of output growth: An update through 2011Q2. *Brookings Papers on Economic Activity*, 2, 385–402.
- Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74, 535–547.
- Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *Journal of Econometrics*, 49, 305–341.
- Smith, R. J., Weale, M. R., & Satchell, S. E. (1998). Measurement error with accounting constraints: Point and interval estimation for latent data with an application to U.K. gross domestic product. *Review of Economic Studies*, 65, 109–134.
- Stone, R., Champernowne, D. G., & Meade, J. E. (1942). The precision of national income estimates. *Review of Economic Studies*, 9, 111–125.
- Summers, R. H. (2015). Demand side secular stagnation. *American Economic Review: Papers and Proceedings*, 105, 60–65.
- Watson, M. W., & Engle, R. F. (1983). Alternative algorithms for estimation of dynamic MIMIC models, factor, and time varying coefficient regression models. *Journal of Econometrics*, 23, 385–400.
- Weale, M. (1992). Estimation of data measured with error and subject to linear restrictions. *Journal of Applied Econometrics*, 7, 167–174.

APPENDIX: PROOF OF THEOREM 1 AND DERIVATION OF EQUATIONS (2) AND (3)

Although for the sake of brevity, we do not discuss frequency-domain ML estimation (see, e.g., [Fiorentini et al., 2018](#)) or Bayesian estimation (e.g., [Durbin & Koopman, 2012](#)), before presenting the proof, it is useful to describe the way estimates are produced. We can obtain numerically equivalent Gaussian MLEs of θ , $\hat{\theta}$, by means of two algorithms. The first one exploits the Kalman filter to recursively compute the one-period ahead conditional means and variances of observables appearing in the log-likelihood function. The second is the EM algorithm, which, for some initial $\hat{\theta}^{(0)}$, updates parameter estimates by iterating over

$$\begin{aligned}\hat{\theta}^{(s)} &= \operatorname{argmax}_{\theta \in \Theta} \mathbb{E}_{\hat{\theta}^{(s-1)}} \left[\sum_{t=1}^T \ln p_\vartheta(\Delta x_t | \Delta x_{1:(t-1)}) + \sum_{t=1}^T \sum_{i=1}^N \ln p_{\psi_i}(\Delta v_{it} | y_{1:T}) \right], \\ \ln p_\vartheta(\Delta x_t | \Delta x_{1:(t-1)}) &= -\frac{1}{2} \left[\ln(2\pi(1-\rho^2)\sigma^2) + \frac{(\Delta x_t - (1-\rho)\mu - \rho\Delta x_{t-1})^2}{(1-\rho^2)\sigma^2} \right], \\ \ln p_{\psi_i}(\Delta v_{it}) &= -\frac{1}{2} \left[\ln(2\pi\psi_i^2) + \frac{\Delta v_{it}^2}{\psi_i^2} \right].\end{aligned}$$

The EM algorithm alternates between smoothing the so-called complete-data likelihood using the current value $\hat{\theta}^{(s-1)}$ for expectation calculations (the E-step), and maximizing the resulting smoothed function to yield a new value $\hat{\theta}^{(s)}$ (M-step). See [Dempster et al. \(1977\)](#), [Ruud \(1991\)](#), and [Watson and Engle \(1983\)](#). If the algorithm converges, we have $\hat{\theta} = \lim_{s \rightarrow \infty} \hat{\theta}^{(s)}$. This EM algorithm is particularly relevant because our proof relies heavily on a generalization of [Louis \(1982\)](#) score formula, which we call EM principle, formalized in [Almuzara et al. \(2019, th. 1\)](#). Consider the functions

$$\begin{aligned}g_\mu(\theta) &= \sqrt{T} \left(\frac{\mathbb{E}_T[\Delta x_t - \rho\Delta x_{t-1}] - \mu}{1-\rho} - \mu \right), \\ g_\rho(\theta) &= \sqrt{T} \mathbb{E}_T[(\Delta x_{t-1} - \mu)(\Delta x_t - \mu) - \rho(\Delta x_{t-1} - \mu)^2], \\ g_\sigma(\theta) &= \sqrt{T} \left(\frac{\mathbb{E}_T[((\Delta x_t - \mu) - \rho(\Delta x_{t-1} - \mu))^2]}{1-\rho^2} - \sigma^2 \right), \\ g_{\psi_i}(\theta) &= \sqrt{T} (\mathbb{E}_T[\Delta v_{it}^2] - \psi_i^2), i = 1, \dots, N.\end{aligned}\tag{A.1}$$

These are proportional to the scaled average scores of the complete-data log-likelihood for the misspecified model. Maximum likelihood estimates $\hat{\theta}$ are characterized by the first-order necessary conditions

$$\mathbb{E}_{\hat{\theta}}[g_{\mu}(\hat{\theta}) | \Delta y_{1:T}] = \mathbb{E}_{\hat{\theta}}[g_{\rho}(\hat{\theta}) | \Delta y_{1:T}] = \mathbb{E}_{\hat{\theta}}[g_{\sigma}(\hat{\theta}) | \Delta y_{1:T}] = \mathbb{E}_{\hat{\theta}}[g_{\psi_i}(\hat{\theta}) | \Delta y_{1:T}] = 0.$$

Define the auxiliary functions

$$\begin{aligned}\tilde{g}_{\mu}(\theta) &= \sqrt{T}(\mathbb{E}_T[\Delta x_t] - \mu), \\ \tilde{g}_{\sigma}(\theta) &= \sqrt{T}(\mathbb{E}_T[(\Delta x_t - \mu)^2] - \sigma^2),\end{aligned}$$

and note that the maximum likelihood estimates for the static model $\tilde{\theta}$ (i.e., the restricted maximum likelihood estimates subject to $\rho = 0$) satisfy

$$\mathbb{E}_{\tilde{\theta}}[\tilde{g}_{\mu}(\tilde{\theta}) | \Delta y_{1:T}] = \mathbb{E}_{\tilde{\theta}}[\tilde{g}_{\sigma}(\tilde{\theta}) | \Delta y_{1:T}] = \mathbb{E}_{\tilde{\theta}}[g_{\psi_i}(\tilde{\theta}) | \Delta y_{1:T}] = 0. \quad (\text{A.2})$$

The first lemma will allow us to replace g_{μ} and g_{σ} by the much simpler \tilde{g}_{μ} and \tilde{g}_{σ} .

Lemma 1. *Let $\hat{\theta}$ be the maximum likelihood estimator for the misspecified model. Under Assumptions 1 and 2,*

$$\begin{aligned}\mathbb{E}_{\hat{\theta}}[g_{\mu}(\hat{\theta}) | \Delta y_{1:T}] &= \mathbb{E}_{\hat{\theta}}[\tilde{g}_{\mu}(\hat{\theta}) | \Delta y_{1:T}] + o_p(1) \\ \text{and } \mathbb{E}_{\hat{\theta}}[g_{\sigma}(\hat{\theta}) | \Delta y_{1:T}] &= \mathbb{E}_{\hat{\theta}}[\tilde{g}_{\sigma}(\hat{\theta}) | \Delta y_{1:T}] + o_p(1).\end{aligned}$$

Proof.

For any $\theta \in \Theta$,

$$g_{\mu}(\theta) - \tilde{g}_{\mu}(\theta) = \frac{\rho(\Delta x_T - \Delta x_0)}{(1-\rho)\sqrt{T}}.$$

One can then show that $\mathbb{E}_{\theta}[\Delta x_0 | \Delta y_{1:T}] = O_p(1)$ and $\mathbb{E}_{\theta}[\Delta x_T | \Delta y_{1:T}] = O_p(1)$, which leads to

$$\mathbb{E}_{\theta}[g_{\mu}(\theta) - \tilde{g}_{\mu}(\theta) | \Delta y_{1:T}] = O_p(1/\sqrt{T}).$$

In particular, this implies that $\mathbb{E}_{\hat{\theta}}[g_{\mu}(\hat{\theta}) - \tilde{g}_{\mu}(\hat{\theta}) | \Delta y_{1:T}] = o_p(1)$. Turning to the score function with respect to σ , for any $\theta \in \Theta$,

$$g_{\sigma}(\theta) + \frac{2\rho}{1-\rho^2} g_{\rho}(\theta) - \tilde{g}_{\sigma}(\theta) = \frac{\rho^2((\Delta x_T - \mu)^2 - (\Delta x_0 - \mu)^2)}{(1-\rho^2)\sqrt{T}}.$$

Since $\mathbb{E}_{\theta}[(\Delta x_0)^2 | \Delta y_{1:T}] = O_p(1)$ and $\mathbb{E}_{\theta}[(\Delta x_T)^2 | \Delta y_{1:T}] = O_p(1)$,

$$\mathbb{E}_\theta[g_\sigma(\theta) + \frac{2\rho}{1-\rho^2} g_\rho(\theta) - \tilde{g}_\sigma(\theta) | \Delta y_{1:T}] = O_p(1/\sqrt{T}).$$

And since $\mathbb{E}_{\hat{\theta}}[g_\rho(\hat{\theta}) | \Delta y_{1:T}] = 0$, we finally get $\mathbb{E}_{\hat{\theta}}[g_\sigma(\hat{\theta}) - \tilde{g}_\sigma(\hat{\theta}) | \Delta y_{1:T}] = o_p(1)$. \blacksquare

Remark. A subtlety in the previous proof is that order-in-probability statements refer to \mathbb{P} while expectations refer to \mathbb{P}_θ . For the sake of brevity, we omit the proof that this difference is inconsequential.

Let $\theta_{\rho=0}$ be the parameter vector θ in which we set $\rho = 0$. In an abuse of notation, we will also occasionally identify $\theta_{\rho=0}$ with the subvector that excludes the $\rho = 0$ component – note that, trivially, $\hat{\theta}$ and $\hat{\theta}_{\rho=0}$ represent the same parameter value. For the Proof of Lemma 2, the following remark will be useful:

Remark. Sample spaces of $\Delta y_{1:T}$, $\Delta x_{0:T}$, and $\Delta v_{1:T}$ are $\mathcal{Y} = \mathbb{R}^{NT}$, $\mathcal{X} = \mathbb{R}^{T+1}$, and $\mathcal{V} = \mathbb{R}^{NT}$. Probability distributions \mathbb{P} and \mathbb{P}_θ , $\theta \in \Theta$, may be taken to be measures on the Borel sets of $\mathcal{X} \times \mathcal{V}$ with probability statements about $\Delta y_{1:T}$ interpreted by means of the inverse image of the mapping in (1). The measure \mathbb{P} and the model \mathcal{P} are then dominated by Lebesgue measure λ on the Borel sets of $\mathcal{X} \times \mathcal{V}$. Consequently, densities exist by the Radon-Nikodym theorem. In contrast, conditional distributions of $\Delta x_{0:T}$ and $\Delta v_{1:T}$ given $\Delta y_{1:T}$ implied by \mathbb{P} and \mathcal{P} are dominated by the σ -finite measure $\lambda_{\Delta y_{1:T}}$ on the Borel sets of the hyperplane defined by (1) for fixed $\Delta y_{1:T}$, rather than by \mathcal{P} . Therefore, conditional densities exist with respect to $\lambda_{\Delta y_{1:T}}$ in that case too.

Lemma 2. Under Assumptions 1 and 2,

$$\sqrt{T} \left(\mathbb{E}_{\hat{\theta}}[\mathbb{E}_T[\Delta x_i] | \Delta y_{1:T}] - \mathbb{E}_{\hat{\theta}_{\rho=0}}[\mathbb{E}_T[\Delta x_i] | \Delta y_{1:T}] \right) = o_p(1),$$

$$\sqrt{T} \left(\mathbb{E}_{\hat{\theta}}[\mathbb{E}_T[\Delta x_i^2] | \Delta y_{1:T}] - \mathbb{E}_{\hat{\theta}_{\rho=0}}[\mathbb{E}_T[\Delta x_i^2] | \Delta y_{1:T}] \right) = o_p(1),$$

$$\sqrt{T} \left(\mathbb{E}_{\hat{\theta}}[\mathbb{E}_T[\Delta v_{it}^2] | \Delta y_{1:T}] - \mathbb{E}_{\hat{\theta}_{\rho=0}}[\mathbb{E}_T[\Delta v_{it}^2] | \Delta y_{1:T}] \right) = o_p(1), i = 1, \dots, N.$$

Proof. Let x denote the latent variables $\{\Delta x_{0:T}, v_{1:N,T}\}$ and y the observables $\Delta y_{1:T}$. Under the restriction $\rho = 0$, the model for x has density

$$p_\eta(x) = b \exp[T \cdot (\eta' S(x) - a(\eta))]$$

with respect to measure λ . Similarly, the density of x given y is an exponential family with density

$$p_\eta(x | y) = b \exp[T \cdot (\eta' S(x) - a(\eta | y))]$$

with respect to measure λ_y . Measures λ and λ_y are defined in the remark above, b is a constant, $\eta = \eta(\mu, \sigma, \psi_{1:N})$ is a function of the original

parameters, $a(\cdot)$ and $a(\cdot | y)$ are functions of η , and the sufficient statistics are $S(x) = \mathbb{E}_T[(\Delta x_t, \Delta x_t^2, \Delta v_{1t}^2, \dots, \Delta v_{Nt}^2)']$.

Define $\hat{S} = \mathbb{E}_{\hat{\theta}}[S(x)]$ and note that if x is such that $S(x) = \hat{S}$, then the densities $p_\eta(x)$ and $p_\eta(x | y)$ are maximized at $\hat{\eta} = \eta(\hat{\mu}, \hat{\sigma}, \hat{\psi}_{1:N})$.

In addition,

$$\hat{S} = \frac{\partial a(\hat{\eta})}{\partial \eta} = \frac{\partial a(\hat{\eta} | y)}{\partial \eta} = \mathbb{E}_{\hat{\theta}_{\rho=0}}[S(x) | y]$$

where this follows from well-known properties of exponential families (Jørgensen & Labouriau, 2012, Th. 1.17 and 1.18). Since $\mathbb{E}_{\hat{\theta}}[S(x) | y] = \hat{S} + o_p(1/\sqrt{T})$ by virtue of Lemma 1, the current lemma immediately follows. ■

The rest of the argument for the asymptotic equivalence between $\hat{\theta}_{\rho=0}$ and $\tilde{\theta}_{\rho=0}$ is standard. Specifically, if G collects the static model estimating equations (5) (suitably scaled by T), and $\bar{\theta}_{\rho=0}$ denotes the (common) probability limit of the two estimators, a Taylor expansion gives

$$o_p(1) = G(\hat{\theta}_{\rho=0}) - G(\tilde{\theta}_{\rho=0}) = [H(\bar{\theta}_{\rho=0}) + o_p(1)] \times \sqrt{T}(\hat{\theta}_{\rho=0} - \tilde{\theta}_{\rho=0}),$$

where $H(\bar{\theta}_{\rho=0})$ is a fixed nonsingular matrix, which in turn implies $\sqrt{T}(\hat{\theta}_{\rho=0} - \tilde{\theta}_{\rho=0}) = o_p(1)$.

We now turn to characterizing the pseudo-true value for the estimator $\hat{\rho}$. From (4) and (a small variation of) Lemma 2, we obtain

$$\mathbb{E}_T[\mathbb{E}_{\hat{\theta}}[(\Delta x_{t-1} - \hat{\mu})(\Delta x_t - \hat{\mu}) | \Delta y_{1:T}]] - \hat{\rho}\hat{\sigma}^2 = o_p(1).$$

Let $\bar{\rho} = \rho_0 + B$ be the probability limit of $\hat{\rho}$. Our discussion of equivalence between dynamic-model and static-model MLE has already established that $\hat{\mu} \xrightarrow{p} \mu_0$, $\hat{\sigma} \xrightarrow{p} \sigma_0$ and $\hat{\psi}_i \xrightarrow{p} \sigma_i$. Thus, $\hat{\rho}\hat{\sigma}^2 \xrightarrow{p} \bar{\rho}\sigma_0^2$. Also let $\bar{\theta}$ be the probability limit of $\hat{\theta}$.

For any $\theta \in \Theta$, we have

$$\begin{aligned} \mathbb{E}_\theta[(\Delta x_{t-1} - \mu)(\Delta x_t - \mu) | \Delta y_{1:T}] &= \text{Cov}_\theta(\Delta x_{t-1}, \Delta x_t | \Delta y_{1:T}) \\ &+ (\mathbb{E}_\theta[\Delta x_{t-1} | \Delta y_{1:T}] - \mu)(\mathbb{E}_\theta[\Delta x_t | \Delta y_{1:T}] - \mu) \\ &= \mathbb{E}_\theta[\text{Cov}_\theta(\Delta x_{t-1}, \Delta x_t | \Delta y_{1:T})] \\ &+ (\mathbb{E}_\theta[\Delta x_{t-1} | \Delta y_{1:T}] - \mu)(\mathbb{E}_\theta[\Delta x_t | \Delta y_{1:T}] - \mu) \\ &= \text{Cov}_\theta(\Delta x_{t-1}, \Delta x_t) - \text{Cov}_\theta(\mathbb{E}_\theta[\Delta x_{t-1} | \Delta y_{1:T}], \mathbb{E}_\theta[\Delta x_t | \Delta y_{1:T}]) \\ &+ (\mathbb{E}_\theta[\Delta x_{t-1} | \Delta y_{1:T}] - \mu)(\mathbb{E}_\theta[\Delta x_t | \Delta y_{1:T}] - \mu) \end{aligned}$$

The second line follows from properties of the normal distribution, while the third follows from a well-known identity for covariances. Now, $\text{Cov}_\theta(\Delta x_{t-1}, \Delta x_t) = \rho\sigma^2$ which leads to

$$\begin{aligned}
o_p(1) &= \mathbb{E}_T[\mathbb{E}_{\hat{\theta}}[(\Delta x_{t-1} - \hat{\mu})(\Delta x_t - \hat{\mu}) | \Delta y_{1:T}] - \hat{\rho}\hat{\sigma}^2] \\
&= \frac{1}{T} \sum_{t=1}^T (\mathbb{E}_{\hat{\theta}}[\Delta x_{t-1} | \Delta y_{1:T}] - \hat{\mu})(\mathbb{E}_{\hat{\theta}}[\Delta x_t | \Delta y_{1:T}] - \hat{\mu}) \\
&\quad - \frac{1}{T} \sum_{t=1}^T \text{Cov}_{\hat{\theta}}(\mathbb{E}_{\hat{\theta}}[\Delta x_{t-1} | \Delta y_{1:T}], \mathbb{E}_{\hat{\theta}}[\Delta x_t | \Delta y_{1:T}])
\end{aligned}$$

Let $T \rightarrow \infty$,

$$\begin{aligned}
&\frac{1}{T} \sum_{t=1}^T (\mathbb{E}_{\hat{\theta}}[\Delta x_{t-1} | \Delta y_{1:T}] - \hat{\mu})(\mathbb{E}_{\hat{\theta}}[\Delta x_t | \Delta y_{1:T}] - \hat{\mu}) \xrightarrow{p} \\
&\quad \text{Cov}(\mathbb{E}_{\bar{\theta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]), \\
&\frac{1}{T} \sum_{t=1}^T \text{Cov}_{\hat{\theta}}(\mathbb{E}_{\hat{\theta}}[\Delta x_{t-1} | \Delta y_{1:T}], \mathbb{E}_{\hat{\theta}}[\Delta x_t | \Delta y_{1:T}]) \xrightarrow{p} \\
&\quad \text{Cov}_{\bar{\theta}}(\mathbb{E}_{\bar{\theta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]).
\end{aligned}$$

We take limits by replacing sample averages by expectations, $\hat{\theta}$ by $\bar{\theta}$, and smoothing with respect to $\Delta y_{1:T}$ by smoothing with respect to $\Delta y_{-\infty:\infty}$. Smoothing with respect to $\Delta y_{1:\infty}$ and $\Delta y_{-\infty:\infty}$ give the same result (the proof of which we omit), but the second turns out to be more convenient.

Both limits are covariances between the smoothed values of Δx_{t-1} and Δx_t obtained using the misspecified model, but in the first case the covariance is taken under the data generating process \mathbb{P} while in the second the covariance is computed using the misspecified model $\mathbb{P}_{\bar{\theta}}$ evaluated at the pseudo-true value $\bar{\theta}$. In summary, $\bar{\rho}$ is characterized by the equation

$$\begin{aligned}
&\text{Cov}(\mathbb{E}_{\bar{\theta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]) \\
&= \text{Cov}_{\bar{\theta}}(\mathbb{E}_{\bar{\theta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]).
\end{aligned}$$

In order to establish the spectral condition (2) we note that the Fourier transform of $\mathbb{E}_{\bar{\theta}}[\Delta x_t | \Delta y_{-\infty:\infty}]$ is the Wiener-Kolmogorov filter,

$$\Delta x_{\infty}(\lambda) = \frac{f_{\bar{\theta}}(\lambda)}{f_{\bar{\theta}}(\lambda) + \sigma_*^2} \frac{\sum_{i=1}^N \sigma_i^{-2} \Delta y_i(\lambda)}{\sum_{i=1}^N \sigma_i^{-2}}$$

where $\Delta y_i(\lambda)$ is the Fourier transform of the time series $\Delta y_{i,-\infty:\infty}$. The filter Δx_{∞} has spectrum

$$\left(\frac{f_{\bar{\theta}}(\lambda)}{f_{\bar{\theta}}(\lambda) + \sigma_*^2} \right)^2 (f_0(\lambda) + \tilde{f}(\lambda))$$

under the data generating process \mathbb{P} and spectrum

$$\left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2} \right)^2 (f_{\bar{\vartheta}}(\lambda) + \sigma_*^2)$$

under the misspecified model $\mathbb{E}_{\bar{\vartheta}}$. Hence, by Fourier inversion,

$$\text{Cov}(\mathbb{E}_{\bar{\vartheta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\vartheta}}[\Delta x_t | \Delta y_{-\infty:\infty}]) = \int_0^{2\pi} \cos(\lambda) \left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2} \right)^2 (f_0(\lambda) + \tilde{f}(\lambda)) d\lambda,$$

$$\text{Cov}_{\bar{\vartheta}}(\mathbb{E}_{\bar{\vartheta}}[\Delta x_{t-1} | \Delta y_{-\infty:\infty}], \mathbb{E}_{\bar{\vartheta}}[\Delta x_t | \Delta y_{-\infty:\infty}]) = \int_0^{2\pi} \cos(\lambda) \left(\frac{f_{\bar{\vartheta}}(\lambda)}{f_{\bar{\vartheta}}(\lambda) + \sigma_*^2} \right)^2 (f_{\bar{\vartheta}}(\lambda) + \sigma_*^2) d\lambda,$$

whence equation (2) follows.

This page intentionally left blank

CHAPTER 2

MARKOV SWITCHING RATIONALITY

Florens Odendahl^a, Barbara Rossi^b and
Tatevik Sekhposyan^c

^aBanco de España, Madrid, Spain

^bICREA-University Pompeu Fabra, Barcelona School of Economics and CREI, Barcelona, Spain

^cTexas A&M University, College Station, Texas, USA

ABSTRACT

The authors propose novel tests for the detection of Markov switching deviations from forecast rationality. Existing forecast rationality tests either focus on constant deviations from forecast rationality over the full sample or are constructed to detect smooth deviations based on non-parametric techniques. In contrast, the proposed tests are parametric and have an advantage in detecting abrupt departures from unbiasedness and efficiency, which the authors demonstrate with Monte Carlo simulations. Using the proposed tests, the authors investigate whether Blue Chip Financial Forecasts (BCFF) for the Federal Funds Rate (FFR) are unbiased. The tests find evidence of a state-dependent bias: forecasters tend to systematically overpredict interest rates during periods of monetary easing, while the forecasts are unbiased otherwise. The authors show that a similar state-dependent bias is also present in market-based forecasts of interest rates, but not in the forecasts of real GDP growth and GDP deflator-based inflation. The results emphasize the special role played by monetary policy in shaping interest rate expectations above and beyond macroeconomic fundamentals.

Keywords: Forecasting; forecast rationality; regression-based tests of forecasting ability; Markov-switching models; survey forecasts; interest rate forecasts

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 35–64

Copyright © 2023 by Florens Odendahl, Barbara Rossi and Tatevik Sekhposyan

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045B002

1. INTRODUCTION

Producing accurate forecasts for economic variables is an important task for both researchers and policymakers alike. A desirable property that forecasts should have is optimality. When evaluating forecasts using a quadratic loss, forecast optimality implies that the forecast error should not be predictable by a constant (unbiasedness), the forecast itself (efficiency), or any information available at the time the forecast is made; otherwise, it is reasonable to conclude that the forecast is suboptimal and can be improved.

However, it is well known that forecasting performance can be unstable over time, and changes in the forecast's quality may be associated with recurring states of the economy. For instance, [Joutz and Stekler \(2000\)](#) find that the Federal Reserve Board's (Fed) forecasts overestimated output growth in slowdowns and recessions and underestimated it in recoveries. Similarly, inflation is typically underpredicted when it is rising and overpredicted when it is declining. [Granziera et al. \(2021\)](#) reach similar conclusions for the European Central Bank's (ECB) inflation forecasts: the ECB tends to overpredict (underpredict) inflation when inflation is below (above) target. [Sinclair et al. \(2010\)](#) show that information on real and inflationary cycles, though incorporated in the Fed's nowcasts, are not incorporated into one-quarter-ahead forecasts.

Evaluating forecasts that may have state-dependent forecast rationality (unbiasedness and efficiency) properties requires particular care since standard tests of absolute and relative forecast evaluations are misleading in the presence of instabilities ([Rossi, 2013](#)). As a consequence, the literature on the evaluation of the absolute (and relative) performance of forecasting models has developed techniques to robustify inference, where instabilities are accounted for non-parametrically; see for instance [Rossi and Sekhposyan \(2016\)](#).

We contribute to this literature by proposing two novel tests that are robust to the presence of time variation. The novelty is that we assume a parametric form of time variation driven by unobserved states; namely, we assume that the forecast errors follow a Markov switching process. Consequently, our tests are more powerful than non-parametric tests of forecast rationality when the forecasting performance varies over time in a regime-switching manner. Our approach is relevant in situations where the rationality of the forecasts depends on exogenous, unobserved (or *a priori* unknown but observable) cycles. An interesting extension would be to adapt the tests to models of endogenous regime-switching, as in [Chang et al. \(2017\)](#).

To demonstrate the empirical relevance of our approach consider [Fig. 1](#), which displays the forecast errors of three-month-ahead (effective) FFR forecasts from the BCFF survey; the forecast error is measured by the difference between the realization and the forecast. The figure also depicts our estimated Markov switching forecast error mean (dash-dotted line).¹ The estimation results imply deviations from unbiasedness during periods of monetary easing (when interest rates, depicted by the dashed black line, are decreasing); this state dependence was identified by the Markov switching model and did not require *ex-ante* specification of the state variable.

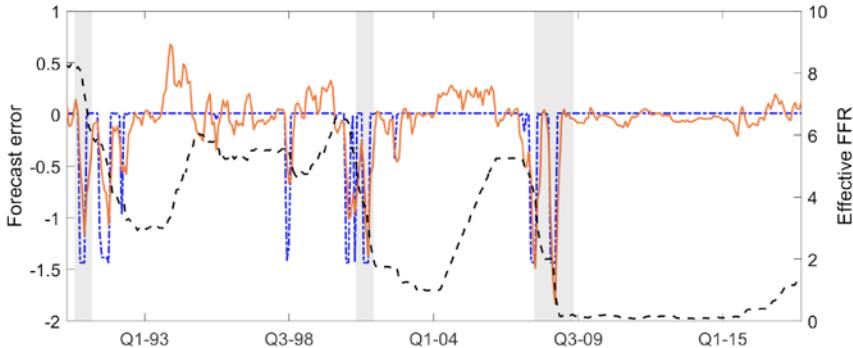


Fig. 1. Forecast Errors with Markov Switching Bias. *Notes:* The solid line (left-hand side y -axis) shows the forecast errors: the difference between the FFR realizations and BCFF's respective three-month-ahead forecasts. The dash-dotted line (left-hand side y -axis) shows a regime-switching unconditional mean estimated using the smoothed state probabilities implied by a two-state Markov switching model; see Section 5 for the description of the model. The dashed line (right-hand side y -axis) shows the FFR level. Grey shaded areas display the National Bureau of Economic Research (NBER) recession periods.

Importantly, our proposed tests can uncover recurring periods of deviations from forecast rationality, even when traditional tests do not reject, on average, over the full sample. Deviations from forecast rationality could be linked to important economic events, such as recessions, financial distress, or other economic circumstances. For instance, Bullard (2016) announced that the St. Louis Fed has abandoned the view of the economy having a single steady-state in favor of a regime-switching world with several steady states. Alternatively, occasional deviations from forecast rationality could be potentially related to information frictions or the way agents learn at different times. Our procedure allows researchers to identify the periods when the forecast was biased as well as quantify the bias, while treating the periods when the absolute performance varies as an unknown state variable.

In this chapter, we propose forecast rationality tests that build on Hansen (1992)'s tests for detecting Markov switching and on the bootstrap procedure proposed in Qu and Zhuo (2021) (we also consider extensions of Garcia, 1998 in the Online Appendix). Testing for Markov switching requires non-standard inference because of several problems. First, the hyper-parameters of the switching process (for instance, the state-to-state transition probabilities) are not identified under the null hypothesis of parameter stability (Davies, 1977, 1987). Second, under the null of parameter stability, the score with respect to the restricted parameters is identically zero, which violates standard regularity conditions imposed to derive an asymptotic chi-squared distribution via a usual second-order Taylor expansion. Consequently, standard Likelihood ratio (LR), Wald, and Lagrange multiplier tests do not have a chi-squared distribution, even asymptotically. Third, the conditional regime probabilities follow a stochastic process that can only be

represented recursively, thus making higher-order Taylor approximations infeasible. Furthermore, there are multiple ways to impose the null of a single state, which further complicates inference, in addition to creating a boundary parameter problem for the null parameter space. Hansen (1992), Garcia (1998), Cho and White (2007), Carter and Steigerwald (2012), and Qu and Zhuo (2021) discuss these issues in detail and make significant contributions, thus shaping our knowledge on how to test for the number of regimes in Markov-switching models.

Our chapter builds on the existing literature and proposes a Markov switching test in a forecast rationality framework, where we impose a joint null hypothesis that there is a single regime and that the relevant parameters are restricted to zero under the null hypothesis. More specifically, we rely on Hansen (1992) who treated the likelihood function as a stochastic process and obtained a lower bound for the LR test. In addition, we use the bootstrap procedure of Qu and Zhuo (2021) to test our null. The bootstrap addresses several of the difficulties associated with testing for Markov switching outlined above and performs well in finite samples. Qu and Zhuo's (2021) bootstrap, though building on Cho and White (2007), does not explicitly address the boundary parameter issue of the composite null that the latter focused on, even though they analyze situations when the transition probabilities are close to the boundary.

In particular, we adapt these tests to our absolute forecast evaluation context and refer to them as the “absolute forecast evaluation – Hansen” (AFE-H) and the “absolute forecast evaluation – bootstrap” (AFE-BS). In addition, the Online Appendix also investigates the “absolute forecast evaluation – Garcia” (AFE-G) test inspired by Garcia (1998). The main difference between our proposed tests relative to the ones in the literature is that Hansen (1992) and Qu and Zhuo (2021) test for Markov switching in the parameters of a model, leaving the values of the parameters unspecified under the null. In contrast, we test for Markov switching directly in the forecast errors by specifying parameter values such that forecast rationality is satisfied under the null. That is, we test for forecast rationality in the full out-of-sample portion of the data against local, regime-switching deviations. More specifically, consider a standard forecast unbiasedness test (Mincer & Zarnowitz, 1969), which evaluates whether the forecast error has a zero mean against the alternative that the mean differs from zero. Instead, under the alternative of our approach, we let the forecast error evolve according to a Markov switching process, and jointly test that the mean of the forecast error is time-invariant and equal to zero.

Aside from the literature on testing for Markov switching, we also relate to a large literature on the evaluation of the absolute predictive performance (Mincer & Zarnowitz, 1969; Rossi & Sekhposyan, 2016; West & McCracken, 1998). Mincer and Zarnowitz (1969) and West and McCracken (1998) assume a constant mean of the forecast error and a constant efficiency parameter, an assumption that is violated in the presence of time variation. Rossi and Sekhposyan (2016) propose a non-parametric test for forecast rationality that is robust to instabilities, which is useful in the situations where rationality is time-varying yet, on average, holds in the full sample. Given the non-parametric nature, their test performs well when time-variation is smooth and persistent, while our proposed tests have stronger

power in the presence of abrupt, short-lived and recurring deviations from rationality. In fact, our tests can detect deviations from rationality that occur for short periods of time, as long as the deviations occur repeatedly.² Note that in our work we focus on model-free or survey-based forecasts, as well as forecasts obtained either with a rolling window with finite size or a recursive window where the contribution of parameter estimation error can be reasonably ignored (for instance, when the estimation sample size is relatively large compared to the evaluation sample size).

We investigate the finite sample properties of our proposed tests with Monte Carlo simulations. The simulations show empirical rejection frequencies close to the nominal size when testing the null hypothesis of unbiasedness and efficiency using the AFE-BS test. For unbiasedness, the AFE-H is well sized in medium to large samples and somewhat undersized when testing for efficiency. In terms of power, the rejection frequencies are similar to the test of [West and McCracken \(1998\)](#) and the Fluctuation rationality test of [Rossi and Sekhposyan \(2016\)](#) for the alternative of a constant deviation from rationality, and clearly outperform both under a Markov switching alternative.

Turning to the empirical analysis, we investigate potential biases in the BCFF survey predictions for the FFR. When we consider the three-month- and six-month-ahead forecast errors, our test rejects unbiasedness in favor of a two-regime model. The estimated regimes indicate that the forecasts are unbiased in the first regime, the one that is prevalent most of the time. However, there is evidence of a second regime in which the forecasters overestimate the FFR. The occurrence of the second regime is associated with monetary policy easing and is not limited to recessionary periods. The biases are present not only in survey forecasts but also in market-based forecasts, suggesting that the lack of forecast rationality is not specific to the survey but inherent to difficulty in forecasting. We investigate the role of disagreement among panelists as well as the role of monetary policy uncertainty, measured via newspaper articles ([Baker et al., 2016](#)), in explaining this regime-dependent behavior. We find no clear association with disagreement, while the regimes appear to be weakly associated with monetary policy uncertainty. Our findings on state-dependent biases can be used to improve the forecasts; for instance, by adjusting for a bias only in monetary easing episodes.

The chapter is organized as follows. Section 2 introduces the econometric framework and formalizes the null hypothesis. Section 3 introduces the proposed test statistics. Section 4 provides a Monte Carlo analysis of the size and power of our proposed procedures, while Section 5 illustrates the usefulness of our test in an empirical analysis. Section 6 concludes.

2. ECONOMETRIC FRAMEWORK

We consider the situation where the researcher has a series of out-of-sample predictions, $y_{t,h}$, made at time t , h -periods into the future, whose corresponding realizations are denoted by y_{t+h} , for $t = 1, \dots, T$. Let $\epsilon_{t,h} \equiv y_{t+h} - y_{t,h}$ denote the forecast error. We are interested in testing whether the forecast error is unbiased,

efficient, and rational (i.e., jointly unbiased and efficient) – while being able to detect regime-switching deviations from the respective forecast rationality property. For simplicity, consider the leading case of a forecast rationality regression with two regimes:

$$\epsilon_{t,h} = \beta x_{t+h} + S_{t+h} \beta_s x_{t+h} + \sum_{i=1}^d \phi_i \epsilon_{t-i,h} + e_{t+h}, \quad (1)$$

where S_{t+h} is a latent, stationary Markov chain with $S_{t+h} \in \{0, 1\}$; e_{t+h} is a mean zero error term with a constant variance; $x_{t+h} = [1, y_{t,h}]'$; and ϕ_i are the lag coefficients, satisfying standard stationarity restrictions, included to control for potential autocorrelation.³ The stationary Markov chain S_{t+h} is characterized by the two state-to-state transition probabilities (p, q) , which take values between zero and one. The vectors $\beta = (\mu, \gamma)$ and $\beta_s = (\mu_s, \gamma_s)$ contain the relevant parameters for rationality regressions. The parameters μ and μ_s are the relevant parameters for the unbiasedness test, while γ and γ_s (the regression coefficients on the forecasts) are the relevant parameters for the efficiency test. Note that the vector x_{t+h} can be extended to contain more regressors, making our test applicable, in general, to all regression-based tests of predictive ability. For notational simplicity, we drop the autoregressive coefficients (ϕ_i) in what follows. This simplification is inconsequential since they are neither parameters of interest nor state-dependent. Throughout the paper, iid denotes independently and identically distributed.

Rationality: The test for rationality is a joint test of unbiasedness and efficiency, and our null and alternative hypothesis are:

$$H_0^R : \beta = \beta_s = 0 \text{ vs. } H_A^R : \beta \neq 0, \beta_s \neq 0, \text{ or } (\beta, \beta_s) \neq 0. \quad (2)$$

In contrast, traditional tests of Markov switching (Carrasco et al., 2014; Garcia, 1998; Hansen, 1992; Qu & Zhuo, 2021) consider the null hypothesis

$$H_0^{\text{MS}} : \beta_s = 0,$$

but leave the value of β unspecified under the null. Traditional tests of forecast rationality, on the other hand, (Mincer & Zarnowitz, 1969; West & McCracken, 1998) consider the model

$$\epsilon_{t,h} = \beta x_{t+h} + e_{t+h},$$

and restrict the value of β to be equal to zero under the null hypothesis, while β_s is not part of the model's parameter space.

Unbiasedness: In the special case of unbiasedness tests, $x_{t+h} = 1$, our null and alternative hypotheses are:

$$H_0^U : \mu = \mu_s = 0 \text{ vs. } H_A^U : \mu \neq 0, \mu_s \neq 0, \text{ or } (\mu, \mu_s) \neq 0. \quad (3)$$

Existing tests for Markov switching test the null hypothesis of no time variation:

$$H_0^{\text{MS}} : \mu_s = 0,$$

but do not impose $\mu = 0$. However, the additional restriction of

$$H_0^U : \mu_s = \mu = 0$$

is important in order to have power against a constant deviation from forecast rationality. Traditional tests for unbiasedness ([Mincer & Zarnowitz, 1969](#); [West & McCracken, 1998](#)) implement the regression

$$\epsilon_{t,h} = \mu + e_{t+h},$$

where the null hypothesis is that μ is equal to zero, and μ_s is not part of the model's parameter space. Consequently, the tests lack power in the case of Markov switching deviations from forecast unbiasedness.

Efficiency: In the special case of efficiency tests, $x_{t+h} = [y_{t,h}]$, our null and alternative hypotheses are

$$H_0^E : \gamma = \gamma_s = 0 \text{ vs. } H_A^E : \gamma \neq 0, \gamma_s \neq 0, \text{ or } (\gamma, \gamma_s) \neq 0. \quad (4)$$

The null in existing tests for Markov switching is

$$H_0^{\text{MS}} : \gamma_s = 0,$$

and the value of γ is unrestricted under the null hypothesis. On the other hand, traditional forecast efficiency tests ([Mincer & Zarnowitz, 1969](#); [West & McCracken, 1998](#)) implement the regression

$$\epsilon_{t,h} = y_{t,h}\gamma + e_{t+h},$$

where the value of γ is restricted to be zero under the null hypothesis, and γ_s is not part of the model's parameter space.

3. TESTING FOR MARKOV SWITCHING RATIONALITY

This section introduces our Markov switching forecast rationality tests, inspired by [Qu and Zhuo \(2021\)](#) and [Hansen \(1992\)](#). We also consider a test inspired by [Garcia \(1998\)](#) in the Online Appendix.

3.1. AFE-BS Test for Forecast Rationality

[Qu and Zhuo \(2021\)](#) showed that when testing for Markov-switching, a parametric bootstrap consistently approximates the asymptotic distribution of the LR test as

long as it correctly reproduces the covariance function of the limiting distribution derived in [Qu and Zhuo \(2021\)](#) Proposition 1.

Recall that the model under the null hypothesis reduces to an AR(d). Following [Qu and Zhuo \(2021\)](#), let $\Lambda_{(p,q)} = \{(p,q) : 0.02 \geq p, q \leq 0.98 \text{ and } p+q \geq 1.02\}$ denote the set of feasible values for (p, q) . Let $\text{LL}_{0,T}$ and $\text{LL}_{A,T,\Lambda_{(p,q)}}$ (note that we subsequently drop the subscript $\Lambda_{(p,q)}$ for notational convenience) denote the log-likelihood under the null and under the alternative hypothesis, respectively.⁴ Let $\text{LR}_T = 2(\text{LL}_{0,T} - \text{LL}_{A,T,\Lambda_{(p,q)}})$ denote the likelihood-ratio.

Parametric bootstrap for testing forecast rationality: In order to construct the LR test, we need to evaluate the likelihood under both the null and the alternative. This requires us to re-sample both $\epsilon_{t,h}$ and x_{t+h} , respecting their covariance structure.⁵ Let $\hat{\phi}_{i,0}$, for $i = 1, \dots, d$, denote the parameter estimates of the autoregressive coefficients under the null. Let $\hat{\phi}_{i,x}$ and $\hat{\phi}_{i,\epsilon_x}$, for $i = 1, \dots, d_x$, denote the parameter estimates of a regression of x_{t+h} on d_x lags of x_{t+h} and $\epsilon_{t,h}$ jointly. Let \hat{e}_{t+h} and $\hat{e}_{x,t+h}$ denote the estimated error term of the regression of $\epsilon_{t,h}$ on its own lags and of the regression of x_{t+h} on its own lags and lags of $\epsilon_{t,h}$, respectively. Furthermore, let $\hat{\Sigma}_e$ denote the covariance matrix of $[\hat{e}_{t+h}, \hat{e}_{x,t+h}]'$, and $d^* = \max(d, d_x)$. Then, for $j = 1, \dots, J$, we proceed with the following steps:

1. Draw $T + d^*$ iid random variables from $N(0, \hat{\Sigma}_e)$ and denote by $\{v_{t,j}^*\}_{t=-d^*+1}^T$ the set of draws.
2. Construct a series $\epsilon_{t,h,j}^*$ and $x_{t+h,j}^*$, for $t = 1, \dots, T$, using $v_{t,j}^*$, $\hat{\phi}_{i,0}$ for $i = 1, \dots, d$, and $(\hat{\phi}_{i,x}, \hat{\phi}_{i,\epsilon_x})$, for $i = 1, \dots, d_x$. We elaborate on this step below.
3. Using $\{\epsilon_{t,h,j}^*, x_{t+h,j}^*\}_{t=1}^T$, compute the bootstrap log-likelihood under the null, $\text{LL}_{0,T,j}^*$, and under the alternative, $\text{LL}_{A,T,j}^*$.
4. Store the bootstrapped likelihood ratio: $\text{LR}_{T,j}^* = 2(\text{LL}_{A,T,j}^* - \text{LL}_{0,T,j}^*)$.

After J iterations, we obtain a set of the bootstrapped likelihood-ratio statistic, $\{\text{LR}_{T,j}^*\}_{j=1}^J$, that approximates the asymptotic distribution.

For the case of forecast unbiasedness, $x_{t+h} = 1$, the researcher only has to re-sample $\epsilon_{t,h}$ and, therefore, $v_{t,j}^*$ is a scalar with $v_{t,j}^* \sim_{\text{iid}} N(0, \hat{\sigma}_{e,0}^2)$, where $\hat{\sigma}_{e,0}^2$ denotes the estimated variance of \hat{e}_{t+h} . Then, if $d > 0$, set $(\epsilon_{-d+1,h,j}^*, \dots, \epsilon_{0,h,j}^*)$ equal to $(1 - \sum_i^d \hat{\phi}_{i,0})^{-1}(v_{-d+1,h,j}^*, \dots, v_{0,h,j}^*)$. We further generate $\epsilon_{t,h,j}^* = \sum_{i=1}^d \hat{\phi}_{i,0} \epsilon_{t-i,h,j}^* + v_{t,h,j}^*$ recursively for $t = 1, \dots, T$; if $d = 0$, set $\epsilon_{t,h,j}^* = v_{t,h,j}^*$.

For the case of forecast efficiency, we need to resample $x_{t+h} = y_{t,h}$ and $\epsilon_{t,h}$ jointly. The bootstrap procedure of [Qu and Zhuo \(2021\)](#) does not directly apply in this case, and they do not recommend using a fixed-regressor bootstrap. Instead, we implement the following procedure following the recommendation of [Qu and Zhuo \(2021\)](#). If the DGP is $y_t = \psi y_{t-1} + u_t$ and u_t is Gaussian white noise, the forecasting model will be $y_{t,1} = \psi y_t$, such that $x_{t+1} = y_{t,1}$. The forecast error subsequently is $\epsilon_{t,1} = y_{t+1} - y_{t,1} = u_{t+1}$. Then, the researcher can re-sample $\epsilon_{t,1}$ and x_{t+1} as follows:

$$\epsilon_{t,1,j}^* \sim_{\text{iid}} N(0, \hat{\sigma}_{e,0}^2), \text{ where } \hat{\sigma}_{e,0}^2 \text{ is estimated using } \hat{e}_{t+h}, \quad x_{t+1,j}^* = \hat{\psi}(x_{t,j}^* + \epsilon_{t-1,1,j}^*),$$

and $x_{0,j}^* \sim_{\text{iid}} N\left(0, \frac{\hat{\psi}^2}{1 - \hat{\psi}^2} \hat{\sigma}_{e,0}^2\right)$.

3.2. AFE-H Test for Rationality

Let α_0 denote the parameter vector under our null hypotheses, formulated in Section 2, and let $\alpha \in A$, with A being a compact metric space, denote a given alternative. Hansen (1992) considers the LR as an empirical process indexed by the parameters of interest, $\alpha = (\beta_s, p, q)$, where (p, q) are transition probabilities, and depending further on the nuisance parameters, $\theta = (\beta, \phi_1, \dots, \phi_d, \sigma)$. To use the strategy of Hansen (1992) for testing our joint null hypothesis, we need to partition the parameter space differently since our null hypotheses specify both β and β_s . Therefore, we cannot treat β as a nuisance parameter, instead, we must add it to the vector of parameters of interest.

Therefore, the three relevant parameter vectors for us are $\alpha = (\beta, \beta_s, p, q)$, $\alpha = (\mu, \mu_s, p, q)$, and $\alpha = (\gamma, \gamma_s, p, q)$, for testing rationality, unbiasedness, and efficiency, respectively. The vector of nuisance parameters reduces to the lag coefficients and the standard deviation, $\theta = (\phi_1, \dots, \phi_d, \sigma)$. The subsequent derivation follows closely Hansen (1992). Let us define

$$\hat{\theta} = \max_{\theta \in \Theta} L_T(\alpha_0, \theta)$$

to be the maximum likelihood estimator (MLE) of the nuisance parameters under the null, α_0 , and let

$$\hat{\theta}(\alpha) = \max_{\theta \in \Theta} L_T(\alpha, \theta(\alpha))$$

denote the MLE of the nuisance parameters under the alternative α . The LR is defined as

$$\widehat{LR}_T(\alpha) = L_T(\alpha, \hat{\theta}(\alpha)) - L_T(\alpha_0, \hat{\theta}),$$

with

$$L_T(\alpha, \hat{\theta}(\alpha)) = \sum_{t=1}^T \ell_t(\alpha, \hat{\theta}(\alpha)) \quad \text{and} \quad L_T(\alpha_0, \hat{\theta}) = \sum_{t=1}^T \ell_t(\alpha_0, \hat{\theta}),$$

where ℓ_t denotes the log-likelihood of observation t , which is allowed to exhibit serial correlation and heterogeneity.⁶ As in Hansen (1992), the LR is split into its expected value, $R_T(\alpha)$, and its deviation from that expectation, denoted by $Q_T(\alpha)$,

$$\widehat{LR}_T(\alpha) = R_T(\alpha) + Q_T(\alpha) + O_p(1),$$

where

$$R_T(\alpha) = E[L_T(\alpha, \theta(\alpha)) - L_T(\alpha_0, \theta)] = E\left[\sum_{t=1}^T [\ell_t(\alpha, \theta(\alpha)) - \ell_t(\alpha_0, \theta)]\right],$$

and

$$Q_T(\alpha) = \sum_{t=1}^T q_t(\alpha) = \sum_{t=1}^T [\ell_t(\alpha, \theta(\alpha)) - \ell_t(\alpha_0, \theta) - E[\ell_t(\alpha, \theta(\alpha)) - \ell_t(\alpha_0, \theta)]],$$

where $q_t(\alpha)$ denotes the deviation of the LR of observation t from its expectation, and θ and $\theta(\alpha)$ denote the large sample values of the MLE of $\hat{\theta}$ and $\hat{\theta}(\alpha)$. Note that $Q_T(\alpha)$ has a mean of zero. The parameter estimation error that is present in the sample analog of $R_T(\alpha)$ and $Q_T(\alpha)$ is included in the term $O_p(1)$. Please see the Online Appendix for details.

Under the null, $R_T(\alpha) \leq 0$ since the value of $R_T(\alpha)$ is maximized at the true parameter α_0 (under the null). It follows that

$$\frac{1}{\sqrt{T}} \widehat{LR}_T(\alpha) \leq \frac{1}{\sqrt{T}} Q_T(\alpha) + o_p(1).$$

Let $V_T(\alpha)$ denote the variance of the $q_t(\alpha)$. For a fixed α , and by standardizing with $V_T(\alpha)$, the zero mean process $Q_T(\alpha)$ converges to a standard Normal distribution by a Central Limit Theorem (CLT):

$$\frac{1}{\sqrt{T}} \frac{Q_T(\alpha)}{V_T^{1/2}(\alpha)} = \frac{1}{\sqrt{T}} Q_T^*(\alpha) \xrightarrow{d} N(0, 1). \quad (5)$$

The asymptotic distribution of the bound, $\frac{1}{\sqrt{T}} Q_T^*(\alpha)$, uniformly over $\alpha \in A$, can be derived by applying an empirical process CLT, using the assumptions stated in Hansen (1992), to eq. (5):

$$\sup_{\alpha \in A} \frac{1}{\sqrt{T}} \widehat{LR}_T^*(\alpha, \theta(\alpha)) \leq \sup_{\alpha \in A} \frac{1}{\sqrt{T}} Q_T^*(\alpha) + o_p(1) \xrightarrow{d} \sup_{\alpha \in A} Q^*(\alpha). \quad (6)$$

The process $Q^*(\alpha)$ is Gaussian with covariance function:

$$K^*(\alpha_1, \alpha_2) = \frac{\sum_{k=-\infty}^{\infty} E q_t(\alpha_1) q_{t+k}(\alpha_2)}{V(\alpha_1)^{\frac{1}{2}} V(\alpha_2)^{\frac{1}{2}}},$$

where $V(\alpha_i)$ denotes the probability limit of the sample analog $V_T(\alpha_i)$.

Since the covariance function, $K^*(\cdot, \cdot)$, depends through $q_t(\alpha_i)$ on the data, critical values cannot be tabulated for the general case. Instead, analog to Hansen (1992), critical values can be approximated by drawing iid Gaussian processes that have covariance function $\hat{K}^*(\cdot, \cdot)$, the empirical counterpart of the unknown $K^*(\cdot, \cdot)$. Doing so is straightforward and requires the simulation of

$$\hat{Q}_T^{*j}(\alpha_i) = \frac{\sum_{m=0}^M \sum_{t=1}^T \hat{q}_t(\alpha_i) v_{t+m}^j}{\sqrt{1+MV_T(\alpha_i)^2}},$$

based on J replications, where the v_{t+m}^j , for $j = 1, \dots, J$, are iid $N(0, 1)$ variates, and $\hat{q}_t(\alpha_i)$ is the empirical counterpart of $q_t(\alpha_i)$, and $\hat{V}_T(\alpha_i)$ is the empirical

counterpart of $V_T(\alpha_i)$. The $\widehat{Q}_T^{*j}(\alpha_i)$ have $\widehat{K}^*(\cdot, \cdot)$ as a covariance function and, hence, approximate the asymptotic distribution. Moreover, Hansen (1996) points out that the likelihood components $q_t(\alpha_i)$ are serially correlated even if the data is iid and, therefore, a Bartlett kernel is used to account for the autocorrelation. The Bartlett's bandwidth parameter, M , can be data dependent; typical choices are $M = T^{1/4}$ or $M = [4(T/100)^{2/9}] + 1$.¹⁷ Critical value are then obtained as percentiles from $\{\widehat{Q}_T^{*j}\}_{j=1}^J$, with $\widehat{Q}_T^{*j} = \sup_{\alpha \in A} \widehat{Q}_T^{*j}(\alpha_i)$.

To obtain a set $\{\widehat{Q}_T^{*j}\}_{j=1}^J$, one has to estimate the model under the alternative over a grid of values for $\alpha = (\beta, \beta_s, p, q)$; let α_{grid} denote the set of grid points. In particular, first, estimate the model under the null and store the log-likelihood values. Second, estimate the model under the alternative using a grid point $\alpha_i \in \alpha_{grid}$ and compute the log-likelihood values. Third, use the log-likelihood values under the null and the alternative to compute $\widehat{LR}_T^*(\alpha_i, \theta(\alpha_i))$, $\{\widehat{q}_t(\alpha_i)\}_{t=1}^T$, and $\widehat{V}_T(\alpha_i)$, where $\widehat{q}_t(\alpha_i)$ is computed by replacing the terms $[\ell_t(\alpha, \theta(\alpha)) - \ell_t(\alpha_0, \theta)]$ and $E[\ell_t(\alpha, \theta(\alpha)) - \ell_t(\alpha_0, \theta)]$ with their sample analogues. Fourth, compute J draws of $\widehat{Q}_T^{*j}(\alpha_i)$ using the $\{\widehat{q}_t(\alpha_i)\}_{t=1}^T$, $\widehat{V}_T(\alpha_i)$, and draws of iid random standard Normal variates $\{v_{t+m}^j\}_{t=1}^T$. Finally, repeat the second to fourth step for all grid points in α_{grid} and compute the supremum of $\widehat{LR}_T^*(\alpha_i, \theta(\alpha_i))$ and $\widehat{Q}_T^{*j}(\alpha_i)$ over the grid values for α_i to obtain the maximum test statistic and the simulated asymptotic distribution.

Table 1 displays the average critical values we obtain when we implement the AFE-H test of unbiasedness and efficiency. Let the data be generated by $y_t = e_t$, where $e_t \sim_{iid} N(0, 1)$. The partitions of the parameter vectors are in this case $\alpha = (\mu, \mu_s, p, q)$, and $\alpha = (\gamma, \gamma_s, p, q)$, respectively. The column denoted by “H” shows the critical values for the original Hansen (1992) null, $H_0^{MS} : \mu_s = 0$ and $H_0^{MS} : \gamma_s = 0$, with $\alpha_U = (\mu_s, p, q)$ and $\alpha_E = (\gamma_s, p, q)$. As the approximation of the asymptotic distribution is data dependent, the numbers are obtained by averaging the critical values over all Monte Carlo replications. The aim of this exercise is not to tabulate critical values, which would be invalid due to the data dependence of the asymptotic distribution, but to show how the additional parameter restriction changes the critical values we obtain. As expected, the critical values of the AFE-H test are larger on average, reflecting the additional restriction of the AFE-H on the null parameter space.

Table 1. Average Critical Values.

Nominal Size	Unbiasedness		Efficiency	
	AFE-H	H	AFE-H	H
1%	3.60	2.80	3.44	3.24
5%	3.01	2.51	2.84	2.64
10%	2.72	2.18	2.53	2.33

Notes: The table shows the average critical values based on our simulations for the proposed AFE-H test and the original Hansen (1992), labeled “H,” test for a standard Normal DGP, a sample size of $T = 500$, and 500 Monte Carlo simulations.

As mentioned, in order to implement the AFE-H test in practice, the researcher needs to decide on the grid values for (p, q) and (β, β_s) . When evaluating the LR process under the Markov switching alternative, for each point on the grid, the researcher will optimize a constrained (imposing the grid point values) likelihood to obtain $\hat{\theta}(\alpha)$. The model under the null hypothesis, on the other hand, is estimated with the constraint that $(\beta = 0, \beta_s = 0)$, i.e., there is no Markov switching present by assumption. The grid points under the alternative serve as a basis for the construction of the test statistics as well the limiting distribution, thus deserving particular interest.

Since (p, q) are bounded below by zero and above by one, the grid choice for (p, q) is only about the number of grid points to consider; we used 12 grid points in our Monte Carlo and did not experience the results to be very sensitive to slightly different choices. In addition, state-to-state transition probabilities in Markov switching models are often well above 0.5, such that the researcher may as well restrict the grid of (p, q) accordingly. The grid choice for (β, β_s) is somewhat more difficult since their domain is not restricted to be between zero and one. Although the grid values for (β, β_s) will vary with the empirical application, the researcher can typically rule out large values for the grid since the left hand side variable is the forecast error, which tends to be small. In general, we recommend to plot the data, and to estimate an unrestricted Markov switching rationality regression to get an idea of where to set the grid for (β, β_s) in practice. In our simulations the performance of the tests are not very sensitive to the choice of the grid points. Relative to [Hansen \(1992\)](#), our proposed procedure could be somewhat more computationally intensive, since it requires evaluating the constrained likelihood with a grid structure for an additional parameter, β .

The estimation of the Markov switching model (with or without grid) is implemented efficiently using the expectation-maximization (EM) algorithm described in [Hamilton \(1990\)](#). However, the researcher can rely on other approaches, as for instance, on the filtering techniques for endogenous switching outlined in ([Chang et al., 2017](#)).

4. MONTE CARLO SIMULATION RESULTS

This section provides Monte Carlo evidence on the finite sample size and power of the unbiasedness and efficiency tests. In all instances, the estimation of the Markov switching model is based on the EM algorithm ([Hamilton, 1990](#)).

4.1. Monte Carlo Results – Unbiasedness

The DGP is the following in this section:

$$y_t = \psi y_{t-1} + u_t, \quad (7)$$

where $|\psi| < 1$ and $u_t \sim_{\text{iid}} N(0, 1)$. We then consider three different forecasting situations that lead to three different cases of forecast errors.

Case 1: Forecasting one-step-ahead with an AR(1)

$$y_{t,1} = \psi y_t \quad \text{and} \quad \epsilon_{t,1} = u_{t+1}, \quad (8)$$

where we, instead, model the forecast error as $\epsilon_{t,1} = \mu + S_{t+1}\mu_s + e_{t+1}$ and set $\psi = 0.5$.

Case 2: Forecasting one-step-ahead with a constant

$$y_{t,1} = c, \quad \epsilon_{t,1} = \psi y_t + u_{t+1}, \quad (9)$$

where we, instead, model the forecast error as $\epsilon_{t,1} = \mu + S_{t+1}\mu_s + \phi_1\epsilon_{t-1,1} + e_{t+1}$, set $c = 0$, and set $\psi = 0.5$.

Case 3: Forecasting multi-step-ahead When forecasting two-periods-ahead with an AR(1) model, the errors will have a MA(1) dynamic in population:

$$y_{t,2} = \psi^2 y_t, \quad \epsilon_{t,2} = (1 + \psi L)u_{t+2}, \quad (10)$$

where we set $\psi = 0.25$.

In the application of our AFE-BS and AFE-H tests, we approximate the MA(1) error dynamics with a Markov switching AR(1) process: $\epsilon_{t,2} = \mu + S_{t+1}\mu_s + \phi_1\epsilon_{t-1,2} + e_{t+2}$.

In all cases, μ denotes the intercept and μ_s is the parameter that changes with the Markov switching regime. S_{t+1} is a stationary Markov chain and $e_t \sim_{\text{iid}} N(0, \sigma^2)$. Note that the Markov switching specification correctly approximates the forecast error's dynamics in Case 1 and Case 2, while the Markov switching model's dynamics are misspecified in Case 3. This case intends to emulate a realistic forecast situation where the researcher has a multiple-step-ahead forecast at hand and controls for potential serial correlation in the forecast error via an autoregressive specification, which is typically easier to estimate than a MA specification.

The null hypothesis of the AFE-BS and AFE-H tests imposes the restriction $\mu = \mu_s = 0$. Panel A of [Table 2](#) shows the size results for AFE-BS and AFE-H tests for a nominal size of 5%. Panel B, instead, shows results for a nominal size of 10%. We also include results for the tests of [West and McCracken \(1998\)](#) (labeled “WM”) and [Rossi and Sekhposyan \(2016\)](#) (labeled “Fluctuation”), which test for constant unbiasedness and time-varying unbiasedness, respectively. As discussed previously, [Rossi and Sekhposyan \(2016\)](#) capture time variation non-parametrically, based on a rolling window estimation.⁸ Overall, the size results of the AFE-BS and AFE-H tests are good, although AFE-BS performs better in small and medium-sized samples relative to the AFE-H. The AFE-H test over-rejects for small and medium-sized samples; however, the size distortions are of a similar magnitude as in [Hansen \(1992\)](#). The mild misspecification in the forecast error dynamics in Case 3 only leads to small size distortions for the AFE-BS and AFE-H tests.⁹

Table 2. Size Results – Forecast Unbiasedness Test.

Panel A. Nominal size 5%										
Test	Case 1			Case 2			Case 3			
	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	
AFE-BS	0.048	0.047	0.058	0.045	0.043	0.045	0.046	0.026	0.032	
AFE-H	0.122	0.082	0.046	0.130	0.058	0.026	0.144	0.076	0.056	
WM	0.046	0.048	0.054	0.107	0.110	0.070	0.062	0.062	0.056	
Fluct.	0.062	0.062	0.060	0.218	0.187	0.126	0.134	0.128	0.081	

Panel B. Nominal size 10%										
Test	Case 1			Case 2			Case 3			
	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	T = 100	T = 200	T = 500	
AFE-BS	0.097	0.109	0.110	0.093	0.097	0.088	0.077	0.068	0.064	
AFE-H	0.176	0.122	0.086	0.182	0.100	0.060	0.198	0.114	0.096	
WM	0.099	0.091	0.099	0.178	0.167	0.133	0.121	0.115	0.103	
Fluct.	0.112	0.112	0.115	0.299	0.262	0.193	0.208	0.207	0.138	

Notes: T denotes the sample size. Cases 1, 2, and 3 refer to the various simulation designs described in Section 4.1. Results are based on 1,000 Monte Carlo replications, except for AFE-H: due to the computational time, these Monte Carlo replications are limited to 500. The results for AFE-H test are based on a 4-tuple of 12 equally-spaced grid points for $(p, q) \in [0.05, 0.95]$ and 20 equally-spaced grid points for $(\mu, \mu_s) \in [-1, 1] \times [-2, 2]$. Results for the AFE-BS test are based on 200 bootstrap replications and $(p, q) \in \Lambda_{(p,q)}$. The window size m for the Fluctuation test is set to $m = T/2$.

To study power, we consider first the alternative of a constant deviation from unbiasedness. The DGP takes the form of

$$y_t = \tilde{\mu} + \psi y_{t-1} + u_t, \quad (11)$$

with $\psi = 0.5$, where the forecasting model is $y_{t,1} = \psi y_t$, such that the forecast error becomes $\epsilon_{t,1} = \tilde{\mu} + u_{t+1}$. The different values for $\tilde{\mu}$ are $[0.20, 0.25, 0.30, 0.35, 0.375, 0.40, 0.45, 0.50]$.

Panel A of Table 3 shows size-adjusted power results for a sample size of $T = 100$ and a nominal size of 5%. As expected, the WM test has the highest power against a constant deviation from the null hypothesis of unbiasedness. However, the AFE-BS and AFE-H test exhibits good power as well and the power increases rapidly with the magnitude of the deviation from unbiasedness. The power of the Fluctuation test is comparable and only slightly worse than that of the AFE-BS test against the constant alternative.

To test for power against the alternative of Markov switching, the DGP takes the form of

$$y_t = \tilde{\mu} + S_t \tilde{\mu}_s + \psi y_{t-1} + u_t, \quad (12)$$

with $\psi = 0.5$, where the forecasting model is $y_{t,1} = \psi y_t$, such that the forecast error becomes $\epsilon_{t,1} = \tilde{\mu} + S_{t+1} \tilde{\mu}_s + u_{t+1}$.

We set the state-to-state transition probabilities of the Markov chain S_t to be $(p, q) = (0.9, 0.9)$ and impose $\tilde{\mu} = -\tilde{\mu}_s / 2$. These parameter choices ensure that the unconditional mean of the series is zero, i.e. $E(\epsilon_{t,1}) = 0$, such that we can

Table 3. Power Results – Unbiasedness.

	Panel A. Constant bias							
	Values of $\bar{\mu}$							
	0.20	0.25	0.30	0.35	0.375	0.40	0.45	0.50
AFE-BS	0.28	0.47	0.63	0.78	0.88	0.90	0.96	0.98
AFE-H	0.32	0.50	0.66	0.78	0.87	0.89	0.96	0.99
WM	0.49	0.71	0.82	0.92	0.97	0.98	0.99	0.99
Fluct.	0.33	0.49	0.65	0.76	0.84	0.88	0.95	0.95
	Panel B. Markov switching bias							
	Values of $\bar{\mu}_s$							
	0.80	1.00	1.20	1.40	1.50	1.60	1.80	2.00
AFE-BS	0.28	0.50	0.73	0.86	0.91	0.96	0.98	1.00
AFE-H	0.14	0.24	0.33	0.53	0.58	0.67	0.80	0.88
WM	0.13	0.16	0.17	0.22	0.21	0.27	0.28	0.23
Fluct.	0.20	0.27	0.32	0.37	0.39	0.42	0.48	0.44

Notes: The values denote the size-adjusted empirical rejection frequency based on 500 Monte Carlo replications. The values for μ and μ_s are given in the first rows of Panel A and B, respectively. The nominal size is 5%. The results for AFE-H are based on a 4-tuple of 12 equally-spaced grid points for $(p,q) \in [0.05, 0.95]$ and 20 equally-spaced grid points for $(\mu, \mu_s) \in [-1, 1] \times [-2, 2]$. Results for the AFE-BS test are based on 200 bootstrap replications and $(p,q) \in \Lambda_{(p,q)}$. The window size m for the Fluctuation test is set to $m = T/2$.

compute the power against the alternative of Markov switching only. The different values that we explore for $\tilde{\mu}_s$ are $[0.80, 1.00, 1.20, 1.40, 1.50, 1.60, 1.80, 2.00]$.

Panel B of Table 3 displays the size-adjusted rejection frequencies at a nominal size of 5%. The AFE-BS and AFE-H tests exhibit strong power against the alternative of Markov switching.

The rejection frequency of the WM test would theoretically be expected to remain at the nominal level of 5%. However, in small samples, it is quite likely to sample one of the states more often than the other, even if the unconditional state probabilities are 0.5, which shifts the sample mean away from zero (this only occurs in small samples with a high regime persistence).

When looking at the Fluctuation test, we find that it does not have strong power against Markov-switching type of time variation. This result is driven by the non-parametric approach of the test, i.e., it has less power against parametric discrete switches. Note, however, that the power results of the Fluctuation test depend to some extent on the window size – smaller windows would likely improve the tests' power under Markov switching. AFE-H exhibits a lower power than AFE-BS; however, note that the grid size used for the test statistic of μ and μ_s could influence the results (though it did not seem to be sensitive to the grid choice in our Monte Carlo exercises).

4.2. Monte Carlo Results – Efficiency

We now turn to test forecast efficiency. Under the null, the DGP is the same as in eq. (7), i.e.,

$$y_t = \psi y_{t-1} + u_t, \quad (13)$$

Table 4. Size Results – Efficiency.

Test	$T = 100$	$T = 200$	$T = 500$	$T = 100$	$T = 200$	$T = 500$
	Nominal size 5%			Nominal size 10%		
AFE-BS	0.055	0.058	0.053	0.098	0.113	0.094
AFE-H	0.020	0.014	0.012	0.036	0.030	0.028
WM	0.059	0.067	0.053	0.114	0.120	0.107
Fluct.	0.050	0.049	0.046	0.094	0.092	0.095

Notes: T denotes the sample size. Results are based on 1,000 Monte Carlo replications, except for AFE-H: due to the computational time, these Monte Carlo replications are limited to 500. The results for AFE-H test are based on a 4-tuple of 12 equally-spaced grid points for $(\gamma, \gamma_s) \in [-1, 1] \times [-2, 2]$ and 20 equally-spaced grid points for $(p, q) \in [0.05, 0.95]$. Results for the AFE-BS test are based on 200 bootstrap replications and $(p, q) \in \Lambda_{(p,q)}$. The window size m for the Fluctuation test is set to $m = T/2$.

where we set $\psi = 0.5$ and $u_t \sim_{\text{iid}} N(0, \sigma_e^2)$. The forecasting model takes the form of $y_{t,1} = \psi y_t$ such that the true forecast error becomes

$$\epsilon_{t,1} = u_{t+1}. \quad (14)$$

We use the following Markov switching specification

$$\epsilon_{t,1} = \gamma y_{t,1} + S_{t+1} \gamma_s y_{t,1} + e_{t+1} \quad (15)$$

to test the null hypothesis: $\gamma = \gamma_s = 0$ in the following regression, where S_{t+1} is a stationary Markov chain and $e_{t+1} \sim_{\text{iid}} N(0, 1)$.

Table 4 shows the size results for AFE-H and AFE-BS tests as well as the WM and the Fluctuation tests. The size results of AFE-BS test are good, even in small samples; the AFE-H somewhat underrejects for small and large samples. The reason for the distortions could be that the critical values are taken from a bound instead of an exact distribution and, therefore, the test is more conservative, or that the test is more sensitive to the choice of the grid for (γ, γ_s) .

To study power, we proceed as follows. Under the alternative of a constant, but non-zero efficiency coefficient, the DGP takes the form

$$y_t = (\psi + \tilde{\gamma}) y_{t-1} + u_t, \quad (16)$$

with the forecasting model being $y_{t,1} = \psi y_t$, such that the forecast error becomes $\epsilon_{t,1} = \tilde{\gamma} y_t + u_{t+1}$, where we let $\tilde{\gamma}$ take the following values: [0.15, 0.20, 0.25, 0.30, 0.35].

Panel A of **Table 5** shows the size-adjusted power results for a sample size of $T = 100$ at a nominal size of 5%. Again the WM test outperforms the other tests in terms of power. However, the AFE-BS and AFE-H have good power against the null of a constant deviation as well.

To test for the alternative of Markov switching, the DGP takes the form

$$y_t = (\psi + \tilde{\gamma}) y_{t-1} + S_t \tilde{\gamma}_s y_{t-1} + u_t, \quad (17)$$

where the forecasting model is $y_{t,1} = \psi y_t$, such that the forecast error becomes $\epsilon_{t,1} = \tilde{\gamma} y_t + S_{t+1} \tilde{\gamma}_s y_t + u_{t+1}$. We set the state-to-state transition probabilities

Table 5. Power Results – Efficiency.

	Constant Deviation					Markov Switching Deviation				
	Values of $\tilde{\gamma}$					Values of $\tilde{\gamma}_s$				
	0.15	0.20	0.25	0.30	0.35	0.30	0.40	0.50	0.60	0.70
AFE-BS	0.30	0.51	0.74	0.91	0.99	0.07	0.22	0.45	0.74	0.92
AFE-H	0.38	0.58	0.80	0.94	1.00	0.06	0.15	0.32	0.59	0.82
WM	0.46	0.67	0.87	0.96	1.00	0.05	0.09	0.12	0.15	0.18
Fluct.	0.44	0.58	0.74	0.88	0.98	0.08	0.12	0.15	0.20	0.27

Notes: The values denote the size-adjusted empirical rejection frequency based on 500 Monte Carlo replications. The values for γ and γ_s are given in the first row of Panel A, and B respectively. The nominal size is 5%. The results for AFE-H test are based on a 4-tuple of 12 equally-spaced grid points for $(p, q) \in [0.05, 0.95]$ and 20 equally-spaced grid points for $(\gamma, \gamma_s) \in [-1, 1] \times [-2, 2]$. Results for the AFE-BS test are based on 200 bootstrap replications and $(p, q) \in \Lambda_{(p,q)}$. The window size m for the Fluctuation test is set to $m = T/2$.

to be $(p, q) = (0.9, 0.9)$, set $\tilde{\gamma} = -\tilde{\gamma}_s/2$, and let $\tilde{\gamma}_s$ take the following values: $[0.30, 0.50, 0.70, 0.90, 1.10]$.¹⁰

Results are shown in Panel B of Table 5. We find that the traditional WM and Fluctuation tests have less power against the alternative of Markov switching efficiency than the AFE-H and AFE-BS tests.

4.3. Discussion

Testing for Markov switching is challenging and both of the proposed tests have advantages and disadvantages.

When using the AFE-H test, the researcher needs to carefully set the grid of parameters of interest. When testing unbiasedness, we recommend plotting the forecast error to decide the grid values. When testing efficiency, setting a grid around the full sample efficiency parameter could be a natural starting point. In addition, the AFE-H has the drawback of being computationally intensive and displaying size distortions in small samples, but the presence of an additional control variable (beyond the autoregressive lags) does not require the researcher to specify a law of motion for this variable.

While the parametric bootstrap procedure requires the researcher to make this additional assumption, which might be difficult in some situations, such as when testing for forecast efficiency of survey forecasts, it addresses many of the technical issues associated with testing for Markov switching. In addition, it has good small sample properties and the researcher only has to search over a grid for the state-to-state transition probabilities.

So far, we have let the variance of e_t be constant; however, that can be relaxed. For instance, the variance can follow a Markov switching process itself. If the variance shares the same regime dynamics as the rationality coefficients, then it can help identify the regime. However, in this case, a rejection of a null hypothesis would not indicate whether the rejection is due to violations of rationality or switches in the variance. Instead, if the variance has its own Markov switching dynamics, then it should be modeled separately. Though testing for switches in the variance might not be of first-order interest in the context of our proposed rationality tests, our framework allows for it nonetheless.

In general, Markov switching models are mixture models and, therefore, the misspecification of the likelihood can lead to size distortions when using likelihood-based tests. Misspecification is less prevalent in the context of rationality tests than when comparing forecasts since forecast error distributions are often reasonably well approximated by a Normal distribution.

5. A MARKOV SWITCHING BIAS IN THE FFR FORECASTS

This section investigates the forecast unbiasedness of the BCFF survey's predictions for the FFR. Significant deviations from forecast unbiasedness by survey participants are important since a state-dependent bias in the interest rate expectations implies that it might be possible to improve prediction in specific periods in time and policymakers such as the central banks can help in the process by improving the communication strategies.

Previous work that found state dependence in forecast errors includes [Joutz and Stekler \(2000\)](#), [Sinclair et al. \(2010\)](#), and [Granziera et al. \(2021\)](#) for various forecasts of the Federal Reserve and the ECB. Studies of forecast rationality of private-sector survey predictions include [Croushore \(2012\)](#) and [Rossi and Sekhposyan \(2016\)](#), who investigate the forecast rationality of US Survey of Professional Forecasters' predictions; the latter find that forecast rationality is time-varying and depends on the sub-sample considered. [Dahlhaus and Sekhposyan \(2018\)](#) consider the BCFF predictions of the FFR, and test forecast rationality in sub-samples, conditional on whether the economy is in a monetary easing or tightening regime; in their work, the regime is observable and measured by lagged interest rate decreases and increases. Our empirical analysis contributes to this literature by revealing state dependence in the BCFF, without having to restrict our consideration to a specific state variable *ex-ante*. Additionally, we show that the state-dependent bias extends to the forecasts implied by the Federal Funds Futures (FFF) markets, i.e., it is not an idiosyncratic feature of the BCFF survey.

The BCFF is conducted monthly and consists of approximately 50 participants in the private financial sector. We focus on the consensus forecast, which is the cross-sectional average of all participants. The predictions are fixed-event forecasts, and we follow [Dahlhaus and Sekhposyan \(2018\)](#) (see also [Chun, 2011](#)) to convert the survey predictions to fixed-horizon forecasts.

In total, the survey data ranges from 1983:M4, the start of the survey, to 2018:M2. In the analysis, we focus on the period starting in 1990:M1 for two reasons. First, the data in the 1980s is quite volatile and contains several outliers. Second, an increase in the Fed's transparency in monetary policy communication at the beginning of the 1990s ([Woodford, 2005](#)) gives rise to potentially confounding structural changes in the forecast error dynamics relative to the earlier period. Lastly, the effective sample size depends on the forecasting horizon.

Let FFR_{t+h} denote the average of the effective FFR in month $t + h$, and let $\text{BCFF}_{t,h}$ denote the h -step prediction of the FFR provided by the BCFF at time t . Then, the forecast error is given by $\epsilon_{t,h} = \text{FFR}_{t+h} - \text{BCFF}_{t,h}$, i.e., it is the difference

of the realization and the forecast. To test for unbiasedness, we specify the following model:

$$\epsilon_{t,h} = \mu + S_{t+h}\mu_s + \sum_{i=1}^d \phi_i \epsilon_{t-i,h} + e_{t+h}, \quad (18)$$

where $S_t \in \{0,1\}$ is a stationary first-order Markov chain, and $e_t \sim_{\text{iid}} N(0, \sigma^2)$. In the following, we denote the case of $S_t = 0$ as regime one and the case of $S_t = 1$ as regime two.

In our baseline specification, we focus on the three-month-ahead forecast error, i.e., $h = 3$. Results for the six-month-ahead forecasts are very similar and are reported in the Online Appendix.

[Tables 6](#) and [7](#) display the results of the AFE-BS and AFE-H test for unbiasedness, i.e., $\mu = \mu_s = 0$ in [eq. \(18\)](#), for $d = 0, 1, 2, 3$.¹¹ For the AFE-BS test we let $(p, q) \in \Lambda_{(p,q)}$ as defined in Section 3.1. For the AFE-H test, we used a 4-tuple of 12 equally-spaced grid points for $(p, q) \in [0.04, 0.96] \times [0.04, 0.96]$, and 20 equally-spaced grid points for $(\mu, \mu_s) \in [-1, 0.2] \times [-2, 0.4]$. For all lag lengths, the AFE-BS and AFE-H test reject the null hypothesis of an unbiased forecast at a significance value below 0.01.

The coefficients p and q , displayed in [Tables 6](#) and [7](#), show the state-to-state transition probabilities of regime one and two, respectively. Across different lag length specifications, regime one is persistent, with a state-to-state transition probability of 96% to 97%, and the forecasts appear to be unbiased as $\mu \approx 0$; a subsequent t-test on μ does not reject the null hypothesis of $\mu = 0$. However, in the second regime, which is considerably less persistent (in [Table 6](#)) when controlling for lags of the forecast error, the forecasters overestimate the future FFR, as the coefficient $\mu + \mu_s$ is large, negative, and significantly different from zero. The forecast bias in absolute terms, i.e., $|\mu + \mu_s|$, is estimated to be around 18 to 50 basis points. Note that while the results are not identical across [Tables 6](#) and [7](#), they have the same implications.¹²

Table 6. AFE-BS Test Results – Three-Month-Ahead Forecast Error.

Model	\hat{p}	\hat{q}	$\hat{\mu}$	$\hat{\mu} + \hat{\mu}_s$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	LR-value	<i>p</i> value
AR(0)	0.98 (0.01)	0.83 (0.06)	0.01 (0.01)	-0.78 (0.02)	–	–	–	146.06	< 0.01
AR(1)	0.97 (0.01)	0.62 (0.11)	0.01 (0.01)	-0.53 (0.02)	0.67 (0.01)	–	–	77.17	< 0.01
AR(2)	0.97 (0.01)	0.60 (0.12)	0.00 (0.01)	-0.50 (0.02)	0.84 (0.03)	-0.17 (0.03)	–	63.55	< 0.01
AR(3)	0.97 (0.01)	0.60 (0.11)	0.00 (0.01)	-0.50 (0.02)	0.81 (0.03)	-0.09 (0.04)	-0.07 (0.03)	64.87	< 0.01

Notes: The sample size is $T = 338$. The displayed coefficients correspond to the maximum obtained under the alternative, using the restriction that $(p, q) \in \Lambda_{(p,q)}$. The column labeled “LR-value” denotes the value of the likelihood-ratio. Numbers in parentheses denote robust standard errors. The column “*p* value” denotes the *p*-value obtained using the approximated asymptotic distribution based on 200 bootstrap replications. \hat{p} denotes the state-to-state transition probability for regime one and \hat{q} denotes the state-to-state transition probability for regime two.

Table 7. AFE-H Test Results – Three-month-Ahead Forecast Error.

Model	\hat{p}	\hat{q}	$\hat{\mu}$	$\hat{\mu} + \hat{\mu}_s$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	AFE-H	p value
AR(0)	0.96 (0.04)	0.92 (0.02)	-0.94 (0.02)	0.20 (0.03)	-	-	-	11.31	< 0.01
AR(1)	0.96 (0.04)	0.88 (0.02)	-0.31 (0.01)	0.14 (0.03)	0.74 (0.03)	-	-	8.49	< 0.01
AR(2)	0.96 (0.08)	0.88 (0.02)	0.01 (0.03)	-0.18 (0.01)	1.03 (0.04)	-0.31 (0.04)	-	7.74	< 0.01
AR(3)	0.96 (0.08)	0.88 (0.02)	0.01 (0.03)	-0.18 (0.01)	1.02 (0.04)	-0.26 (0.05)	-0.05 (0.03)	8.25	< 0.01

Notes: The sample size is $T = 338$. The displayed coefficients correspond to the coefficients obtained when maximizing the likelihood over the finite grid of (p, q, μ, μ_s) of the AFE-H statistic. Numbers in parentheses denote robust standard errors. “AFE-H” denotes the value of the test statistic. The column “p value” denotes the p-value obtained from the simulated asymptotic distribution. The results for AFE-H are based on a 4-tuple of 12 equally-spaced grid points for $(p, q) \in [0.04, 0.96]$ and 20 equally-spaced grid points for $\mu, \mu_s \in [-1, 0.2] \times [-2, 0.4]$. \hat{p} denotes the state-to-state transition probability for regime one and \hat{q} denotes the state-to-state transition probability for regime two.

Fig. 2 plots the smoothed regime probabilities (solid lines) of the AR(3) model of Table 6 against the forecast error and the FFR. The left y-axis denotes the scale of the regime probability, whereas the right y-axis denotes the scale of the forecast error and the FFR. The dashed line displays the forecast error (rescaled by a

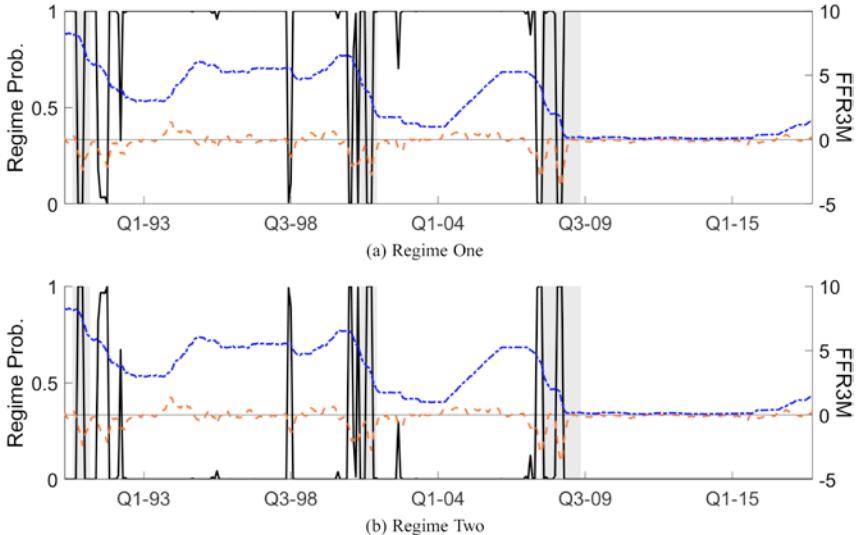


Fig. 2. Regime Probabilities. Notes: The left y-axis denotes the regime probability. The right y-axis denotes the value of the forecast error and the FFR. The solid line displays the smoothed regime probabilities of the Markov switching model with three lags, defined in eq. (18). The dashed line displays the forecast error. We rescaled the forecast error by a factor of two, to increase the legibility of the plot. The dash-dotted line displays the FFR and grey shaded areas display NBER recession periods.

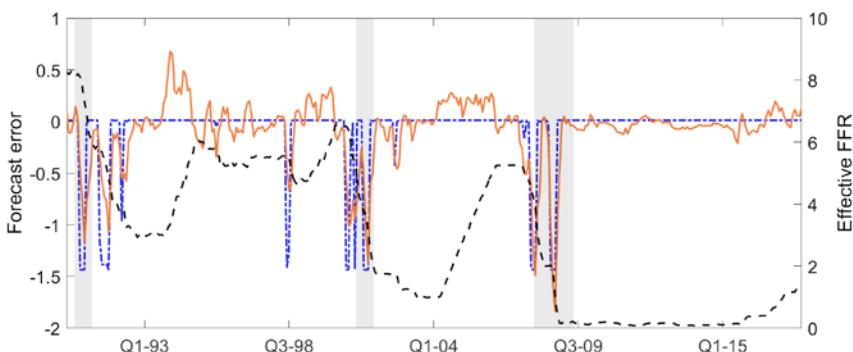
factor of two to increase the legibility of the plot). The dash-dotted line displays the FFR, while grey shaded areas display NBER recession periods. An increase of the probability of regime two is associated with monetary easing, but is not limited to recessionary periods. In particular, in the early 1990s, around 1998, and before the Great Recession in 2007–2009, the probability co-moves with changes in the FFR although the economy was not in a recession according to the NBER. Overall, the regimes appear to be well identified, in the sense that most regime probabilities are close to zero or one.

[Fig. 3](#) plots the forecast error, $\epsilon_{t,3}$, against the time-varying unconditional mean of the AR(3) model of [Table 6](#), given by $\hat{\mu}(1-\hat{\phi}_1-\hat{\phi}_2-\hat{\phi}_3)^{-1} + \hat{S}_{t+3}\hat{\mu}_s(1-\hat{\phi}_1-\hat{\phi}_2-\hat{\phi}_3)^{-1}$, using the smoothed state probabilities for \hat{S}_{t+3} . The figure shows that the switches in the unconditional mean alone can account for much of the recurring negative realizations of the forecast error.

In comparison to [eq. \(18\)](#), [West and McCracken \(1998\)](#) full-sample test for unbiasedness considers the null hypothesis $H_0 : \mu = 0$ in the model $\epsilon_{t,h} = \mu + e_{t+h}$, where e_t is a zero mean error term. Applying [West and McCracken \(1998\)](#) to the three-month-ahead forecast error does not reject the null of $\mu = 0$; the p-value is around 0.6.¹³

The non-parametric Fluctuation test by [Rossi and Sekhposyan \(2016\)](#) rejects the null hypothesis of unbiasedness at the 5% level with the rolling window size m chosen to be at the one third of the total out-of-sample period. However, Markov switching model results can identify the potential states driving the bias, making our results more useful when trying to bias-correct forecasts or make policy decisions in particular states of the world.

Our empirical results are closely related to [Dahlhaus and Sekhposyan \(2018\)](#). The authors evaluate forecast unbiasedness of FFR forecast errors of the BCFF and find that the bias seems to be mainly present in periods of monetary easing. However, since there is no common definition of “periods of monetary easing,”



[Fig. 3](#). Time-Varying Mean. *Notes:* The solid line (left-hand side y-axis) displays the forecast error. The dash-dotted line (left-hand side y-axis) displays the time-varying unconditional mean of the AR(3) model of [Table 6](#), using the smoothed state probabilities for \hat{S}_{t+3} . The dashed line (right-hand side y-axis) shows the FFR level and grey shaded areas display NBER recession periods.

the authors first have to define a state variable to identify their subsamples. In contrast, although the Markov switching approach proposed here finds similar periods of a negative forecast bias, it does so without having to define the state variable *ex-ante*. Instead, the periods are identified via the latent state of the regime-switching model.

Regime switching bias in market-based forecast errors: So far, our analysis focused on the BCFF forecasts, which are collected from forecasters working in the financial sector. If their forecasts are indeed representative of what major financial institutions expect, then their forecasts could be correlated with the futures market's expectation of the FFR. Thus, we might expect similar regime switches and deviations for forecast unbiasedness in FFF as well.

To investigate whether this is the case, we constructed three- and six-month-ahead monthly forecast errors using prices of FFF for the period of January 1995 to February 2018.¹⁴ FFF settle on the average effective FFR of the respective h -step-ahead target month and, therefore, provide a benchmark market-based measure of FFR expectations. We compute the h -step-ahead forecast error as the average effective FFR in month $t + h$ minus the FFF settlement price of the last trading day of month t . For instance, the March 31, 2006, settlement price of the three-month-ahead FFF is evaluated against the average effective FFR of June 2006.¹⁵ Fig. A.1 plots the forecast error based on the BCFF prediction (labeled FE-BCFF) against the forecast error based on the FFF prices (labeled FE-FFF) for the three-month-ahead periods. Note that the forecast errors of the BCFF and the FFF are very similar, suggesting that the information set of the BCFF panelists and the agents in futures market are indeed very similar. In the Online Appendix, we show that this is also the case for the six-month-ahead forecast errors.

Tables A.1 and A.2 in the Appendix show the results for testing for a Markov switching bias in the FFF implied forecast error. Results are very similar to the BCFF results displayed in Tables 6 and 7, respectively. In fact, Fig. A.2 shows that the respective regime probabilities estimated on the three-month-ahead forecast error produced by the FFF (solid line, RP-FFF) and by the BCFF (dashed line, RP-BCFF) are very similar.

Regime switches and forecast disagreement: We also investigate whether the bias is related to the panelists' disagreement about the future FFR. In fact, if the forecast error biases are systematically correlated with disagreement, then point forecasts may reflect a shift in the marginal forecaster between "hawks," who always over-predict the interest rate, and "doves," who always under-predict the interest rate. To that end, we computed the difference between the top-10-average and bottom-10-average forecasts of the panelists as in Andrade et al. (2016) and Dahlhaus and Sekhposyan (2018). Fig. 4 shows plots the disagreement of the forecasters against the regime probabilities. The figure suggests that while sometimes the disagreement and the regime probabilities co-move, there is no systematic correlation between disagreement and point forecast errors.

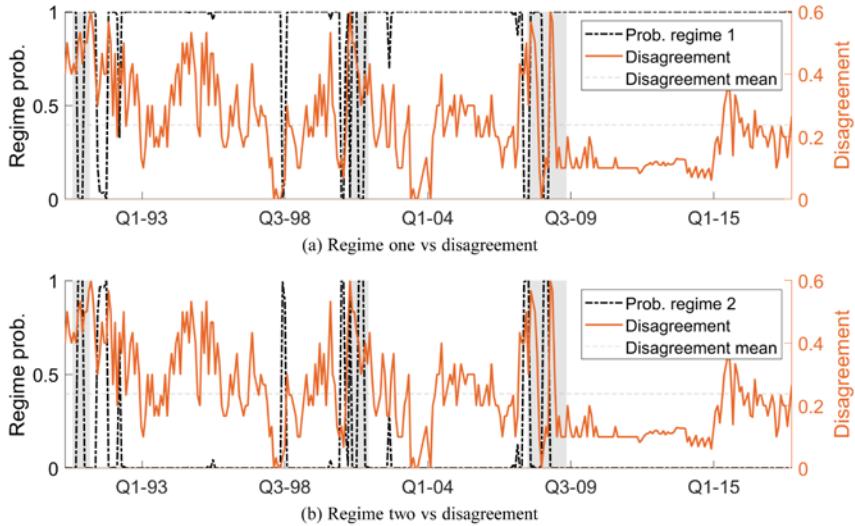


Fig. 4. Three-Month-Ahead Forecast Error: Regime Probabilities Versus Disagreement. *Notes:* The left y -axis shows the scale of the smoothed regime probability, estimated using the model defined in eq. (18), with three lags, on the BCFF FFR forecast error. The right y -axis shows the disagreement between BCFF panelists. Grey shaded areas display NBER recession periods.

Regime switches and monetary policy uncertainty: Moreover, we analyze whether the bias of BCFF panelists' FFR forecast is more generally related to uncertainty about monetary policy. Our measure of monetary policy uncertainty is the “MPU” index of Baker et al. (2016), which is constructed using newspaper articles of the 10 major US newspapers. Fig. 5 plots the estimated regime probabilities (solid line, left y -axis) against the MPU index (dashed line, right y -axis). Spikes in the MPU index before the zero lower bound (ZLB) period tend to coincide with an increase in the probability of the second regime, notably around the two US recessions in our sample as well as in 1998 around the Fed intervention triggered by the collapse of Long Term Capital Management.

Forecast errors of real output growth and inflation rates: The BCFF panelists also provide forecasts for US real gross domestic product (GDP) and the GDP deflator growth (from hereon referred to as inflation). To investigate whether the bias in the FFR forecast is associated with a bias in the corresponding macroeconomic forecasts, we computed the average forecast error of real GDP growth and inflation conditional on the regimes estimated on the BCFF FFR forecast errors, denoted by $\hat{S}_{t+h}^{\text{FFR}}$. Then, we estimate the following regression for the forecast error of real GDP growth and inflation:

$$\epsilon_{t,h} = \mu + \mu_s \hat{S}_{t+h}^{\text{FFR}} + \sum_{i=1}^3 \phi_i \epsilon_{t-i,h} + e_{t+h}, \quad (19)$$

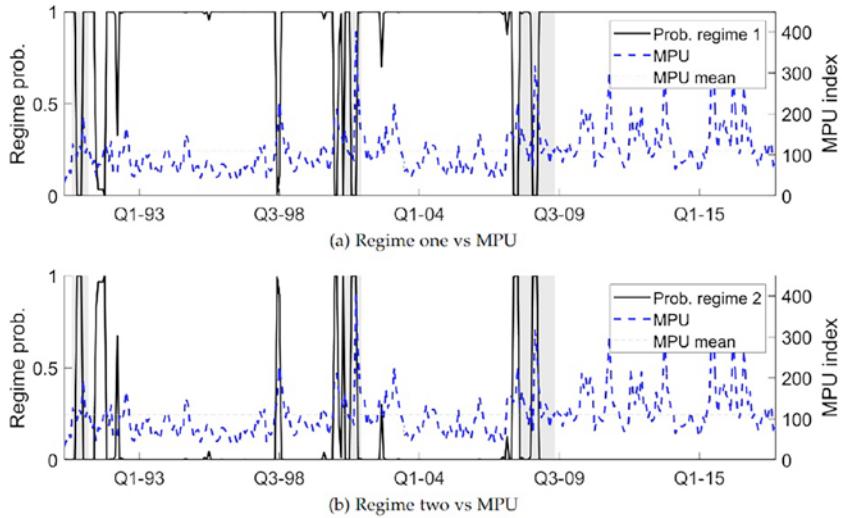


Fig. 5 Three-Month-Ahead Forecast Error: Regime Probabilities Versus MPU.

Note: The left y -axis shows the scale of the smoothed regime probability, estimated using the model defined in eq. (18), with three lags, on the BCFF FFR forecast error. The right y -axis shows the MPU index. Grey shaded areas display NBER recession periods.

where $e_{t+h} \sim_{\text{iid}} N(0, \sigma_e^2)$. Results are reported in Table 8, which shows the estimated coefficients, $\hat{\mu}$ and $\hat{\mu} + \hat{\mu}_s$, and the p -value of a t -test of $\mu = 0$ and an F-test on $\mu + \mu_s = 0$. The point estimates of $\hat{\mu} + \hat{\mu}_s$ are negative for both the real GDP growth and inflation forecast errors, i.e., periods of overestimation of the FFR coincide with periods of overestimation of real GDP growth and inflation. Note, however, that we cannot reject the null hypothesis of $\mu + \mu_s = 0$ at conventional significance levels and that part of these results are driven by the large negative forecast error during the Great Recession of 2008 to 2009. For the GDP deflator-based inflation we find that $\hat{\mu}$ is also negative and significantly different from zero, pointing to a potential constant bias in the forecast error.

Additional robustness analyses: In the Online Appendix, we show that our results are robust to both the exclusion of the ZLB period after the Great Recession and the use of the six-month-ahead (instead of the three-month-ahead) forecast errors.

Table 8. Results for Output Growth and Inflation.

	GDP Growth		Inflation	
	$\hat{\mu}$	$\hat{\mu} + \hat{\mu}_s$	$\hat{\mu}$	$\hat{\mu} + \hat{\mu}_s$
Parameter values	0.030 (0.316)	-0.426 (0.117)	-0.112 (0.002)	-0.375 (0.122)

Notes: The sample size is $T = 338$. Values in parenthesis denote p -values based on the HAC variance estimator of Newey and West (1987) with a bandwidth of $T^{1/4}$.

Besides addressing the question of what causes the bias, the results also have potential implications for monetary policy communication. While prior to the 1990s many in the Fed believed that policy effectiveness depended on surprising the market (Poole, 2005), the current consensus is rather along the opposite lines: it is a central bank's job to transparently manage expectations (see Woodford, 2005 for a discussion). In the words of Goodfriend (1991): "By making itself more predictable to the markets, the central bank makes market reactions to monetary policy more predictable to itself. And that makes it possible to do a better job of managing the economy." From that perspective, a systematic over-prediction of the policy rate during monetary easings suggests that there is room for improvements in the Fed's communication strategy.

6. CONCLUSION

Despite ample evidence on state dependence in prediction errors, existing forecast rationality tests either rely on non-parametric techniques to account for the time-variation or treat the states as observable when thinking of forecast optimality. We propose a framework for forecast evaluation that is able to detect state-dependent deviations from forecast rationality where the states are unknown *a priori*. Overall, our tests exhibit good size and power properties in Monte Carlo simulations, although they somewhat underreject when testing forecast efficiency. We show that, in the presence of Markov switching, the new tests outperform available alternatives, which in general have weak power when the time-variation takes the form of regime-switching.

While we focus on a two-state Markov switching structure, we expect our results to generalize to n-states. We leave this analysis to future work due to the fact that, in practice, Markov switching models are most commonly estimated in a two-state environment and that the computational costs make implementation in the presence of more than two states difficult in practice.

In an empirical investigation of the forecast unbiasedness of the BCFF survey, for the sample period from 1990 to 2018, our results show that the predictions exhibit a Markov switching bias when forecasting the three- and six-month-ahead FFR. While we find no evidence in favor of a constant deviation from unbiasedness in the full sample, we do provide evidence that participants tend to systematically overestimate the FFR in monetary easing episodes. We show that a similar state-dependent bias is also present in market-based forecasts of interest rates, but not in the forecasts of real GDP growth and GDP deflator-based inflation.

NOTES

1. See Section 5 for more details.
2. In a two-state model, the expected duration of regime j is given by $1/(1 - p_{jj})$, where p_{jj} is the state-to-state transition probability of regime j . Therefore, for instance, the expected duration of a state with a state-to-state transition probability as high as 90% is only 10 periods.

3. The constant variance assumption could be relaxed by allowing for conditional heteroskedasticity. More specifically, the case where the variance follows a Markov switching process is discussed in Section 4.3.

4. To reduce the computational costs of the estimation procedure under the alternative, we proceed as follows. In a first step, we maximize the log-likelihood of the model under the alternative without taking into account the restrictions on (p, q) embedded in $\Lambda_{(p,q)}$. If the obtained maximum implies values for (p, q) outside of the feasible set $\Lambda_{(p,q)}$, we resort to estimating the model over a 2-tuple of 10 equally-spaced grid values for (p, q) in $[0.02, 0.98]$; otherwise, we proceed with the maximum obtained in the first step.

5. We assume normal errors to illustrate the bootstrap procedure, since a normality assumption is the leading case for Markov switching applications.

6. The serial correlation is restricted to near-epoch dependence.

7. In addition, in applications of the AFE-H, the researcher can easily simulate the asymptotic distribution for different values of M to gauge the impact of the serial correlation on the critical values. Our results were not sensitive to the choice of M , which is in line with the findings of [Hansen \(1996\)](#).

8. Note that the size distortions in small samples for the Fluctuation test come from the fact that although the true DGP is an AR(1), we use a HAC estimator of [Newey and West \(1987\)](#) with a bandwidth equal to $T^{1/4}$ to control for the autocorrelation.

9. Markov switching tests are generally not robust to misspecification under the null hypothesis. In unreported results, we found that for a more severe misspecification, i.e. large values of the MA(1) coefficient in Case 3, both the AFE-BS and AFE-H show size distortions.

10. Again, these parameter choices ensure that we can compute the power against the alternative of Markov switching efficiency only.

11. A rational forecast would exhibit maximum serial correlation length of $h-1$, i.e., in this case two. We show results for a maximum of three lags to be robust against rejections of the null hypothesis due to other type of misspecifications.

12. Remember that AFE-H is estimated over a finite grid for both (p, q) and μ, μ_s , which may lead to a slightly different maximum than the estimation that is not based on a finite grid.

13. These results hold when additionally controlling for lags of the forecast error and using HAC standard errors of [Newey and West \(1987\)](#) with a bandwidth $T^{1/4}$.

14. We start our sample in 1995 due to data availability.

15. Prices on non-trading days are substituted by the respective price of the most recent previous trading day ([Swanson, 2006](#)).

ACKNOWLEDGMENTS

We thank the editor Yoosoon Chang and the three anonymous referees for their constructive comments and suggestions. Part of this research was carried out while Tatevik Sekhposyan was a Visiting Fellow at the Federal Reserve Bank of San Francisco, whose hospitality is gratefully acknowledged. The views expressed are those of the authors and do not necessarily reflect the views of the Banco de España, the Eurosystem, the Federal Reserve Bank of San Francisco or the Board of Governors of the Federal Reserve System. Barbara Rossi acknowledges Financial support from the Spanish Ministry of the Economy and Competitiveness and from the Spanish Agencia Estatal de Investigación (AEI), through the Severo Ochoa Programme for Centres of Excellence in R&D (Barcelona School of Economics CEX2019-000915-S) as well as the Spanish Ministry of Science and Innovation under grant PID2019-107352GB-100.

REFERENCES

- Andrade, P., Crump, R. K., Eusepi, S., & Moench, E. (2016). Fundamental disagreement. *Journal of Monetary Economics*, 83, 106–128.
- Baker, S., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131, 1593–1636.
- Bullard, J. (2016). *The St. Louis Fed's new characterization of the outlook for the U.S. Economy*. Commentary, Federal Reserve Bank of St. Louis.
- Carrasco, M., Hu, L., & Ploberger, W. (2014). Optimal test for Markov switching parameters. *Econometrica*, 82, 765–784.
- Carter, A., & Steigerwald, D. (2012). Testing for regime switching. *Econometrica*, 80, 1809–1812.
- Chang, Y., Choi, Y., & Park, J. Y. (2017). A new approach to model regime switching. *Journal of Econometrics*, 196, 127–143.
- Cho, J. S., & White, H. (2007). Testing for regime switching. *Econometrica*, 75, 1671–1720.
- Chun, A. L. (2011). Expectations, bond yields, and monetary policy. *Review of Financial Studies*, 24, 208–247.
- Croushore, D. (2012). *Forecast bias in two dimensions* [Federal Reserve Bank of Philadelphia Working Paper 12-9, Federal Reserve Bank of Philadelphia].
- Dahlhaus, T., & Sekhposyan, T. (2018). *Monetary policy uncertainty: A tale of two tails* [Staff Working Paper 2018-50], Bank of Canada. <https://www.bankofcanada.ca/2018/09/staff-working-paper-2018-50/>
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64, 247–254.
- Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74, 33–43.
- Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review*, 39, 763–788.
- Goodfriend, M. (1991). Interest rates and the conduct of monetary policy. *Carnegie-Rochester Conference on Public Policy*, 34, 7–30.
- Granziera, E., Jalasjoki, P., & Paloviita, M. (2021). *The bias and efficiency of the ECB inflation projections: A state dependent analysis* [Norges Bank Working Paper 1/2021. Norges Bank].
- Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45, 39–70.
- Hansen, B. E. (1992). The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 7, 61–82.
- Hansen, B. E. (1996). Erratum: The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics*, 11, 195–198.
- Joutz, F., & Stekler, H. (2000). An evaluation of the predictions of the Federal Reserve. *International Journal of Forecasting*, 16, 17–38.
- Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In J. A. Mincer (Ed.), *Economic forecasts and expectations: Analysis of forecasting behavior and performance* (pp. 3–46). New York: National Bureau of Economic Research.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–708.
- Poole, W. (2005). *After greenspan: Whither fed policy? Speech 1*. Federal Reserve Bank of St. Louis.
- Qu, Z., & Zhuo, F. (2021). Likelihood ratio based tests for Markov regime switching. *The Review of Economic Studies*, 88, 937–968.
- Rossi, B. (2013). Advances in forecasting under model instability. In G. Elliot & A. Tmmermann, (Eds.), *Handbook of economic forecasting* (Vol. 2b, Chap. 21, pp. 1203–1324). Elsevier Publications.
- Rossi, B., & Sekhposyan, T. (2016). Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. *Journal of Applied Econometrics*, 31, 507–532.
- Sinclair, T. M., Joutz, F., & Stekler, H. (2010). Can the fed predict the state of the economy? *Economic Letters*, 108, 28–32.

- Swanson, E. (2006). Have increases in Federal Reserve transparency improved private sector interest rate forecasts? *Journal of Money, Credit, and Banking*, 38, 791–819.
- West, K. D., & McCracken, M. W. (1998). Regression-based tests of predictive ability. *International Economic Review*, 39, 817–840.
- Woodford, M. (2005). Central Bank communication and policy effectiveness. In J. Hole (Ed.), *Proceedings - Economic Policy Symposium* (pp. 399–474). Federal Reserve Bank of Kansas City.

APPENDIX. A EMPIRICAL RESULTS USING FEDERAL FUNDS FUTURES

Table A.1. AFE-BS Test Results — Three-Month-Ahead Market Forecast Error.

Model	\hat{p}	\hat{q}	$\hat{\mu}$	$\hat{\mu} + \hat{\mu}_s$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	LR-value	p value
AR(0)	0.98 (0.01)	0.98 (0.01)	-0.75 (0.02)	0.01 (0.01)	-	-	-	112.54	< 0.01
AR(1)	0.98 (0.01)	0.51 (0.17)	0.01 (0.01)	-0.63 (0.02)	0.59 (0.01)	-	-	80.51	< 0.01
AR(2)	0.97 (0.01)	0.51 (0.16)	0.00 (0.01)	-0.59 (0.02)	0.73 (0.03)	-0.16 (0.03)	-	73.69	< 0.01
AR(3)	0.97 (0.01)	0.52 (0.15)	0.00 (0.01)	-0.59 (0.02)	0.71 (0.03)	-0.12 (0.04)	-0.04 (0.03)	72.80	< 0.01

Notes: The sample size is $T = 277$. The displayed coefficients correspond to the maximum obtained under the alternative, using the restriction that $(p, q) \in \Lambda_{(p,q)}$. The column labeled “LR-value” denotes the value of the likelihood-ratio. Numbers in parentheses denote robust standard errors. The column “p value” denotes the p-value obtained using the approximated asymptotic distribution based on 200 bootstrap replications. \hat{p} denotes the state-to-state transition probability for regime one and \hat{q} denotes the state-to-state transition probability for regime two.

Table A.2. AFE-H Test Results – Three-Month-Ahead Market Forecast Error.

Model	\hat{p}	\hat{q}	$\hat{\mu}$	$\hat{\mu} + \hat{\mu}_s$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	AFE-H	p value
AR(0)	0.92 (0.96)	0.88 (0.75)	-0.05 (0.21)	0.01 (0.10)	-	-	-	11.01	< 0.01
AR(1)	0.96 (1.61)	0.88 (0.82)	-0.31 (0.80)	-0.49 (0.25)	0.54 (0.27)	-	-	9.11	< 0.01
AR(2)	0.96 (0.40)	0.88 (0.28)	-0.05 (0.13)	0.01 (0.09)	1.03 (0.06)	-0.33 (0.04)	-	9.10	< 0.01
AR(3)	0.96 (0.07)	0.88 (0.02)	0.01 (0.02)	-0.68 (0.01)	0.66 (0.03)	-0.09 (0.05)	-0.04 (0.03)	9.22	< 0.01

Notes: The sample size is $T = 277$. The displayed coefficients correspond to the coefficients obtained when maximizing the likelihood over the finite grid of (p, q, μ, μ_s) of the AFE-H statistic. Numbers in parentheses denote robust standard errors. “AFE-H” denotes the value of the test statistic. The column “p value” denotes the p-value obtained from the simulated asymptotic distribution. The results for AFE-H are based on a 4-tuple of 12 equally-spaced grid points for $(p, q) \in [0.04, 0.96]$ and 20 equally-spaced grid points for $\mu, \mu_s \in [-1, 0.2] \times [-2, 0.4]$. \hat{p} denotes the state-to-state transition probability for regime one and \hat{q} denotes the state-to-state transition probability for regime two.

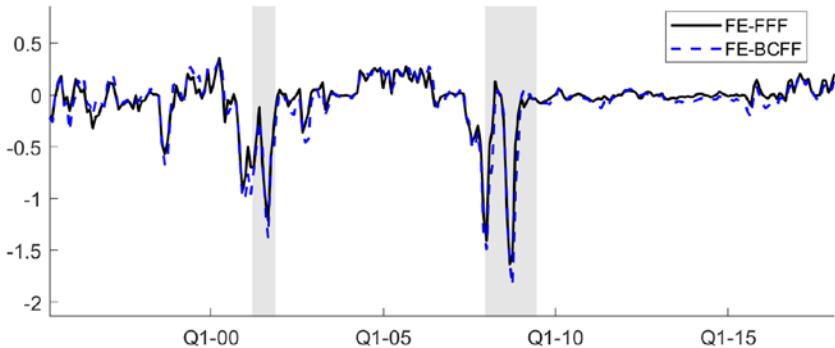


Fig. A.1. Three-Month-Ahead Forecast Errors Based on Market and Survey Predictions. *Notes:* The forecast errors implied by the Federal Funds Futures are, denoted by “FE-FFF,” whereas the forecast are errors of the BCFF survey forecasts, are denoted by “FE-BCFF.” Grey shaded areas display NBER recession periods.

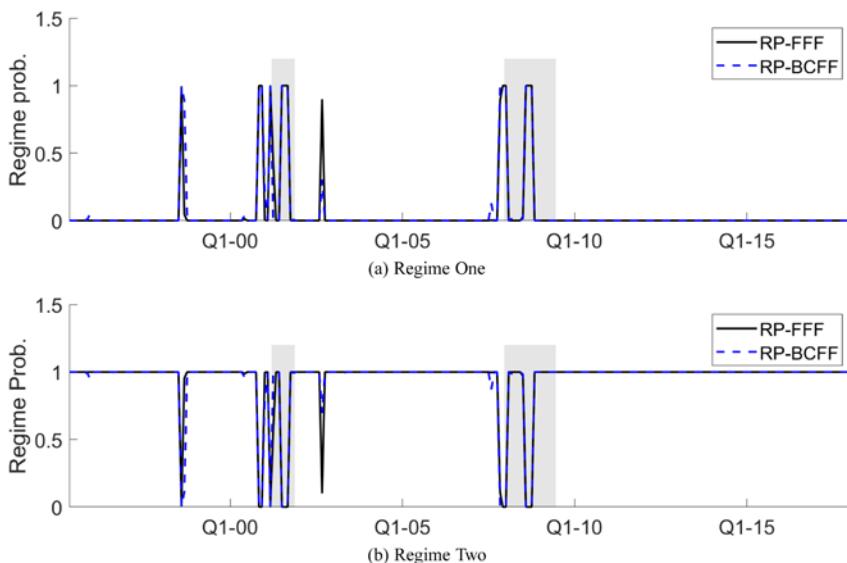


Fig. A.2. Regime Probabilities Implied by Three-Month-Ahead Market and Survey Predictions. *Notes:* The smoothed regime probabilities estimated on the forecast errors implied by the Federal Funds Futures are, denoted by “RP-FFF,” whereas the smoothed regime probabilities estimated on the forecast errors implied by the BCFF are denoted by “RP-BCFF”. In both cases, we obtained the smoothed regime probabilities from the Markov switching model, defined in eq. (18), with three lags. Grey shaded areas display NBER recession periods.

CHAPTER 3

THE ECONOMETRICS OF OIL MARKET VAR MODELS

Lutz Kilian and Xiaoqing Zhou

Federal Reserve Bank of Dallas, United States

ABSTRACT

Oil market VAR models have become the standard tool for understanding the evolution of the real price of oil and its impact on the macro economy. As this literature has expanded at a rapid pace, it has become increasingly difficult for mainstream economists to understand the differences between alternative oil market models, let alone the basis for the sometimes divergent conclusions reached in the literature. The purpose of this survey is to provide a guide to this literature. Our focus is on the econometric foundations of the analysis of oil market models with special attention to the identifying assumptions and methods of inference.

Keywords: Identification; model specification; elasticity; Bayesian estimation; structural VAR; instruments; textual analysis

JEL Codes: Q43; Q41; C36; C52

1. INTRODUCTION

In the last decade, structural vector autoregressive (VAR) models of the global oil market have become the standard tool for understanding the evolution of the real price of oil and its effect on the macro economy. These models have helped create a consensus among researchers and policymakers that shifts in the global demand for oil are the primary determinant of the real price of oil. As the oil market literature

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 65–95

Copyright © 2023 by Lutz Kilian and Xiaoqing Zhou

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:[10.1108/S0731-90532023000045B003](https://doi.org/10.1108/S0731-90532023000045B003)

has grown at a rapid pace, there has been a proliferation of alternative approaches to modeling the real price of oil, some refining existing models of the global oil market and others challenging this consensus. As a result, it has become increasingly difficult for mainstream economists to understand the differences between alternative oil market models, let alone the basis for the sometimes divergent conclusions reached in the literature. The purpose of this survey is to provide a guide to this literature.¹

Our focus is on the econometric foundations of the analysis of oil market models with special attention to the identifying assumptions and methods of inference. We not only explain how the workhorse models in this literature have evolved, but also examine alternative oil market VAR models. Our review is intended to help the reader understand why the latter models sometimes have generated unconventional, puzzling or erroneous conclusions. Our analysis is of interest not only to applied researchers interested in modeling oil markets and their relationship with the domestic economy, but also to applied econometricians interested in structural VAR modeling more generally.

The remainder of the chapter is organized as follows. In Section 2, we trace the evolution of the oil market literature from its origins to the workhorse model of Kilian and Murphy (2014). We also discuss extensions to larger-dimensional models. Section 3 reviews some commonly used methods of estimating oil market VAR models and the question of how to conduct inference about the structural impulse responses and similar statistics. It also discusses the arguments for and against time-varying coefficient models. Section 4 examines some recent controversies about the specification and estimation of oil market VAR models. Section 5 discusses extraneous measures of oil demand and supply shocks. The concluding remarks are in Section 6.

2. CONVENTIONAL IDENTIFICATION STRATEGIES IN OIL MARKET MODELS

VAR models of the global oil market evolved in three main stages, starting with the work of Kilian (2006, 2008a, 2009) which exploited short-run exclusion restrictions. This approach was followed by the introduction of static and dynamic inequality restrictions as an alternative to short-run exclusion restrictions, as exemplified by Lippi and Nobili (2012), Kilian and Murphy (2012) and Inoue and Kilian (2013). Finally, Kilian and Murphy (2014) extended this framework to allow oil price expectations to have an effect on the real price of oil through shifts in storage demand. This last extension is crucial because it addresses concerns about previous oil market models being informationally deficient. It also provides a direct link to the literature on modeling oil futures markets (see Alquist & Kilian, 2010; Knittel & Pindyck, 2016).² In this section, we review these benchmark models and their extensions.

2.1. Short-Run Exclusion Restrictions

The importance of disentangling demand and supply shocks in oil markets was first pointed out in Barsky and Kilian (2002). A quantitative framework for this

task was provided in [Kilian \(2008a, 2009\)](#) who proposed a monthly structural VAR model of the global oil market since 1973 that includes the percent change in global oil production ($\Delta prod$), an index of cyclical variation in global real economic activity derived from bulk dry cargo ocean freight rates (rea), and the log of the real price of oil ($rpoil$). Variation in these data is explained based on three shocks: (1) a shock to the amount of oil pumped out of the ground (“oil supply shock”), (2) a shock to the demand for all industrial commodities including crude oil (“aggregate demand shock”), and (3) a residual oil demand shock (“oil-market specific demand shock”) designed to capture precautionary oil demand shocks (see [Alquist & Kilian, 2010](#)). The latter shock may also be interpreted as a preference shock. For example, an increased preference for smaller, more fuel-efficient automobiles would result in lower demand for oil, given the same level of global real activity. Thus, there are two oil demand shocks and one oil supply shock in this model.

The identifying restrictions are imposed on the structural impact multiplier matrix, B_0^{-1} with ij th element b_{ij}^0 , that links the vector of reduced-form errors, u_t , to the vector of structural shocks, w_t :

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{pmatrix} b_{11}^0 & 0 & 0 \\ b_{21}^0 & b_{22}^0 & 0 \\ b_{31}^0 & b_{32}^0 & b_{33}^0 \end{pmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix} \quad (1)$$

The real price of oil is allowed to respond instantaneously to all three structural shocks. The impact price elasticity of oil supply is assumed to be zero, which means that global oil production does not respond to either demand shock on impact ($b_{12}^0 = b_{13}^0 = 0$). The model also postulates that an oil-specific demand shock that raises the real price of oil does not lower global real activity within the same month ($b_{23}^0 = 0$).³ There are no restrictions on the responses at longer horizons.⁴

Although [Kilian \(2009\)](#) imposes a recursive ordering on the structural impact multiplier matrix, this structure is not atheoretical. Every exclusion restriction can be motivated on economic grounds. Notably, the assumption of a zero impact price elasticity of oil supply is consistent with evidence about how OPEC oil producers historically have set production levels (see [Kilian, 2009](#)). It is also consistent with the theoretical result that oil producers in response to higher prices adjust their production levels only with a delay, because adjusting the level of oil production is costly (see [Anderson et al., 2018](#)). Likewise, the restriction on the response of global real economic activity can be rigorously defended (see [Kilian & Zhou, 2018](#)).⁵

2.2. Long-Run Exclusion Restrictions

Long-run identifying restrictions are rarely used in oil market models. [Stürmer's \(2018\)](#) analysis of the oil market between 1840 and 2014 is an exception. His model relies on the same three variables as in [Kilian \(2009\)](#), except that all

variables are annual and global real GDP growth takes the place of the global real activity index. Since the data are annual, the short-run identifying restrictions employed by Kilian (2009) are not economically plausible. Stürmer therefore develops a different approach to identification.

There are no restrictions on the short-run effects of the structural shocks on the model variables. Instead, it is assumed that persistent global economic expansions are associated with gains in total factor productivity and hence permanently raise the level of global real output. These expansions are also associated with sustained increases in the real price of oil because they raise the aggregate demand for all industrial commodities. Persistently high real oil prices in turn stimulate technological innovation in oil extraction and discoveries of new oil deposits. Thus, these oil demand shocks also tend to have a permanent effect on the level of oil production. In contrast, oil supply shocks are associated with strikes, industrial accidents, natural disasters, and political unrest or wars. Such shocks typically have only short-lived effects on global real output. Hence, it makes sense to model their effect on real output as transitory, while allowing oil supply shocks to have permanent effects on the production of oil.⁶ Finally, oil-market specific demand shocks such as oil inventory demand shocks do not affect either global real output or oil production in the long run. These assumptions yield three exclusion restrictions on the long run level response of global real activity and oil production that suffice to identify all three shocks.

2.3. Sign Restrictions

One limitation of the Kilian (2009) model is that it suggests that the spike in the price of oil following the invasion of Kuwait in 1990 was entirely caused by the oil-specific demand shock, which does not seem credible, given the large oil supply disruption that took place in 1990. This prompted Kilian and Murphy (2012, 2014) to explore alternative identification schemes that restrict the sign of selected impulse responses rather than imposing zero restrictions on B_0^{-1} . One advantage of sign-identified structural VAR models is that they allow the impact price elasticity of oil supply to be close to zero without requiring it to be literally zero (see Kilian & Murphy, 2012).⁷

2.3.1. The Kilian and Murphy (2012) Model

Kilian and Murphy (2012) proposed a model that is analogous to Kilian (2009) except that the identification is based on sign restrictions rather than exclusion restrictions.

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{pmatrix} - & + & + \\ - & + & - \\ + & + & + \end{pmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix} \quad (2)$$

The rationale of these sign restrictions is that an unexpected decline in the supply of oil shifts the oil supply curve to the left along the oil demand curve, causing

global oil production to fall, the real price to increase, and global real activity to fall. In contrast, a positive aggregate demand shock raises global real activity and the real price of oil and stimulates global oil production. Finally, an increase in oil-specific demand (such as an increase in oil inventory demand) causes the real price of oil to increase, while stimulating global oil production and causing a decline in global real activity.

One key difference from other early examples of sign-identified oil market models such as Peersman and van Robays (2009) and Lippi and Nobili (2012) is that Kilian and Murphy (2012) emphasized the importance of bounding the impact price elasticity of oil supply, which they defined as the ratio of the impact responses of global oil production and of the real price of oil to an oil demand shock.⁸ They observed that there is a consensus in the literature that the one-month price elasticity of oil supply is near zero. Not imposing a bound on this elasticity means that we consider models with impact price elasticities of oil supply as high as 2 as plausible as models that imply elasticities close to 0. That approach is clearly unreasonable. Kilian and Murphy (2012) proposed an upper bound of 0.0258 on the one-month price elasticity of oil supply based on historical evidence. A detailed discussion of the derivation of this bound can be found in Kilian (2022a). Kilian and Murphy concluded that it is not possible to generate large responses of the real price of oil to oil supply shocks, once the oil supply elasticity is restricted in this manner.⁹ Once a reasonable supply elasticity bound is imposed, the responses to oil supply shocks in the Kilian and Murphy (2012) model are broadly similar to those reported in Kilian (2009). As Kilian and Murphy showed, even with a supply elasticity as high as 0.08, oil supply shocks would explain only 10% of the variation in the real price of oil.

Inoue and Kilian (2013) extend the Kilian and Murphy (2012) model to include additional dynamic sign restrictions. They restrict the real price of oil to be positive for the first year in response to unanticipated oil supply disruptions and in response to positive oil demand shocks. The validity of the identifying assumptions in the models of Kilian and Murphy (2012) and Inoue and Kilian (2013) has been empirically assessed and confirmed in a number of studies including Lütkepohl and Netšunajev (2014), Herwartz and Plödt (2016) and Hafner, Herwartz and Wang (2022).

Antolin-Diaz and Rubio-Ramirez (2018) show how more precise estimates of the Kilian and Murphy (2012) model may be obtained by imposing additional narrative restrictions in Bayesian estimation. Narrative restrictions refer to restrictions on the signs or relative magnitudes of structural shocks or of historical decompositions during selected periods, for which extraneous evidence exists. For example, we know from oil industry sources that there was a surge in storage demand for oil between May and December 1979. Thus, it makes sense to impose the restriction that this feature also holds in the estimated oil market model. Similarly, we know that in August 1990 a negative oil supply shock took place when Iraq invaded Kuwait. Oil production in both countries ceased and the real price of oil rose. Any model that does not reproduce this feature clearly would not be credible. We also know that the sharp increase in the real price of oil in mid-1990 was not caused by increased flow demand, and we know that there must have been a simultaneous increase in storage demand,

raising the real price of oil further, because otherwise oil inventories would have fallen sharply in response to the oil supply disruption. These considerations provide further restrictions on the historical decomposition of the real price of oil.

A closely related frequentist approach to imposing narrative restrictions on structural shocks has been proposed by [Ludvigson et al. \(2020\)](#). Both [Antolin-Diaz and Rubio-Ramirez \(2018\)](#) and [Ludvigson et al. \(2020\)](#) apply their approach to model (2) and confirm the substantive conclusions of [Kilian and Murphy \(2012\)](#).

2.3.2. The [Kilian and Murphy \(2014\)](#) Model

The [Kilian and Murphy \(2014\)](#) model generalizes the [Kilian and Murphy \(2012\)](#) model by explicitly identifying shocks to storage demand (also referred to as speculative demand or inventory demand shocks). This extension is made possible by the inclusion of a proxy for global crude oil inventories, constructed by scaling US crude oil inventories, as reported by the EIA, by the ratio of OECD petroleum inventories over US petroleum inventories.¹⁰ Global oil inventories are best expressed in changes (Δinv) rather than growth rates.¹¹ The importance of modeling storage demand is that it allows oil price expectations to affect the real price of oil, even when oil price expectations cannot be directly observed. This fact allows the structural VAR model to capture shifts in oil price expectations not captured by the information set of the VAR model.¹² For added clarity, the oil supply and aggregate demand shocks in this model are relabeled as flow supply shocks and flow demand shocks, respectively.

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \\ u_t^{\Delta inv} \end{pmatrix} = \begin{pmatrix} - & + & + & b_{14}^0 \\ - & + & - & b_{24}^0 \\ + & + & + & b_{34}^0 \\ - & - & + & b_{44}^0 \end{pmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow demand}} \\ w_t^{\text{storage demand}} \\ w_t^{\text{other oil demand}} \end{pmatrix} \quad (3)$$

Effectively, this model decomposes the residual oil demand shock in [Kilian and Murphy \(2012\)](#) into a storage demand shock and another oil demand shock that represents, for example, shocks to inventory technology and preferences as well as idiosyncratic shocks to the US Strategic Petroleum Reserve that are not otherwise accounted for. This other oil demand shock is implicitly defined as the complement to the first three structural shocks and has no explicit structural interpretation. Negative flow supply and positive flow demand shocks are associated with declines in oil inventories, as refiners smooth the production of refined products. A positive storage demand shock, in contrast, increases oil inventories, the real price of oil and global oil production on impact, while lowering global real activity.¹³

As in [Kilian and Murphy \(2012\)](#), the one-month price elasticity of oil supply is bounded by 0.0258. The estimates are robust to relaxing this bound to 0.04 (see, e.g., [Zhou, 2020](#); [Herrera & Rangaraju, 2020](#); [Inoue & Kilian, 2022b](#)). In addition, the one-month price elasticity of oil demand is bounded by the fact that this

elasticity cannot exceed the corresponding long-run oil demand elasticity. The latter elasticity is bounded by 0.8 based on extraneous microeconomic estimates (see [Hausman & Newey, 1995](#); [Kilian & Zhou, 2022b](#)). One legacy of the analysis in [Kilian and Murphy \(2012, 2014\)](#) is that it drew attention to the importance of oil supply and oil demand elasticities for interpreting the relationship between prices and quantities in the global oil market. Indeed, much of the recent controversy about how to model oil markets evolves around this question (see, e.g., [Baumeister & Hamilton, 2019](#); [Caldara et al., 2019](#); [Kilian, 2022a](#)).¹⁴

In addition, [Kilian and Murphy \(2014\)](#) restricted the dynamic responses of global oil production and global real activity to a negative flow supply shock to be negative, and the response of the real price of oil to be positive for the first 12 months. This joint dynamic restriction can be motivated on economic grounds. Finally, [Kilian and Murphy \(2014\)](#) impose what nowadays would be referred to as narrative restrictions on the historical decomposition of the real price of oil (see [Antolin-Diaz & Rubio-Ramirez, 2018](#)). Such restrictions are especially important when conducting inference in this class of models (see, e.g., [Kilian & Zhou, 2020a, 2022a](#); [Zhou, 2020](#); [Inoue & Kilian, 2022b](#)).

In model (3), storage takes place above the ground and is driven by the demand side of the oil market. Most storage is held by refineries in oil importing countries. Although it is possible to view the stock of oil left below the ground as another form of oil inventories, the latter type of inventories is economically distinct because it cannot be used by refineries to smooth the production of refined products in the event of an unexpected shortfall of domestic oil production or oil imports. Below-ground inventories are important in their own right because oil producers expecting a higher price for future deliveries may withhold oil from the market and accumulate inventories below the ground, resulting in a reduction in flow supplies, a higher spot price, lower oil consumption and hence lower global real activity. Such a speculative supply shock, however, is observationally equivalent to any other disruption of flow supplies, say due to civil strife or a strike in the oil industry. Thus, Kilian and Murphy stress that, for all practical purposes, speculative supply shocks and flow supply shocks cannot be separately identified within existing global oil market models.¹⁵

2.3.3. Larger-Dimensional Extensions of the Kilian–Murphy Framework

As with any structural model, the validity of structural oil market VAR models hinges on the premise that the model does not omit any important determinants of the real price of oil. An obvious concern is that decomposing the structural shocks further within an extended VAR model may change the substantive conclusions obtained in the baseline model. In general, it is difficult to assess the importance of omitted structural shocks, short of extending the model to include additional variables and imposing the restrictions required to identify the additional structural shocks.

A case in point is the analysis in [Kilian and Murphy \(2014\)](#) who showed that the substantive conclusions in [Inoue and Kilian \(2013\)](#) are largely robust to including the change in global oil inventories and explicitly modeling storage

demand. They also found, however, that for specific episodes such as the 1990 oil price spike there are noticeable differences in the interpretation of the data.

By the same token, we need to ask how robust the conclusions in Kilian and Murphy (2014) are to further extensions of their model. This question could not be addressed at the time that paper was written, because the estimation period was too short to consider models with more than four variables. There are several more recent studies, however, that suggest that the substantive conclusions of Kilian and Murphy (2014) are robust to extending their model further. For example, Kilian and Zhou (2020a) confirm these findings based on a model that explicitly allows for shocks to the US Strategic Petroleum Reserve. Cross et al. (2022) reach the same conclusion when differentiating between storage demand shocks driven by changes in the expected price of oil and changes in oil price uncertainty. Finally, Kilian and Zhou (2022a) confirm the robustness of the conclusions of Kilian and Murphy (2014) to modeling exogenous variation in the US real market rate of interest and in the trade-weighted US real exchange rate.

2.3.4. Incomplete Oil Market Models

There are also examples of studies that seek to simplify the benchmark models (1), (2) and (3) in the interest of greater parsimony. Just as we compared the Kilian–Murphy framework to larger-dimensional extensions of that framework in Section 2.3.3, we need to ask whether such lower-dimensional models are capable of recovering at least approximately the estimates from the benchmark models (1), (2), and (3). If they do not, these lower-dimensional models must be considered incomplete and misleading. A case in point is the study of Bjørnland and Zhulanova (2019) who rely on a global oil market model of the form

$$\begin{pmatrix} u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_{11}^0 & 0 \\ b_{21}^0 & b_{22}^0 \end{bmatrix} \begin{pmatrix} w_t^{\text{flow demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}. \quad (4)$$

Bjørnland and Zhulanova's discussion makes it clear that they have in mind a simplified version of model (1) in which there are only two shocks, namely a flow demand shock and an oil-specific demand shock, which is identified by a delay restriction.¹⁶ They argue that the oil supply shock in model (1) may be dropped, given that Kilian (2009) showed that this shock plays only a modest role. This argument is problematic. Although oil supply shocks play only a modest role on average, for specific episodes such as the shale oil boom after 2010 their effect may be quantitatively important. Moreover, there is strong evidence for the importance of including oil inventories in oil market models and of modeling storage demand explicitly (see Kilian & Murphy, 2014; Kilian & Lee, 2014; Kilian, 2017; Herrera & Rangaraju, 2020). Thus, model (4) is mis-specified in that it conflates the structural shocks in the data generating process.

This type of problem may also arise in sign-identified models. For example, the identification of the oil market block in [Baumeister and Peersman \(2013b\)](#) boils down to the sign-identified model.

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rpoil} \end{pmatrix} = \begin{pmatrix} - & + \\ + & + \end{pmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow demand}} \end{pmatrix}, \quad (5)$$

which is easily recognized as a restricted version of the model in [Kilian and Murphy \(2014\)](#) with two of the four structural shocks suppressed. It is immediately clear that such a model conflates the distinct shocks contained in the fully specified model. More importantly, even if we restrict attention to the flow supply shock and treat this model as partially identified, there is no reason for the underspecified model (5) to be able to recover the flow supply shock in the data generating process because the information set is different.

2.3.5. Understanding the Impact of Oil Demand and Oil Supply Shocks on Domestic Macroeconomic Aggregates

Oil market VAR models are typically estimated on monthly data, because the exclusion restrictions and elasticity bounds used for identification tend to be more credible at monthly frequency than at quarterly or annual frequency. If we are interested in the effect of global oil demand and oil supply shocks on a monthly US macroeconomic aggregate such as US industrial production, a natural approach is to specify a block recursive VAR model with the domestic variable ordered last. A good example is the model of [Kilian and Park \(2009\)](#), who extended model (1) to include monthly US real stock returns under the maintained assumption that global oil market variables are predetermined with respect to the US stock market, which implies that $b_{14}^0 = b_{24}^0 = b_{34}^0 = 0$. The latter assumption is supported by evidence in [Kilian and Vega \(2011\)](#).

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \\ u_t^{ret} \end{pmatrix} = \begin{pmatrix} b_{11}^0 & 0 & 0 & 0 \\ b_{21}^0 & b_{22}^0 & 0 & 0 \\ b_{31}^0 & b_{32}^0 & b_{33}^0 & 0 \\ b_{41}^0 & b_{42}^0 & b_{43}^0 & b_{44}^0 \end{pmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \\ w_t^{\text{other shocks to stock returns}} \end{pmatrix} \quad (6)$$

Here the upper left 3×3 matrix contained in B_0^{-1} represents the oil market block and the lower right 1×1 matrix is the domestic block. Obviously, the same approach would work, if the oil market block were identified based on sign restrictions, except that the conventional approach to Bayesian inference would have to be modified, as described in [Arias et al. \(2018\)](#).¹⁷ One can include without loss of generality additional variables in the lower right block as long as one restricts attention to the responses of these variables to oil demand and oil supply shocks.¹⁸

Sometimes this block-recursive VAR approach is not feasible. One common situation is that we are interested in the effect of global oil demand and oil supply shocks recovered from monthly VAR models on quarterly or annual US macroeconomic aggregates. One approach to this problem could be to estimate a mixed-frequency VAR model (see Ghysels, 2016; Chudik & Giorgiadis, 2022). A simpler approach proposed by Kilian (2009) is to sum (or, equivalently, average) the monthly structural shocks obtained from model (1) by quarter and to run a quarterly distributed-lag model second-stage regression. For example, consider the second-stage model

$$\Delta gdp_i = \alpha_i + \beta_{0,i} w_t^i + \beta_{1,i} w_{t-1}^i + \dots + \beta_{h,i} w_{t-h}^i + v_{i,t}, \quad i = 1, 2, 3,$$

where $v_{i,t}$ denotes the possibly serially correlated and heteroskedastic regression innovation in equation i , w_t^i denotes the sum of the three realizations of the structural VAR shock i in quarter t , and $\partial gdp_{t+j} / \partial w_t^i = \partial gdp_t / \partial w_{t-j}^i = \beta_{j,i}$. Inference on the impulse responses may be conducted using HAC standard errors or the block bootstrap. This approach works because the structural shocks are mutually uncorrelated at monthly frequency and approximately mutually uncorrelated at lower frequency. Similar temporal aggregation schemes are commonly used in the literature on aggregating high-frequency monetary policy shocks. Efficiency may be gained by including current and lagged values of all shocks:

$$\begin{aligned} \Delta gdp_i = & \alpha + \beta_{0,1} w_t^1 + \beta_{1,1} w_{t-1}^1 + \dots + \beta_{h,1} w_{t-h}^1 + \\ & \beta_{0,2} w_t^2 + \beta_{1,2} w_{t-1}^2 + \dots + \beta_{h,2} w_{t-h}^2 + \\ & \beta_{0,3} w_t^3 + \beta_{1,3} w_{t-1}^3 + \dots + \beta_{h,3} w_{t-h}^3 + v_t, \end{aligned}$$

and imposing a common intercept.¹⁹

This two-stage approach has been widely employed in the literature (e.g., Kilian et al., 2009; Kilian & Hicks, 2013; Bützer et al., 2016). The same two-stage approach could be applied to sign-identified models, except that in this case inference is complicated by the fact that we need to evaluate the second-stage regression for each admissible draw from the posterior of the oil market model (see Herrera & Rangaraju, 2020).

Another common situation in which block-recursive VAR models cannot be applied occurs when the US macroeconomic aggregate of interest is not available for the full estimation period, but only for a comparatively short time span. In that case, it makes sense to recover the oil demand and oil supply shocks from a structural VAR model estimated on the full sample, but to fit the second-stage distributed lag model on the shorter subsample, even when the US variable of interest is available at monthly frequency.

Because crude oil is traded in US dollars, modeling the propagation of global oil demand and oil supply shocks to the US economy is comparatively straightforward. The impact of oil demand and oil supply shocks on other net oil-importing economies, in contrast, will in addition depend on the value of the real

exchange rate. The fact that the real price of oil in domestic consumption units depends on the value of the real exchange rate, however, makes it essential to include the bilateral US dollar real exchange rate in the block recursive VAR model. It is not clear how to employ second-stage distributed lag models in that case.

3. ESTIMATION AND INFERENCE

Oil market models identified by short-run or long-run exclusion restrictions are typically estimated by the least-squares method or by GMM with inference based on bootstrap methods, possibly making allowance for conditional heteroskedasticity in the error term (see [Kilian & Lütkepohl, 2017](#)). In contrast, sign-identified oil market VAR models in the literature have typically been estimated by Bayesian methods. Because these models are identified by inequality restrictions, they do not generate unique point estimates of the impulse responses, but a potentially large set of admissible models that are consistent with the data and satisfy the identifying restrictions.

3.1. Standard Bayesian Inference in Sign-Identified Models

The conventional approach to estimating sign-identified oil market models is based on [Rubio-Ramirez et al. \(2010\)](#), [Arias et al. \(2018\)](#), and [Antolin-Diaz and Rubio-Ramirez \(2018\)](#). Consider a set of reduced-form VAR parameters consisting of the slope parameters, A , and of the reduced-form error covariance matrix, Σ , from which we can compute the lower triangular Cholesky decomposition, P , with positive elements on the diagonal. Candidate solutions for sign-identified models are created by generating at random draws for the orthogonal matrix Q such that PQ represents a candidate solution for the structural impact multiplier matrix B_0^{-1} . Algorithms for generating random draws for Q are discussed in [Kilian and Lütkepohl \(2017\)](#). This procedure is repeated for each of many random draws from the posterior distribution of (A, Σ) . Any combination of A and PQ is a model solution and is associated with a set of structural impulse responses. Model solutions that generate structural impulse response functions that satisfy all identifying restrictions on the impulse responses are considered admissible and are retained. All other solutions are discarded.²⁰

The traditional approach to summarizing this set of admissible models has been to report so-called posterior median response functions. This response function is constructed by connecting the posterior medians obtained from the marginal posterior distribution of each impulse response to form a line across the horizons of a given response function. As noted by [Fry and Pagan \(2011\)](#), [Kilian and Murphy \(2012\)](#), [Inoue and Kilian \(2013\)](#), [Kilian and Lütkepohl \(2017\)](#), and [Kilian \(2022b\)](#), among others, this approach is questionable because the shape of median response function may look substantially different from that of any of the impulse response function that could conceivably be produced by the underlying structural VAR model. [Inoue and Kilian \(2022a\)](#) prove that the posterior median response function is not in general the Bayes estimator.²¹

An alternative approach is to derive the Bayes estimator of the impulse response vector under quadratic and under absolute loss, for example, as in [Inoue and Kilian \(2022a\)](#).

Likewise, pointwise error bands constructed from the quantiles of the marginal distribution of each impulse response are invalid for quantifying the estimation uncertainty about impulse response vectors (e.g., [Sims & Zha, 1999](#); [Inoue & Kilian, 2013, 2022a](#); [Montiel Olea & Plagborg-Møller, 2019](#)). One alternative is to construct approximations to the lowest posterior risk joint credible regions under the relevant loss function that capture the uncertainty about the vector of impulse responses, that provide information about likely departures from the path of the estimated impulse response function, and that are guaranteed to be contained in the set of feasible impulse responses functions (see [Inoue & Kilian, 2022a](#)). Another option is to construct Bayesian joint error bands, as discussed in [Montiel Olea and Plagborg-Møller \(2019\)](#). The key difference between these methods and the traditional approach of reporting posterior median or mean response functions is the focus on joint rather than marginal Bayesian inference.²²

3.2. The Role of the Prior in the Conventional Approach

The conventional approach to Bayesian inference described in Section 3.1 involves postulating a Gaussian-inverse Wishart prior for the reduced form parameters and a uniform (or Haar) prior for the orthogonal matrix Q . Identifying restrictions are imposed directly on the structural impulse responses or on transformations of these responses. [Baumeister and Hamilton \(2015\)](#) recently reiterated the well-known point that this two-step approach may be unintentionally informative about the prior for the structural impulse response functions.

[Baumeister and Hamilton \(2015\)](#) argue that a Haar prior on Q is virtually always unintentionally informative for the marginal prior distribution of the impact responses. However, they never formally derive the prior for the vector of impulse responses implied by the conventional uniform-Gaussian-inverse Wishart prior specification for the VAR model parameters. They only discuss the impulse response distribution conditional on the maximum likelihood estimate of A and Σ , ignoring the prior uncertainty about the latter parameters in the conventional approach. [Inoue and Kilian \(2022b\)](#) show that the impulse response prior implied by the conventional approach differs substantially from the results reported in Baumeister and Hamilton's work.

Nor do Baumeister and Hamilton establish that the impulse response posterior is driven by the Haar prior in general. While there are examples in which the Haar prior dominates the impulse response posterior, Inoue and Kilian examine several sign-identified structural VAR models including the oil market model of [Kilian and Murphy \(2014\)](#) and find no evidence that the impulse response priors implied by the use of conventional priors on A , Σ and Q are necessarily unintentionally informative or economically implausible. Moreover, they find that the impulse response posterior in many typical applications is dominated by the data rather than the Haar prior. This evidence suggests that recent concerns about the

prior for Q have been overstated and that the credibility of these models must be assessed on a case-by-case basis.

3.3. An Alternative Bayesian Approach to Estimating Sign-Identified Oil Market Models

Baumeister and Hamilton (2015) also propose an alternative approach to estimating sign-identified oil market VAR models that dispenses with the Haar prior for Q . In short, their proposal is to specify priors on the parameters of the structural VAR process

$$B_0 y_t = B_1 y_{t-1} + \dots + B_p y_{t-p} + w_t,$$

where w_t is zero mean Gaussian white noise. Abstracting from the lagged coefficients, this involves specifying priors directly on the elements of the matrix B_0 . It is also possible to impose priors on selected elements of B_0^{-1} .²³

The seeming advantage of Baumeister and Hamilton's method is that we can impose nondegenerate priors on the parameters of the model. Rather than imposing that some element of B_0 is positive, for example, we can impose an explicit prior distribution of how the probability mass is distributed in the positive region. The problem is that, in practice, the existing literature provides little guidance about the nature of this prior. Baumeister and Hamilton therefore tend to rely on priors such as truncated Student-t distributions with no explicit economic rationale.

It is important to recognize that the alternative Bayesian approach proposed by Baumeister and Hamilton suffers from exactly the same problem of potentially inadvertently informative impulse response priors as the conventional approach (see Inoue & Kilian, 2022b). The reason is that the impulse responses in this case are defined as a nonlinear function of the structural parameters, B_0, \dots, B_p , much like the impulse responses in the conventional approach are defined as a nonlinear function of A, Σ , and Q . The same point applies when putting additional priors on elements of B_0^{-1} . Thus, there is nothing to choose between their approach and the conventional approach on a priori grounds. If the impulse responses is what we care about, it becomes essential to assess the impulse response prior implied by Baumeister and Hamilton's approach using the tools developed in Inoue and Kilian (2022b) and to defend that prior.²⁴

Three other studies have tried to directly address the problem of potentially unintentionally informative priors for Q . None of these approaches has been employed in the oil market literature to date. One approach is to adapt the prior on the reduced-form parameters, given a uniform prior on Q , to ensure that the implied prior on the structural impulse responses is flat (see Arias et al., 2015). Of course, a flat impulse response prior may not be an economically sensible prior in practice. Another approach is to directly specify a possibly informative prior on the structural impulse responses, as proposed by Plagborg-Møller (2019). The latter approach dispenses with reduced-form priors, but the construction of economically sensible joint priors on the structural responses is nontrivial in general. Finally, Giacomini and Kitagawa (2021) propose an alternative impulse response

estimator that avoids taking a stand on the prior for Q , allowing users to assess the sensitivity of the estimates to the universe of alternative priors for Q . Their approach leaves unanswered what the joint impulse response prior is and whether it is economically sensible.

3.4. How Credible are Time-Varying Coefficient Models of the Oil Market?

A long-standing question is whether to allow the coefficients of global oil market models to vary over time or not. There is a multitude of reasons to expect smooth structural change in the coefficients of models of the global oil market. Examples include temporary capacity constraints in storage and production, transportation bottlenecks, changes in market structure and contract structure over time, changes in the share of oil in value added reflecting conservation and diversification away from oil products, and possibly the development of oil futures markets in the 1980s. This does not necessarily mean that the linear model is a poor approximation, but it raises the question of whether we need to consider the possibility of smooth structural change.

[Baumeister and Peersman \(2013a,b\)](#), in particular, made the case for estimating global oil market models as time-varying coefficient (TVC)-VAR models, building on [Primiceri \(2005\)](#).

The coefficients in TVC-VAR models are typically expressed as latent random walk processes and estimated using Bayesian methods. The error covariance matrix is also allowed to change over time. Impulse responses in TVC-VAR models are dependent on the history of the data and the magnitude of the structural shocks and have to be evaluated by Monte Carlo integration.²⁵ The use of TVC-VAR models is restricted to the construction of impulse responses and counterfactual histories (see [Kilian & Lütkepohl, 2017](#)). Historical decompositions and variance decompositions are not well defined, making it more difficult to interpret the model estimates.

There are a number of other caveats about the use of TVC-VAR models. First, there is evidence that the error bands for the nonlinear impulse responses in this class of models tend to be so wide that they convey essentially no information (see, e.g., [Herrera & Rangaraju, 2020](#)).

Second, a common misperception in applied work is that time-variation in the estimated coefficients of a TVC-VAR model is evidence of time variation in the real world. This view is mistaken because the parameters that govern the smooth structural change in the model coefficients are not identified when there is no time variation. In fact, evidence of time variation is expected due to overfitting, even when the linear model is correct.²⁶ There does not appear to exist a formal statistical test for the absence of time variation in TVC-VAR models.²⁷

Third, for computational reasons, TVC-VAR models of a given dimension can only be estimated allowing for a small number of autoregressive lags. [Baumeister and Peersman \(2013a\)](#), for example, restrict the maximum lag order of their VAR model to four quarterly lags. Since it is well known how important including a sufficiently large number of lags is for accurately estimating the impulse responses in global oil market models, this raises the question of whether any differences

from the response estimates in linear VAR models reflect departures from linearity or unreasonably tight restriction on the largest autoregressive lag order.

Fourth, the specification and identification of the Baumeister and Peersman oil market model is superficially similar to that in Kilian and Murphy (2012). It should be noted, however, that Baumeister and Peersman (2013a) had to relax the oil supply elasticity bound in order to obtain any solutions at all for their econometric model. Their upper bound for the quarterly supply elasticity ranges from 0.6 to 1, respectively, which is unrealistically high compared with the consensus view that the quarterly oil supply elasticity is close to zero.

In defense of their approach, Baumeister and Peersman (2013a) suggest that the impact supply elasticity may have declined since the 1970s and early 1980s. Upon reflection, this argument is not compelling. The cost constraints that discourage oil producers from substantially adjusting production within the quarter do not vary much over time. Baumeister and Peersman present no argument for why the supply elasticity in the 1970s and 1980s would have exceeded conventional benchmarks. Thus, even if we take Baumeister and Peersman's arguments for a decline in the elasticity at face value, the question remains of how to explain their incredibly large estimates of the supply elasticity in the 1970s (with posterior median values as high as 0.85). Likewise, their estimates of the one-quarter price elasticity of oil demand for the 1970s and early 1980s (with posterior medians as low as -0.65) strain credulity.²⁸

4. RECENT CONTROVERSIES ABOUT MODELING OIL MARKETS

Recently, some studies have claimed to have overturned the consensus based on mainstream oil market models that oil supply shocks have not been an important determinant of oil price fluctuations. These claims are closely linked to the definition of the impact price elasticities of oil demand and oil supply in global oil market models and the choice of the elasticity priors.

4.1. How to Define the Impact Price Elasticity of Oil Demand

Some researchers have uncritically applied to the oil market a textbook definition of the price elasticity of demand intended for goods that are not storable. Since crude oil is storable, the amount of oil produced in a given period may be consumed in a refinery or put into storage, (see Kilian & Murphy, 2014). Thus, defining the oil demand elasticity based on an accounting identity that equates oil production with oil consumption at each point in time without accounting for the change in oil inventories, as in Baumeister and Hamilton (2019), for example, is incorrect. This observation applies when using the elasticity concept underlying Kilian and Murphy (2014), which focuses on the change in the use of oil in response to the oil price increase caused by an exogenous shift in oil supply. It also applies when using the textbook definition of the oil demand elasticity in Baumeister and Hamilton (2019), which holds constant the response of all other model variables in defining the elasticity.

This point has important implications in practice. First, the impact price elasticity of oil demand is not explicitly defined in oil market models such as Kilian (2009) or Kilian and Murphy (2012) that do not include oil inventories. Attempts to estimate this elasticity anyway, without accounting for changes in oil inventories, will produce spurious estimates. Second, the claim that a low price elasticity of oil supply necessarily implies an unrealistically high price elasticity of oil demand, which is central to the analysis in Baumeister and Hamilton (2019) and Caldara et al. (2019) is sensitive to how the elasticity is defined (see Kilian, 2022a). Third, the four-variable oil market model specification proposed in Baumeister and Hamilton (2019) mismeasures oil consumption growth, invalidating their estimate of the price elasticity of oil demand, as shown in Kilian (2022a).

4.2. Is There Evidence for a Much Higher Impact Price Elasticity of Oil Supply?

Kilian and Murphy (2012) established that the effect of oil supply shocks on the real price of oil is necessarily modest, if the impact price elasticity of oil supply in model (2) is close to zero, as suggested by economic theory and extraneous evidence. A low impact oil supply elasticity ensures that oil demand shocks are the primary determinant of the variability in the real price of oil. This point has been reaffirmed in the context of model (3) by Herrera and Rangaraju (2020).

Baumeister and Hamilton (2019) propose an alternative model of the global oil market that allows the impact price elasticity of oil supply to be unbounded from above. Their posterior median oil supply elasticity estimate of 0.15 is 22 standard errors above the extraneous elasticity estimate in Newell and Prest (2019) for the United States, for example. They show that, not surprisingly, given this large supply elasticity value, the response of the real price of oil to oil supply shocks is larger in their model than in conventional oil market models.²⁹ Herrera and Rangaraju (2020) establish that this conclusion is not robust. In particular, under any prior that bounds the impact price elasticity of oil supply in line with conventional views of the magnitude of this elasticity, the response of the real price of oil to oil supply shocks in Baumeister and Hamilton's model is similar to that obtained in Kilian and Murphy's model.

Thus, the magnitude of the upper bound on the value of the one-month oil supply elasticity is central for this controversy. Baumeister and Hamilton (2019) argue for not imposing any upper bound, citing evidence in Bjørnland et al. (2021) that this elasticity could be as high as 0.9. They also cite a study by Caldara et al. (2019) that produced estimates larger than conventional elasticity bounds. Kilian (2022a) discusses the limitations of these studies and reviews evidence from other studies, concluding that there is no support for a global oil supply elasticity as large as claimed in these studies.³⁰

4.3. Is the Price Elasticity of Oil Demand Positive?

Most oil market VAR models including Kilian and Murphy (2014) impose the assumption that the impact price elasticity of oil demand is negative. Sockin

[and Xiong \(2015\)](#) make the striking claim that these models are inherently misspecified because, according to their own analysis, the price elasticity of oil demand actually is positive rather than negative. This view is based on a theoretical model of informational frictions. In short, Sockin and Xiong's premise is that rising oil futures prices signal a stronger global economy. Their theoretical model postulates that economic agents have no other means of detecting whether the global economy is booming but to observe the change in oil futures prices. This friction ensures that agents in their model habitually confuse increases in the oil price driven by oil supply shocks or storage demand shocks with increases driven by flow demand shocks.

Sockin and Xiong show that the informational content of increases in oil futures prices in their theoretical model can be so strong that it offsets the dampening effect of higher oil costs on manufacturing activity, resulting in a positive impact price elasticity of oil demand. In other words, Sockin and Xiong argue that higher oil prices induce manufacturing firms to buy more oil. Since standard global oil market VAR models impose a negative oil demand elasticity, they are misspecified under Sockin and Xiong's assumptions.

Sockin and Xiong apply their model to the period of the sustained commodity price boom between 2003 to mid-2008. They argue that an exogenous increase in speculative trading in oil futures markets that was reflected in higher oil futures prices caused manufacturing firms to increase their demand for raw materials such as crude oil in anticipation of an economic boom. Such a nonfundamental demand shift would look like a flow demand shock in the [Kilian and Murphy \(2014\)](#) model rather than a shock to speculative demand.

There are many reasons to be skeptical of the Sockin–Xiong model. First, even if we accept the premise that manufacturing firms may at times be confused about the state of the global economy, the argument that manufacturing firms collectively increased their demand for commodities and their output based on the false premise of a booming global economy for five years without realizing that there was no demand for their products is not credible. A model in which agents are confused for five years at a time and do not learn from their mistakes is a model of irrational expectations. In the real world, firms that make such systematic mistakes go out of business.

Second, there is no extraneous evidence that the one-month price elasticity of oil demand is positive. Nor is there evidence that economic agents predicted a sustained economic boom between 2003 and mid-2008, as maintained by Sockin and Xiong. In fact, [Kilian and Hicks \(2013\)](#) document that professional real GDP growth forecasters systematically underestimated global growth between 2003 and mid-2008.

Third, the very existence of the informational friction postulated in Sockin and Xiong's model is suspect. It is by no means necessary for economic agents to directly observe oil demand and oil supply shocks in order to detect a global economic expansion or decline. There are many indicators of global real economic activity that are readily available to economic agents and allow them to learn whether there actually is an economic boom or not without having to rely exclusively on industrial commodity prices (see [Kilian & Zhou, 2018](#)).

This makes it implausible that manufacturing firms would ever blindly rely on industrial commodity prices as indicators of the direction of the global economy. If they do not, then the mechanisms described by [Sockin and Xiong \(2015\)](#) are irrelevant.

Finally, it should be noted that there is no support for Sockin and Xiong's premise that the increased participation of financial traders in oil futures markets was an exogenous event, as opposed to an endogenous response to fundamental demand or supply pressures and rising expected returns on oil. Nor is there support for the premise that these financial traders necessarily took positions that raised the oil futures price. For a comprehensive review of this debate, the reader is referred to [Fattouh et al. \(2013\)](#).

4.4. Has the Shale Oil Revolution Undermined the Stability of Global Oil Market Models?

An important question is whether the US shale oil revolution that took place after 2008 has undermined the stability of global oil market models.³¹ [Kilian \(2017\)](#) directly addresses this question by constructing a counterfactual for global oil production in the absence of US shale oil production. The study notes that the shale oil revolution can be represented as a sequence of classical shocks to the technology of producing crude oil. [Kilian \(2017\)](#) shows that the sequence of global flow supply shocks that would be required to remove the shale oil component from global crude oil production involves shocks that are neither unusually large nor unusually serially correlated by historical standards. Thus, there is no reason to expect these shocks to have undermined the stability of the coefficients of the structural model.

Another common misperception is that the shale oil revolution must have increased the value of the one-month price elasticity of flow supply. Not only is shale oil only a small fraction of global oil production, limiting the effects of the shale oil supply elasticity on the one-month price elasticity of oil supply in global oil market models, but [Newell and Prest \(2019\)](#) provide independent microeconomic evidence that the one-month price elasticity of supply for shale oil is close to zero, much like the corresponding elasticity for conventional crude oil. This conclusion is also supported by direct evidence from industry sources (see [Kilian, 2022a](#)).

4.5. How to Measure the Global Business Cycle in Oil Market Models

An important question in models of the global oil market is how to measure the global business cycle at monthly frequency. [Kilian \(2009\)](#) proposed a widely used measure of global real economic activity that is based on a proxy for the volume of shipping of industrial raw materials. It is well known that changes in trade volumes need not line up with changes in real output, as measured by world real GDP or world industrial production (see [Kilian & Zhou, 2018](#)).³² Not only is the relationship between global industrial production and the volume of shipping potentially unstable over time, but the timing of these indices differs

because industrial commodities tend to be shipped before changes in industrial production take place. [Funashima \(2020\)](#) confirms that the Kilian index is a leading indicator for global industrial production. The evolution over time of these indices also differs, because the Kilian index embodies an expectational component that is missing in indices of global industrial production because the decision to ship these raw materials is made in expectation of future industrial production (see [Kilian & Zhou, 2018](#); [Kilian, 2019](#)). Thus, there is no a priori reason for these indices to behave similarly, although it has been shown that most of the substantive results in Kilian and Murphy are robust to replacing the Kilian index by the OECD+6 industrial production index created by the OECD (see [Zhou, 2020](#)).³³

An important question, in practice, is whether to express indices of global real economic activity in deviations from trend or in growth rates. The [Kilian \(2009, 2019\)](#) business cycle index is stationary by construction (see [Kilian & Zhou, 2018](#)). It therefore does not make sense to difference this business cycle index.³⁴ When using alternative indicators of global real activity such as the index of OECD+6 industrial production originally created by the OECD, it has been common to express this index in deviations from a log-linear time trend. This approach is natural if we are interested in capturing the global business cycle. The alternative of expressing indicators of global real activity as month-by-month growth rates eliminates long cycles in the real price of oil that are characteristic of commodity markets. It also overemphasizes the high-frequency variation in the data. As a result, log-differencing tends to downplay the importance of flow demand shocks.³⁵

4.6. How to Transform the Real Price of Oil

A recurring question among applied researchers is whether to express the real price of oil in global oil market models in log-levels or in growth rates. There has been no apparent trend in the log real price of oil since 1974. Thus, the conventional approach of expressing the real price of oil in log-levels has the advantage that standard frequentist inference about the estimates of the impulse responses under weak conditions will remain asymptotically valid at short horizons, even when the underlying process contains one or more unit roots (and the variables are possibly cointegrated) (see [Inoue & Kilian, 2020](#)). The same is true for forecast error variance decompositions at any finite horizon. In contrast, differencing the real price of oil when it actually is stationary causes these estimates to be inconsistent and inference to be invalid. Likewise, from a Bayesian point of view, inference remains valid whether the real price of oil is $I(0)$ or $I(1)$. Only the construction of historical decompositions requires the user to take a stand on the order of integration of the real price of oil.

Unit root tests (or for that matter so-called stationarity tests) are not able to discriminate between the $I(0)$ and the $I(1)$ hypothesis for realistic sample sizes, but the fact that the real price of oil has ultimately reverted back to its mean for almost four decades, defying predictions of permanent highs and permanent lows time and again, is suggestive of a slowly mean-reverting $I(0)$ process.

5. NON-TRADITIONAL APPROACHES TO IDENTIFYING OIL DEMAND AND OIL SUPPLY SHOCKS

While the bulk of structural oil market models seeks to recover structural shocks based on the information set provided by a VAR model, there is a smaller literature seeking to exploit extraneous estimates of oil supply and oil demand shocks. Such extraneous shock measures may be used as an external VAR instrument, as an internal VAR instrument, or as regressors in distributed lag models or local projection models (see [Plagborg-Møller & Wolf, 2021](#)).

5.1. Historical Counterfactuals for OPEC Events

Given the importance attached to OPEC actions in the early oil market literature, there has been much interest in identifying exogenous oil supply shocks in OPEC countries. For example, [Hamilton \(2003\)](#) proposed a simple measure of OPEC oil supply shortfalls caused by exogenous geopolitical events. [Kilian \(2008b\)](#) observes that not all of the events considered by Hamilton are plausibly exogenous and points out that Hamilton's OPEC oil supply shock measure is based on indefensible assumption about the timing and magnitude of the OPEC oil supply disruptions. [Kilian \(2008b\)](#) proposes an improved measure that articulates explicit counterfactuals about how the production of other OPEC members would have evolved in the absence of geopolitical events in selected OPEC member countries. [Kilian's \(2008b\)](#) exogenous OPEC oil supply shock series has been updated by [Bastianin and Manera \(2018\)](#).³⁶ [Kilian \(2006\)](#) uses [Kilian's \(2008b\)](#) OPEC oil supply shock measure as an internal instrument within an extended version of the [Kilian \(2009\)](#) oil market model and shows that explicitly modeling OPEC oil supply shocks provides no value added. [Montiel Olea et al. \(2021\)](#) establish the robustness of the results in [Kilian \(2009\)](#) to using the same OPEC oil supply shock measure as an external instrument.

In practice, the usefulness of both the [Kilian \(2008b\)](#) and the [Hamilton \(2003\)](#) measure of exogenous OPEC oil supply shocks and related shock measures is limited for two reasons. First, [Kilian \(2008b\)](#) stresses that all such measures lack predictive power for the real price of oil. This not only contradicts [Hamilton's \(2003\)](#) claim that these oil supply shocks explain major oil price fluctuations, but means that these shocks are weak instruments for the real price of oil, which complicates estimation and inference. This point has been reinforced by more formal evidence in [Kilian \(2008a\)](#), [Montiel Olea et al. \(2021\)](#) and [Kilian \(2022a\)](#).³⁷

Second, [Kilian \(2006, 2009\)](#) emphasize that OPEC members such as Saudi Arabia and to a lesser extent Kuwait and the UAE have had a history of responding to exogenous production shortfalls in other OPEC countries by expanding their own production, so the net shortfall of oil often is much smaller than it seems at first sight. Examples include the Iranian Revolution of 1978/79 and the invasion of Kuwait in 1990. Moreover, the importance of OPEC has changed since the 1970s. OPEC oil production shortfalls since the 1980s have been increasingly offset by production increases in non-OPEC countries, making it important to measure oil supply shocks at the global level rather than at the OPEC level. This evidence, along with the increasing recognition that OPEC oil supply shocks

are weak instruments at best, has led to the demise of the literature on OPEC oil supply shocks caused by geopolitical events. There have been a number of other efforts in the recent literature, however, to construct extraneous estimates of oil demand and oil supply shocks.

5.2. Oil Supply News Shocks

One increasingly popular approach has been to rely on event studies. For example, [Käenzig \(2021\)](#) uses changes in the 6-month oil futures price on days of OPEC announcements as an instrument to identify what he calls an oil supply news shock. His premise is that a change in the oil futures price at high frequency reflects a shift in oil price expectations rather than in the risk premium. These daily shocks are aggregated to monthly frequency by summing the daily shocks over a given month. An increase in oil price expectations (referred to as negative supply news by Käenzig) is associated with an immediate increase in the spot price of oil, a gradual fall in oil production and an increase in oil inventories.

It can be shown that the oil supply news shock discussed in [Käenzig \(2021\)](#) is a special case of a storage demand shock driven by the change in oil price expectations, as discussed in [Kilian and Murphy \(2014\)](#).³⁸ Käenzig's approach is not without limitations, however. For example, the assumption of no change in the risk premium around OPEC announcements need not be correct in practice. Likewise, the temporal aggregation of the daily shock measure to monthly frequency is ad hoc. Nor is it clear which horizon of the term structure of oil futures we should focus on in constructing this shock. Finally, [Degasperi \(2021\)](#) shows that Käenzig's shock measure captures revisions in expectations about oil demand based on the OPEC news release. A more natural interpretation of this shock thus would be as a shock to oil price expectations.

This comment also applies to other oil supply news measures discussed in the literature. For example, [Arezki et al. \(2017\)](#) treat large oil discoveries as news shocks about future oil output. Oil discoveries by construction leave the flow of oil production unaffected for years to come. Instead, they matter because they shift expectations of future oil production and hence expectations of future oil prices. Thus, these shocks do not represent shocks to the flow supply of oil, but to storage demand. It is worth pointing out that oil supply news measures such as giant oil discoveries are poor proxies for exogenous shifts in oil price expectations. The link between oil discoveries and oil price expectations is nonlinear in general. An unexpected giant oil discovery will not move oil price expectations, when expected oil demand is low and the market is well supplied, but may lower oil price expectations substantially, when expected oil demand is high. Just because an oil discovery has a large effect on the economy of the oil producer (and possibly on the real price of oil) in one episode does not necessarily mean that it will do so in another episode. Thus, the coefficients of a regression of changes in the real price of oil on this shock measure will be time-varying. The same concern applies to other oil supply news measures such as indices of attacks on oil shipping. If we force the coefficients to be time-invariant, we obtain a linear approximation, the accuracy of which is sensitive to the unmodeled demand side of the oil market

over the estimation period. In contrast, storage demand shocks, as measured in the [Kilian and Murphy \(2014\)](#) model avoid this instability by focusing on the shift in storage demand associated with a shift in oil price expectations. This avoids having to measure oil price expectations or having to model the nonlinear process that generates oil price expectations.

5.3. Forecast Revisions for Global Growth Forecasts

Measuring shocks to the flow demand for oil associated with the global business cycle is even harder than measuring supply shocks or storage demand shocks. [Kilian and Hicks \(2013\)](#) construct proxies for global flow demand shocks based on revisions of forecasts of real GDP growth made by professional forecasters, as recorded by the Economist Intelligence Unit. They document that professional forecasters between 2003 and mid-2008 persistently underestimated global growth, mainly because they underestimated growth in emerging Asia. This pattern is consistent with the pattern of flow demand shocks recovered by the [Kilian and Murphy \(2014\)](#) model.

5.4. The Narrative Approach to Identifying Oil Supply and Oil Demand Shocks

An alternative is the narrative approach to identification. An early example is [Cavallo and Wu \(2012\)](#). Their approach involves a manual audit of articles published in the *Oil Daily*, the *Oil & Gas Journal*, and the *Monthly Energy Chronology* between 1984 and 2007. Cavallo and Wu use introspection informed by articles in these industry publications to attribute daily changes in the spot price of oil (or, alternatively, the residual of a regression of the change in the spot price on the oil futures spread) to a subset of 22 different types of oil-market related events such as OPEC announcements on oil production, US oil inventory announcements, political developments in the Middle East, or oil production and transportation disruptions that are considered exogenous with respect to the oil price. If there is more than one event on a given day, say, n events, $1/n$ of the change in the price of oil on that day is attributed to each event. Cavallo and Wu then proceed to construct monthly averages of the daily oil price changes that they attribute to exogenous events. They abstract from inflation in defining their exogenous oil price shocks.

One obvious concern with this methodology is that the events in question are not necessarily the cause of these oil price changes. For example, an announcement about oil inventories (or of OPEC production plans) would be expected to have no effect on the price of oil to the extent that the announcement was expected (e.g., [Ye & Karali, 2016](#)). Nor is it clear to what extent political events in an OPEC country cause the price of oil to move. Another concern is that many of the events in question are not plausibly exogenous with respect to the change in the price of oil. For example, OPEC production plans are endogenous with respect to the state of the global economy. Nor is there any justification for giving equal weight to different events. Finally, one has to have a lot of faith in the ability of the authors (and, implicitly, in the ability of the oil journalists whose articles form the basis of the analysis) to solve the underlying identification problem.

The problem faced by these journalists is little different from that faced by sports commentators or stock market pundits explaining the latest results after the fact. There is a natural tendency to list all possible explanations with little regard to the consistency of the explanation over time. It is these explanations that form the raw material for the analysis in [Cavollo and Wu \(2012\)](#). Many of these problems carry over to more recent narrative approaches to identifying oil demand and oil supply shocks.

5.5. Text-Based Measures of Oil Demand and Oil Supply Shocks

[Cavollo and Wu's \(2012\)](#) analysis focuses on measuring exogenous oil price shocks associated with oil demand and oil supply shocks. It does not provide direct measures of the actual demand or supply shocks, only of their impact on the oil price. The direct measurement of oil supply and/or oil demand shocks based on textual analysis is discussed in Datta and Dias (2019), who propose using a systematic and fully automated procedure for gathering information from news articles in two oil industry publications by the Energy Intelligence Group. Their objective is measuring the supply- and demand-driven components of oil price movements. By categorizing words in these articles into expressions linked to “oil supply,” “oil demand,” “increase,” and “decrease,” they construct oil supply and oil demand indicators, demeaned values of which they treat as proxies for oil demand and oil supply shocks.

Datta and Dias' (2019) approach is subject to important drawbacks. First, their claim that text searches allow them to circumvent delays in the availability of data on oil production and global real activity is not persuasive. For example, given the six-month delay in the availability of global oil production data, oil industry journalists clearly know as little as economists about the current level of global oil production. They have no advance knowledge of the data.

Second, there is reason to be skeptical of the reliability of these text classifications. Just because the process is automated does not mean that it reads the data correctly. For example, policymakers routinely mislabel oil production as “supply” and oil consumption as “demand.” Since oil consumption may drop in response to an oil supply disruption, it is easy to see how a mechanical text search could confuse oil supply with oil demand. Another concern is that Datta and Dias' method does not accommodate negations such as “not a supply increase,” which will be coded as a “supply increase,” overturning the meaning of the original text. Another problematic situation arises with texts such as “OPEC oil production will increase its production next year by less than expected.” The textual search will classify this sentence as a “supply increase,” when in reality it reflects a decline in expected oil production, which, all else equal, causes an increase in storage demand.

Third, an index of the frequency of certain words or phrases in the press, on television, or in social media (say the phrase “decline in oil supply”) is not a shock, but an endogenous variable, calling into question Datta and Dias' (2019) interpretation of their word counts. The idea that the unpredictable component of the demand (supply) indicator would be a natural measure of the demand

(supply) shock is a misconception that arises from a semantic confusion about the terms “demand” and “supply.”³⁹ Clearly, there is no reason for the magnitude of this index (or its unpredictable component) to be proportionate to the magnitude of actual oil supply shocks.

Fourth, as the oil market VAR literature has shown, there are many different types of oil demand shocks with different effects on the real price of oil. Lumping all these demand shocks together in textual analysis, as proposed by Datta and Dias (2019), is not likely to produce a sensible measure of oil demand shocks.

Finally, there is no reason to presume that the press always correctly identifies supply and demand shifts. For example, the oil supply disruption that occurred as a result of Hurricane Rita and Katrina in 2005, when offshore oil platforms in the Gulf of Mexico were shut down, was negligible on a global scale. The main effect of these hurricanes was to shut down the US refining industry along the Gulf Coast. This shutdown by construction represented a negative supply shock for the US gasoline market, but a negative demand shock for the global oil market. Nevertheless, this event was frequently incorrectly characterized as a negative oil supply shock in the media. Problems such as these are pervasive in textual analysis.⁴⁰ Likewise, it has been common in the media to attribute oil price increases (as well as decreases) to actions taken or not taken by OPEC, even when the basis of these attributions is not clear. Similarly, the media attention given to the peak-oil hypothesis in discussing rising oil prices has never matched the empirical support for this hypothesis. Yet another example is the extensive public debate about the role of speculative demand in oil markets in the 2000s, triggered by the so-called Masters Hypothesis, which at no point was supported by hard evidence (see Fattouh et al., 2013).

6. CONCLUSION

As this survey has illustrated, there is a plethora of approaches to identifying oil demand and oil supply shocks and of alternative oil market VAR models, reflecting differences in the data, the reduced-form specification, the identifying assumptions, and the estimation method. What is remarkable about the oil market VAR literature to date is how robust its central findings have been across a broad range of model specifications. There are a few insights in the global oil market literature that stand out because they have been reaffirmed time and again. First, although the economics profession has been preoccupied with measuring the impact of oil supply shocks on the real price of oil for three decades, there is robust evidence that the effect of oil demand shocks on the real price of oil is quantitatively more important than that of oil supply shocks. This is true even accounting for the role of the US shale oil revolution in recent years. Second, there is no credible evidence at this point that the oil price fluctuations of the 1970s and early 1980s were primarily caused by exogenous supply disruptions in the Middle East, calling into question the treatment of these episodes in many macroeconomics textbooks. Third, it is not only shocks to the flow demand for oil that matter, but also shocks to the storage demand for oil driven by shifts in oil price expectations.

Such expectations shocks were largely ignored by the literature until a few years ago. Explicitly accounting for oil price expectations helps understand how political events have shaped the oil price and provides a fresh perspective on historical events in the oil market.

This does not mean that the current models have resolved these questions once and for all. There continues to be much interest in exploring new identification schemes, new econometric approaches and new identifying information for oil market models. For example, [Plagborg-Møller \(2019\)](#), [Lanne and Luoto \(2020\)](#), [Braun \(2021\)](#) and [Angelini et al. \(2021\)](#) recently proposed innovative approaches to achieving identification in structural VAR models that are relevant for oil market studies. There is also interest in incorporating nonlinearities into structural VAR models of the oil market. Time will tell which of these ideas will stand the test of time and which will remain interesting thought experiments.

NOTES

1. We do not address in this survey studies of the effects of global oil price shocks on the economy that do not differentiate between oil demand and oil supply shocks.

2. Variations of the [Kilian and Murphy \(2014\)](#) framework have been used in a range of recent studies including [Kilian and Lee \(2014\)](#), [Juvenal and Petrella \(2015\)](#), [Kilian \(2017\)](#), [Baumeister and Hamilton \(2019\)](#), [Herrera and Rangaraju \(2020\)](#), [Zhou \(2020\)](#), [Kilian and Zhou \(2020a, 2022a\)](#), and [Cross et al. \(2022\)](#).

3. [Kilian and Zhou \(2018\)](#) establish the validity of this assumption for the measure of global real activity utilized by [Kilian \(2009\)](#). Whether this assumption holds for alternative measures of global real activity is less clear.

4. There is a strong case to be made for imposing the additional restriction $b_{21}^0 = 0$, because, based on the same reasoning that suggests that $b_{23}^0 = 0$, oil supply shocks should not affect global real activity on impact. [Kilian and Lütkepohl \(2017\)](#) show that this overidentifying restriction cannot be rejected at conventional significance levels. When the [Kilian \(2009\)](#) model is estimated by GMM subject to the overidentifying restriction, the impulse response estimates are indistinguishable from the original estimates in [Kilian \(2009\)](#). The reason is that even without the overidentifying restriction, the estimate of b_{21}^0 is close to zero.

5. While the identifying restrictions in the recursive [Kilian \(2009\)](#) model are not testable, they may be treated as overidentifying restrictions within an oil market VAR model identified by heteroskedasticity. [Lütkepohl and Netšunajev \(2014\)](#) report not being able to reject these overidentifying restrictions. Similarly, [Herwartz and Plödt \(2016\)](#) establish that the impulse response estimates obtained from a non-Gaussian oil market VAR model under the stronger assumption of independent (rather than mutually uncorrelated) structural shocks are similar to those reported in [Kilian \(2009\)](#). In addition, [Angelini et al. \(2021\)](#) provide support for the structure of the [Kilian \(2009\)](#) model based on a proxy VAR model.

6. This contrasts with the traditional approach in macroeconomics of viewing domestic demand shocks as having no long-run effect on real output and identifying any domestic shock that affects real output in the long as a supply or productivity shock (see [Blanchard & Quah, 1989](#)).

7. It should be noted that sign-identified oil market models cannot be used to validate the exclusion restrictions in oil market models, because the latter models are not nested by sign-identified models (see [Kilian & Lütkepohl, 2017](#)).

8. There are different elasticity concepts in use in the oil market literature. For a review of what the appropriate use of each concept is see [Kilian \(2022a\)](#).

9. This conclusion has recently been confirmed for other oil market models (see [Herrera & Rangaraju, 2020](#)).

10. This proxy excludes Chinese crude oil inventories, which became important sometime after 2006, when China started building its strategic oil reserve. The Kilian and Murphy proxy also excludes crude oil stored on oil tankers at sea. A simple way of addressing this problem of measurement is to combine this proxy with the proprietary global crude oil inventory series compiled by the Energy Intelligence Group (EIG), which includes these and other missing components of global crude oil inventories starting in the 2000s. Sensitivity analysis in [Kilian and Lee \(2014\)](#) and [Kilian \(2022b\)](#) using this EIG inventory series confirms the substantive conclusions in [Kilian and Murphy \(2014\)](#).

11. This stabilizes the variance of the oil inventory series. It also facilitates the computation of the price elasticity of oil demand and the imposition of identifying restrictions based on economic theory.

12. Abstracting from changes in the risk premium, expectations shifts would also be captured by oil futures prices, but for much of the estimation period in [Kilian and Murphy \(2014\)](#), the oil futures market did not exist. [Kilian and Murphy \(2014\)](#) show that the oil futures spread, when it does exist, does not contain more information than already captured by the information set of the [Kilian and Murphy \(2014\)](#) model. This amounts to testing whether the VAR model is fundamental in the sense described in [Kilian and Lütkepohl \(2017\)](#).

13. [Kilian and Murphy \(2014\)](#) originally did not impose the restrictions $b_{41}^0 < 0$ and $b_{42}^0 < 0$, but adding these restrictions tends to sharpen the results when conducting inference.

14. It should be noted that the rationale for adding elasticity bounds is not to make the impulse response error bands narrower, although this may be a side-effect, but to eliminate as inadmissible models that are economically implausible. Including these models in the admissible set would bias the estimates of the oil market model (see [Kilian & Murphy, 2012](#)).

15. An empirical test of the proposition that storage below the ground responds to price signals is discussed in [Kilian \(2022a\)](#).

16. In discussing their model, we deliberately abstract from the possible inclusion of a second block of domestic variables and the additional identifying assumptions required to separately identify domestic and global shocks. We also abstract from the fact that the global real activity measure may be any of the indicators discussed in [Kilian and Zhou \(2018\)](#) or the leading principal component of a panel of real activity indicators.

17. Further examples of Bayesian estimation and inference in block recursive oil market model can be found in [Kilian and Zhou \(2020a, 2022a\)](#).

18. This does not mean that the shocks in the lower-right block cannot be identified. One example is [Kilian \(2010\)](#) who jointly modeled the global oil market block and a US gasoline market block. Another example is [Kilian and Zhou's \(2022a\)](#) model of oil prices, exchange rates and interest rates.

19. Alternatively, one could also estimate these responses based on linear local projections, but that approach would be less parsimonious. It should be noted that, as in the local projection literature, using residuals from first-stage regressions creates a generated regressor problem, which is typically ignored in practice.

20. When sign restrictions are augmented by zero restrictions on elements of B_0^{-1} or by narrative restrictions, these posterior draws typically are reweighted based on the importance samplers proposed by [Arias et al. \(2018\)](#) and [Antolin-Diaz and Rubio-Ramirez \(2018\)](#).

21. The same problem applies to the vector of pointwise medians of the oil supply shock series obtained from the [Baumeister and Hamilton \(2019\)](#) model. Likewise, vectors of pointwise quantiles of forecast error variance decompositions, of linear combinations of impulse responses, and of historical decompositions are subject to this problem.

22. Another approach sometimes used in the literature is to report the admissible model that comes closest to an extraneous demand elasticity estimate and to construct credible sets based on the admissible models that come closest to this estimate (see, e.g., [Kilian & Lee, 2014; Kilian, 2017](#)). The drawback of this penalty function approach is that we must be confident in the extraneous demand elasticity estimate.

23. A key difference from Kilian and Murphy (2014) is that Baumeister and Hamilton, like Caldara et al. (2019), define the impact price elasticities of oil supply and oil demand in terms of the parameters of the matrix B_0 rather than as functions of impulse responses. Elasticities in their VAR framework are defined as the response of the variable in question to an exogenous demand or supply shift, holding constant the responses of all other model variables. Since extraneous elasticity estimates typically do not control for these responses, it is not possible to appeal to these estimates in motivating elasticity priors within this framework. In contrast, Kilian and Murphy define the impact price elasticity allowing for other model variables to respond to the shock in question, which matches the way extraneous elasticity estimates are constructed. The fact that this elasticity concept does not match the textbook definition of an elasticity does not detract from its usefulness for imposing identifying restrictions on the structural VAR model. It only means that we must not interpret these elasticities as measuring the slopes of the short-run demand and supply curves (see Kilian, 2022a).

24. Inoue and Kilian (2022b) show that the central tendency of the impulse response prior implied by Baumeister and Hamilton's (2019) prior on the structural parameters of their baseline oil market model is highly economically implausible, unlike the impulse response prior in the Kilian and Murphy model. For example, Baumeister and Hamilton's prior implies that an unexpected exogenous expansion of global real economic activity lowers the real price of oil, contrary to the implications of standard economic models (see Hamilton, 2009).

25. A common mistake in applied work is to report the impulse responses of the TVC-VAR model conditional on the date t estimate of the model coefficients without accounting for the expected evolution of the model coefficients.

26. The degree of overfitting can be controlled to some extent by choosing the prior. However, this makes the estimates of TVC-VAR models sensitive to the specification of the prior distribution. This sensitivity is rarely reported in applied work.

27. An informal diagnostic would be to compute the 95% joint error band for the responses to a one-standard deviation shock from the corresponding linear VAR model. This allows us to assess whether the nonlinear response functions generated by the TVC-VAR model are contained within this error band. If so, time variation is not likely to be important.

28. A partial explanation may be that the impact elasticity of oil demand in Baumeister and Peersman (2013a) is incorrectly defined in that it omits oil inventories (see Section 4).

29. It should be noted that there are other concerns about their model specification and data, as reviewed in Kilian and Zhou (2020b) and Kilian (2022b), but these features are not the main driver of their results and hence can be ignored here.

30. There is also indirect evidence against the high oil supply elasticity estimate in Baumeister and Hamilton (2019). Braun (2021) shows that when exploiting the non-Gaussianity of the error term in Baumeister and Hamilton's (2019) four-variable oil market VAR model for the identification, the importance of the oil supply shock is greatly reduced and their model produces conclusions similar to Kilian and Murphy (2014). Similarly, Carriero et al. (2021) find that the Baumeister and Hamilton (2019) model supports an oil supply elasticity below 0.0258, when incorporating the heteroskedasticity of the errors into the identification.

31. For a review of the US shale oil revolution, the reader is referred to Kilian (2016).

32. Alternative proxies for global real economic activity that recognize that industrial commodities are globally traded are indices of global real industrial commodity prices (e.g., Alquist et al., 2020; Delle Chiaie et al., 2022).

33. Not all models are robust to the choice of the index. For example, the estimates in Baumeister and Hamilton (2019) are sensitive to replacing their industrial production index by the Kilian index (see Herrera & Rangaraju 2020).

34. Hamilton (2021) suggests that there is statistical evidence against the data transformations applied by Kilian (2009, 2019) and that the raw data underlying this index should be log-differenced. This claim has been refuted in Kilian (2022b).

35. An alternative is to express indices of global real activity as cumulative growth rates over one year, as discussed in Kilian and Zhou (2018), or over two years, as proposed by

[Hamilton \(2021\)](#). It is unclear what the rationale of the latter procedure is, except as a descriptive statistic. By construction, cumulative growth rates do not provide a good measure of the business cycle at a given point in time.

36. Unlike Hamilton, Kilian allows OPEC oil supply shocks to be negative as well as positive. He draws attention to the fact that political events such as wars paradoxically may stimulate oil production rather than necessarily curtailing it. For example, during the Iran–Iraq War, both of these countries sought to increase their oil exports in order to finance purchases of military equipment from abroad.

37. The same concern applies to the oil supply shock series constructed in [Caldara et al. \(2019\)](#), as shown in [Kilian \(2022a\)](#).

38. This proposition is testable. Since we know that unexpected oil supply disruptions cause a decline in oil inventories, whereas positive storage demand shocks cause an increase in oil inventory holdings, it is immediately clear from Känzig's impulse response estimates that OPEC announcements that raise oil futures prices affect the real price of oil through storage demand rather than through changes in the flow supply of oil. This conclusion is also consistent with the sharp increase in the real price of oil following the supply news shock. Although Känzig's shock measure captures only a subset of the storage demand shocks identified in recent structural oil market models, it corroborates the dynamic responses to storage demand shocks uncovered by [Kilian and Murphy \(2014\)](#).

39. The term “demand” among policymakers and in news reports is short-hand for global real activity or oil consumption. Clearly, a surprise change in these variables could be caused by any combination of oil demand or oil supply shocks. The underlying structural shocks are not identified. The same is true for unexpected changes in the “supply” of oil, which in press reports refers to oil production. Thus, the resulting “supply shock” measure would not be a structural shock either.

40. Another good example is Iraq's invasion of Kuwait in August 1990, which represented both an oil supply disruption and a positive shock to storage demand, reflecting expectations of future oil supply disruptions and hence rising oil prices. The latter explanation received much less attention in the media, yet has been shown to be as important as the oil supply disruption for understanding the spike in the real price of oil in 1990/91 (see [Kilian & Murphy, 2014](#)).

ACKNOWLEDGMENTS

The views expressed in this chapter are our own and should not be interpreted as reflecting the views of the Federal Reserve Bank of Dallas or any other member of the Federal Reserve System. We thank Ana Maria Herrera, the editors and the referee for helpful discussions.

REFERENCES

- Alquist, R., Bhattachari, S., & Coibion, O. (2020). Commodity-price comovement and global economic activity. *Journal of Monetary Economics*, 112, 41–56.
- Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, 25, 539–573.
- Anderson, S. T., Kellogg, R., & Salant, S. W. (2018). Hotelling under pressure. *Journal of Political Economy*, 126, 984–1026.
- Angelini, G., Cavaliere, G., & Fanelli, L. (2021). *An identification strategy for Proxy-SVARs with weak proxies*. University of Bologna.
- Antolin-Díaz, J., & Rubio-Ramírez, J. F. (2018). Narrative sign restrictions for SVARs. *American Economic Review*, 108, 2802–2839.

- Arezki, R., Ramey, V., & Sheng, L. (2017). News shocks in open economies: Evidence from giant oil discoveries. *Quarterly Journal of Economics*, 132, 103–155.
- Arias, J., Rubio-Ramirez, J. F., & Waggoner, D. F. (2015). *Inference based on SVARs identified with sign and zero restrictions: Theory and applications*. Federal Reserve Board.
- Arias, J., Rubio-Ramirez, J. F., & Waggoner, D. F. (2018). Inference based on SVARs identified with sign and zero restrictions: Theory and applications. *Econometrica*, 86, 685–720.
- Barsky, R. B., & Kilian, L. (2002). Do we really know that oil caused the great stagflation? A monetary alternative. In B. Bernanke & K. Rogoff (Eds.), *NBER Macroeconomics Annual 2001* (pp. 137–183). NBER.
- Bastianin, A., & Manera, M. (2018). How does stock market volatility react to oil shocks? *Macroeconomic Dynamics*, 22, 666–682.
- Baumeister, C., & Hamilton, J. D. (2015). Sign restrictions, vector autoregressions, and useful prior information. *Econometrica*, 83, 1963–1999.
- Baumeister, C., & Hamilton, J. D. (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and oil demand shocks. *American Economic Review*, 109, 1873–1910.
- Baumeister, C., & Peersman, G. (2013a). The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *Journal of Applied Econometrics*, 28, 1087–1109.
- Baumeister, C., & Peersman, G. (2013b). Time-varying effects of oil supply shocks on the U.S. economy. *American Economic Journal: Macroeconomics*, 5, 1–28.
- Bjørnland, H. C., Nordvik, F. M., & Rohrer, M. (2021). Supply flexibility in the shale patch: Evidence from North Dakota. *Journal of Applied Econometrics*, 36, 273–292.
- Bjørnland, H. C., & Zhulanova, J. (2019). *The shale oil boom and the US Economy: Spillovers and time-varying effects*. BI Business School.
- Blanchard, O., & Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, 79, 655–673.
- Braun, R. (2021). *The importance of supply and demand for oil prices: Evidence from non-Gaussianity*. Bank of England.
- Bützer, S., Habib, M. M., & Stracca, L. (2016). Global exchange rate configurations: Do oil shocks matter? *IMF Economic Review*, 64, 443–470.
- Caldara, D., Cavallo, M., & Iacoviello, M. (2019). Oil price elasticities and oil price fluctuations. *Journal of Monetary Economics*, 103, 1–20.
- Carriero, A., Marcellino, M., & Tornese, T. (2021). *Blended Identification in Structural VARs*. Bocconi University.
- Cavallo, M. & Wu, T. (2012). *Measuring oil-price shocks using market-based information* [International Monetary Fund Working Paper No. 12/19].
- Chudik, A., & Georgiadis, G. (2022). Estimation of impulse response functions when shocks are observed at a higher frequency than outcome variables. *Journal of Business and Economic Statistics*, 40, 965–979.
- Cross, J. L., Nguyen, B. H., & Tran, T. D. (2022). The role of precautionary and speculative demand in the global market for crude oil. *Journal of Applied Econometrics*, 37, 882–895.
- Datta, D. D., & Dias, D. A. (2019). *Oil shocks: A textual analysis approach*. Federal Reserve Board.
- Degasperi, R. (2021). *Identification of expectational shocks in the oil market using OPEC announcements*. University of Warwick.
- Delle Chihaie, S., Ferrara, L., & Giannone, D. (2022). Common factors of commodity prices. *Journal of Applied Econometrics*, 37, 461–476.
- Fattouh, B., Kilian, L., & Mahadeva, L. (2013). The role of speculation in oil markets: What have we learned so far? *Energy Journal*, 34, 7–33.
- Fry, R., & Pagan, A. R. (2011). Sign restrictions in structural vector autoregressions: A critical review. *Journal of Economic Literature*, 49, 938–960.
- Funashima, Y. (2020). Global economic activity indexes revisited. *Economics Letters*, 193, 109269.
- Ghysels, E. (2016). Macroeconomics and the reality of mixed frequency data. *Journal of Econometrics*, 193, 294–314.
- Giacomini, R., & Kitagawa, T. (2021). Robust Bayesian inference about set-identified models. *Econometrica*, 89, 1519–1556.

- Hafner, C. M., Herwartz, H. & Wang, S. (2022). Causal inference with (partially) independent shocks and structural signals on the global crude oil market. University of Göttingen.
- Hamilton, J. D. (2003). What is an oil shock? *Journal of Econometrics*, 113, 363–398.
- Hamilton, J. D. (2009). Understanding crude oil prices. *Energy Journal*, 30, 179–206.
- Hamilton, J. D. (2021). Measuring global economic activity. *Journal of Applied Econometrics*, 36, 293–303.
- Hausman, J. A., & Newey, W. K. (1995). Nonparametric-estimation of exact consumers' surplus and deadweight loss. *Econometrica*, 63, 1445–1476.
- Herrera, A. M., & Rangaraju, S. K. (2020). The effect of oil supply shocks on U.S. economic activity: What have we learned? *Journal of Applied Econometrics*, 35, 141–159.
- Herwartz, H., & Plödt, M. (2016). The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach. *Journal of International Money and Finance*, 61, 30–44.
- Inoue, A., & Kilian, L. (2013). Inference on impulse response functions in structural VAR models. *Journal of Econometrics*, 177, 1–13.
- Inoue, A., & Kilian, L. (2020). The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics*, 215, 450–472.
- Inoue, A., & Kilian, L. (2022a). Joint Bayesian inference about impulse responses in VAR models. *Journal of Econometrics*, 231(2), 457–476.
- Inoue, A., & Kilian, L. (2022b). *The role of the prior in estimating VAR models with sign restrictions*. Federal Reserve Bank of Dallas.
- Juvenal, L., & Petrella, I. (2015). Speculation in the oil market. *Journal of Applied Econometrics*, 30, 621–649.
- Käñzig, D. (2021). The macroeconomic effects of oil supply news: Evidence from OPEC announcements. *American Economic Review*, 111, 1092–1125.
- Kilian, L. (2006). *Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market* [CEPR Discussion Paper No. 5994].
- Kilian, L. (2008a). The economic effects of energy price shocks. *Journal of Economic Literature*, 46, 871–909.
- Kilian, L. (2008b). Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy? *Review of Economics and Statistics*, 90(2), 216–240.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99, 1053–1069.
- Kilian, L. (2010). Explaining fluctuations in U.S. gasoline prices: A joint model of the global crude oil market and the U.S. retail gasoline market. *Energy Journal*, 31, 87–104.
- Kilian, L. (2016). The impact of the shale oil revolution on U.S. oil and gas prices. *Review of Environmental Economics and Policy*, 10, 185–205.
- Kilian, L. (2017). The impact of the fracking boom on Arab oil producers. *Energy Journal*, 38, 137–160.
- Kilian, L. (2019). Measuring global real economic activity: Do recent critiques hold up to scrutiny? *Economics Letters*, 178, 106–110.
- Kilian, L. (2022a). Understanding the estimation of oil demand and oil supply elasticities. *Energy Economics*, 107, 105844.
- Kilian, L. (2022b). Facts and fiction in oil market modeling. *Energy Economics*, 110, 105973.
- Kilian, L., & Hicks, B. (2013). Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *Journal of Forecasting*, 32, 385–394.
- Kilian, L., & Lee, T. K. (2014). Quantifying the speculative component in the real price of oil: The role of global oil inventories. *Journal of International Money and Finance*, 42, 71–87.
- Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.
- Kilian, L., & Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models. *Journal of the European Economic Association*, 10, 1166–1188.
- Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29, 454–478.
- Kilian, L., & Park, C. (2009). The impact of oil price shocks on the U.S. Stock market. *International Economic Review*, 50, 1267–1287.

- Kilian, L., & Vega, C. (2011). Do energy prices respond to U.S. macroeconomic news? A test of the hypothesis of predetermined energy prices. *Review of Economics and Statistics*, 93, 660–671.
- Kilian, L., & Zhou, X. (2018). Modeling fluctuations in the global demand for commodities. *Journal of International Money and Finance*, 88, 54–78.
- Kilian, L., & Zhou, X. (2020a). Does drawing down the U.S. strategic petroleum reserve help stabilize oil prices? *Journal of Applied Econometrics*, 35, 673–691.
- Kilian, L., & Zhou, X. (2020b). Oil supply shock redux? Federal Reserve Bank of Dallas.
- Kilian, L., & Zhou, X. (2022a). Oil prices, exchange rates and interest rates. *Journal of International Money and Finance*, 126, 102679.
- Kilian, L., & Zhou, X. (2022b). Heterogeneity in the pass-through from oil to gasoline prices: A new instrument for estimating the price elasticity of gasoline demand. Federal Reserve Bank of Dallas.
- Knittel, C. R., & Pindyck, R. S. (2016). The simple economics of commodity price speculation. *American Economic Journal: Macroeconomics*, 8, 85–110.
- Lanne, M., & Luoto, J. (2020). Identification of economic shocks by inequality constraints in Bayesian structural vector autoregression. *Oxford Bulletin of Economics and Statistics*, 82, 425–452.
- Lippi, F., & Nobili, A. (2012). Oil and the macroeconomy: A quantitative structural analysis. *Journal of the European Economic Association*, 10, 1059–1083.
- Ludvigson, S. C., Ma, S., & Ng, S. (2020). *Shock-restricted structural vector autoregressions*. Columbia University.
- Lütkepohl, H., & Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs. *Journal of Applied Econometrics*, 29, 479–496.
- Montiel Olea, J. L., & Plagborg-Møller, M. (2019). Simultaneous confidence bands: Theory, implementation, and an application to SVARs. *Journal of Applied Econometrics*, 34, 1–17.
- Montiel Olea, J. L., Stock, J. H., & Watson, M. W. (2021). Inference in structural vector autoregressions identified with an external instrument. *Journal of Econometrics*, 225, 74–87.
- Newell, R., & Prest, B. (2019). The unconventional oil supply boom: Aggregate price response from microdata. *Energy Journal*, 40, 1–30.
- Peersman, G., & van Robays, I. (2009). Oil and the Euro area economy. *Economic Policy*, 24, 603–651.
- Plagborg-Møller, M. (2019). Bayesian inference on structural impulse response functions. *Quantitative Economics*, 10, 145–184.
- Plagborg-Møller, M., & Wolf, C. K. (2021). Local projections and VARs estimate the same impulse responses. *Econometrica*, 89, 955–980.
- Primiceri, G. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72, 821–852.
- Rubio-Ramirez, J. F., Waggoner, D., & Zha, T. (2010). Structural vector autoregressions: Theory of identification and algorithms for inference. *Review of Economic Studies*, 77, 665–696.
- Sims, C. A., & Zha, T. (1999). Error bands for impulse responses. *Econometrica*, 67, 1113–1156.
- Sockin, M., & Xiong, W. (2015). Informational frictions and commodity markets. *Journal of Finance*, 70, 2063–2097.
- Stürmer, M. (2018). 150 years of boom and bust—What drives mineral commodity prices. *Macroeconomic Dynamics*, 22, 702–717.
- Ye, S., & Karali, B. (2016). The informational content of inventory announcements: Intraday evidence from the crude oil futures market. *Energy Economics*, 59, 349–364.
- Zhou, X. (2020). Refining the workhorse oil market model. *Journal of Applied Econometrics*, 35, 130–140.

This page intentionally left blank

PART II

FINANCIAL ECONOMETRICS

This page intentionally left blank

CHAPTER 4

QUANTILE IMPULSE RESPONSE ANALYSIS WITH APPLICATIONS IN MACROECONOMICS AND FINANCE

Whayoung Jung^a and Ji Hyung Lee^b

^a*Korea Capital Market Institute, Seoul, Republic of Korea*

^b*Department of Economics, University of Illinois, Urbana, IL, USA*

ABSTRACT

This chapter studies the dynamic responses of the conditional quantiles and their applications in macroeconomics and finance. The authors build a multi-equation autoregressive conditional quantile model and propose a new construction of quantile impulse response functions (QIRFs). The tool set of QIRFs provides detailed distributional evolution of an outcome variable to economic shocks. The authors show the left tail of economic activity is the most responsive to monetary policy and financial shocks. The impacts of the shocks on Growth-at-Risk (the 5% quantile of economic activity) during the Global Financial Crisis are assessed. The authors also examine how the economy responds to a hypothetical financial distress scenario.

Keywords: Quantile impulse response; growth-at-risk; monetary policy; financial shocks; downside risk; quantile regressions

JEL Classification: C22

1. INTRODUCTION

The conditional mean of an outcome variable has been the primary object of study in economics, as it summarizes the central response to explanatory variables. The scientific interests of policymakers and researchers, however, go beyond the conditional mean. Extreme events and business cycles have significant effects on the economy, so we also need to study the tail or shoulder of the outcome distribution. It is therefore important to obtain a more complete picture of the dynamic responses of the conditional distribution.

As an alternative to the conventional mean regression, [Koenker and Bassett \(1978\)](#) proposed quantile regression (QR). Since QR estimates heterogeneous regression coefficients for a response variable across its conditional distribution, it provides a richer interpretation in regression analysis (see [Koenker \(2005\)](#) for the textbook treatment). Recent development in time series QR models enables researchers to study dynamics at various parts of an outcome distribution.

This chapter investigates how quantiles of endogenous variables respond over time to a shock in a vector autoregressive (VAR) model. We build a QR model accommodating important dynamics of macro/financial time series data. Certain cross-sectional and time series characteristics, such as dispersion and persistence, are important in their distributional evolution. While those characteristics are not fully measured by observable variables, they can be effectively captured using quantiles. Thus, we modify an autoregressive conditional quantile specification, and model the conditional quantile of innovations in a VAR model as a function of past observable variables as well as its own past quantiles. We adopt the CAViaR-type model by [Engle and Manganelli \(2004\)](#), but extend it to include the level impact of macro variables and the effect of time-varying volatility, see Remark 2.1 for a detailed discussion.

Based on the proposed multi-equation QR model, we construct QIRFs applying the quantile framework to impulse response functions (IRFs). Yet there is no consensus on how to define the quantile response. We suggest an alternative definition of the QIRF which is compatible with the conventional mean IRF. This QIRF describes a heterogeneous shock-response mechanism across the distribution, complementing the conventional IRF analysis in macro/finance. Thus the QIRF provides effective tools for empirical policy analysis.

We also provide estimation and statistical inference for the QIRF. We suggest a three-step estimation procedure based on the conventional VAR and a QR model. Moreover, valid econometric inferential tools based on both asymptotics and the residual-based moving block bootstrap (MBB) are provided. While some researchers estimate quantile responses using the local projection method proposed by [Jordà \(2005\)](#), theoretical results for the application of local projections to QR models have not been fully explored. This chapter contributes to statistical inference of quantile responses.

In our empirical application, we first investigate quantile responses of US macroeconomic variables to monetary policy and financial shocks. We find that economic activity has the most heterogeneous response across quantiles, while the response of financial variables is relatively homogeneous. An expansionary

monetary policy shock shifts the distribution of economic activity to the right. The shock significantly reduces downside risk to growth but merely affects upside risk. On the contrary, a financial shock shifts the economic activity distribution to the left. The left tail quantiles are substantially more responsive than the median or upper quantiles. This empirical result is in line with [Adrian et al. \(2019\)](#) who show that deteriorating financial conditions strongly increase the downside risks to growth, but not the upside risks. Moreover, a monetary policy shock has much more persistent effects on *Growth-at-Risk*, defined as the conditional 5% quantile of economic activity, than on its mean. The dynamic response of *Growth-at-Risk* to a financial shock decays in a similar way to the mean IRF of economic activity.

Secondly, we quantitatively assess how much downside and upside risks to growth were affected by the financial and monetary policy shocks during and after the Global Financial Crisis (GFC). Financial shocks during August 2007–June 2009 decreased the 5% quantile of Chicago Fed National Activity Index (CFNAI) by 1.5 on average over 2008–2010.¹ However, the decrease in its 95% quantile due to the shocks over the same period was much less: −0.8. Monetary policy shocks during July 2009–December 2015 increased the 5% quantile by 0.4 on average over 2010–2015. The increase suggests the unconventional monetary policy after the GFC effectively reduced downside risks to growth. On the other hand, the upper quantile was hardly affected by the monetary policy.

Thirdly, a measure of financial conditions (National Financial Conditions Index, NFCI) exhibits explosive dynamics at its right tail quantiles (tighter financial conditions).² When severe financial conditions continue for several months, it creates substantial downside risk to the economy. This locally explosive behavior of financial conditions illustrates that a sharp deterioration of financial markets may lead to a financial crisis in a short period of time.

This chapter relates to several strands of literature. From the perspective of econometrics, we extend times series QR models describing heterogeneous dynamics at different parts of the conditional distribution. Over the past few decades, QR methods have been widely applied to time series models to study asymmetric dynamics. [Koenker and Zhao \(1996\)](#), [Xiao and Koenker \(2009\)](#), [Koenker and Xiao \(2006\)](#), and [Xiao \(2009\)](#) estimate QR models for autoregressive conditional heteroskedasticity (ARCH), generalized ARCH (GARCH), autoregressive (AR), and cointegrated processes, respectively. While the conditional quantile is modeled as a linear function of past observations in those models, [Engle and Manganelli \(2004\)](#) develop autoregressive conditional quantile specifications in which the conditional quantile depends not only on past observations but also on unobservable past conditional quantiles. [White et al. \(2015\)](#), WKM henceforth further extend the model to multivariate and multi-quantile models.³ Modifying their autoregressive conditional quantile specification, our QR model effectively incorporates latent information such as dispersion and persistence of distribution into the evolution of the conditional distribution.

This chapter is also related to [Chang et al. \(2021a\)](#) and [Chang et al. \(2021b\)](#) who study time series of cross-sectional distributions. [Chang et al. \(2021a\)](#) analyze

the effects on income distribution of macroeconomic policy shocks, and [Chang et al. \(2021b\)](#) study the dynamics of the global temperature distributions. Our chapter modifies the usual VAR approach to study the conditional quantile dynamics, while [Chang et al. \(2021a\)](#) and [Chang et al. \(2021b\)](#) develop new methodology of mixed autoregression which combines the conventional VAR and functional autoregression (FAR). Please also see [Chang et al. \(2016\)](#) and [Chang et al. \(2020\)](#) for the earlier related studies.

In terms of empirical applications, our chapter closely relates to recent literature investigating asymmetric impacts of economic state variables on upside and downside risks to growth.⁴ [Adrian et al. \(2019\)](#) find that the lower quantiles of economic activity are substantially affected by financial conditions, while the upper quantiles are not. [Adrian et al. \(2018\)](#) study the effects of financial conditions on Growth-at-Risk defined as the conditional 5% quantile of GDP growth. They show that looser financial conditions increase the lower quantiles of GDP growth in the short run, but decrease the lower quantiles in the medium term. [Loria et al. \(2019\)](#) estimate how upside and downside risks to growth respond to various shocks. They find that the lower quantiles of GDP growth are affected more than other quantiles by all shocks under study (monetary policy, credit spread, and productivity shocks). Although we use a different QR model and a different data set, our empirical findings are largely consistent with these studies.

This chapter also contributes to recent studies constructing QIRFs. There have been a few papers investigating the evolution of the conditional quantile in response to a shock. WKM propose the pseudo-QIRF to investigate the response of Value-at-Risk (VaR) based on a GARCH model. The QIRF of [Montes-Rojas \(2019\)](#) estimates the response of quantile paths applying the vector directional quantile model to vector autoregression. The QIRF proposed by [Chavleishvili and Manganelli \(2019\)](#) describes the impact of a shock on the quantile of future quantiles in a VAR model.⁵ [Han et al. \(2022\)](#) estimates the quantile response of financial asset returns under the GARCH framework using the local projection method, and [Lee et al. \(2021\)](#) investigate the quantile response of macroeconomic variables in a VAR framework. Adopting the generalized impulse response function by [Koop et al. \(1996\)](#), this chapter suggests the QIRF which is conceptually comparable to the standard mean IRF. Section 3.2 compares our QIRF to the existing QIRFs in recent studies.

The rest of the chapter is organized as follows. Section 2 introduces the QR model for the innovations in a VAR model. Section 3 proposes the definition and construction of the QIRF, with a brief comparison to the recent literature. Section 4 discusses estimation of the model and QIRF, then provides inferential methods based on asymptotics and the residual-based MBB. In Section 5, we study QIRFs of the US economy. In particular, we study the dynamic quantile responses of economic activity during and after the GFC. We also examine the quantile responses of macroeconomic variables in a distress scenario where a deterioration of financial conditions continues. Section 6 concludes, and the online supplement (Jung & Lee, 2022) complements this chapter with supporting assumptions, proofs, and technical derivations.

2. QR MODEL FOR A STRUCTURAL VAR ANALYSIS

In this section, we introduce the autoregressive conditional quantile model in a VAR framework. Let $\mathbf{y}_t = [y_{1t}, y_{2t} \dots y_{nt}]^\top$ be an $n \times 1$ vector of variables of interest. In a typical structural VAR analysis, \mathbf{y}_t consists of the conditional mean and the innovations:

$$\mathbf{y}_t = \underbrace{A_0 + A_1 \mathbf{y}_{t-1} + \dots + A_p \mathbf{y}_{t-p}}_{\text{The conditional mean}} + \underbrace{\mathbf{u}_t}_{\text{The innovations}}, \quad (1)$$

where $\mathbf{u}_t = [u_{1t}, u_{2t} \dots u_{nt}]^\top = \Theta_0 \boldsymbol{\epsilon}_t$ for a vector of structural shocks $\boldsymbol{\epsilon}_t = [\epsilon_{1t}, \epsilon_{2t} \dots \epsilon_{nt}]^\top \sim (0, I)$ such that $\Sigma_{\mathbf{u}} = \Theta_0 \Theta_0^\top$. The conditional mean and the innovations determine the location and shape of the outcome distribution, respectively.

While the conditional mean is usually the main interest in the VAR literature, researchers begin to pay more attention to time-varying volatility of the innovations. A large literature in macroeconomics employs stochastic volatility (SV) models to take account of volatility dynamics exhibited in macroeconomic variables. In the SV model, variations in volatility are attributed to *a random process*.⁶ However, recent studies suggest that the conditional volatility can be accounted for by observable state variables. For example, [Adrian et al. \(2019\)](#) find that the conditional volatility of GDP growth is correlated with its conditional mean and financial conditions. Their finding suggests that variations in GDP growth volatility can be explained by GDP growth and financial conditions.

In this chapter, we build a QR model explaining systematic dynamics of the innovations. While most of the VAR studies assume a specific distribution such as the multivariate Gaussian for \mathbf{u}_t , we model the conditional quantile of the innovations without such assumptions. Moreover, we allow for the conditional heteroskedasticity of unknown form as in [Brüggemann et al. \(2016\)](#). We take an empirically driven modeling approach using the QR model. This modeling framework allows us to investigate asymmetry in the downside and upside risks, unlike the typical model with symmetric second-moment dynamics. When evolution of the distribution is accompanied by skewness dynamics, this QR approach can be more effective.

Define a natural filtration $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$. All information available at time t is represented by the information set \mathcal{F}_t . For $i = 1, 2, \dots, n$ and $\tau \in (0, 1)$, let $Q_{u_{it}}(\tau | \mathcal{F}_{t-1})$ denote the τ -quantile of u_{it} conditional on \mathcal{F}_{t-1} such that $Pr(u_{it} \leq Q_{u_{it}}(\tau | \mathcal{F}_{t-1}) | \mathcal{F}_{t-1}) = \tau$. Our QR model describes the evolution of the conditional distribution using the following autoregressive conditional quantile specifications. For u_{it} , its conditional τ -quantile is modeled as a function of \mathbf{y}_{t-1} and its own past quantiles:

$$Q_{u_{it}}(\tau | \mathcal{F}_{t-1}) = c_{i,\tau} + \mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1} + \sum_{k=1}^l [b_{k,i,\tau} (Q_{u_{i,t-k}}(\tau_U | \mathcal{F}_{t-k-1}) - Q_{u_{i,t-k}}(\tau_L | \mathcal{F}_{t-k-1})) + d_{k,i,\tau} Q_{u_{i,t-k}}(\tau | \mathcal{F}_{t-k-1})], \quad (2)$$

for some τ_U and τ_L such that $0 < \tau_L < \tau_U < 1$. The vector coefficient $\mathbf{a}_{i,\tau}$ measures how the conditional quantile responds to observable economic state variables, \mathbf{y}_{t-1} , typically correlated with the business cycle. Note that $\mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1}$ represents the impacts of macro variables on the shape of the distribution, not on the location as the conditional mean of u_{it} is zero.

The summation terms in (2) describe autoregressive dynamics of u_{it} along its own quantiles, which are elaborated in the following remark.

Remark 2.1. In Equation (2), $b_{k,i,\tau}$ and $d_{k,i,\tau}$ represent the effect of dispersion of the conditional distribution and the quantile persistence, respectively. As a measure of dispersion, we use the distance between the conditional τ_H -quantile and τ_L -quantile. To get the idea, consider a scale model of $y_t = \sigma_t \varepsilon_t$ with $\sigma_t \in \mathcal{F}_{t-1}$ and $\varepsilon_t \sim \text{iid } F_\varepsilon$, then it is easy to show that

$$Q_{u_{it}}(\tau_H | \mathcal{F}_{t-1}) - Q_{u_{it}}(\tau_L | \mathcal{F}_{t-1}) = (F_\varepsilon^{-1}(\tau_H) - F_\varepsilon^{-1}(\tau_L))\sigma_t,$$

where $F_\varepsilon^{-1}(\cdot)$ is the inverse cumulative distribution function (CDF). Therefore, $Q_{u_{it}}(\tau_H | \mathcal{F}_{t-1}) - Q_{u_{it}}(\tau_L | \mathcal{F}_{t-1})$ is the conditional volatility scaled by $F_\varepsilon^{-1}(\tau_H) - F_\varepsilon^{-1}(\tau_L)$.⁷ Since the distance between the conditional τ_H and τ_L quantiles serves as a measure of volatility, appropriate values need to be chosen for τ_H and τ_L . In this chapter, we let $\tau_H = 84\%$ and $\tau_L = 16\%$, which correspond to left and right shoulders of a distribution.⁸ The proposed QR model provides rich flexibility in modeling the evolution of the conditional distribution. The lagged conditional quantiles incorporate information not captured by observable variables. Dispersion and persistence play important roles in distribution dynamics in practice. Thus, we use the past conditional quantiles to include the dispersion and autoregressive terms (persistence).

A stable VAR(p) process $\{\mathbf{y}_t\}$ has the following Wold moving-average representation: $\mathbf{y}_t = \mu + \sum_{k=0}^{\infty} \Phi_k \mathbf{u}_{t-k}$ where $\mu = \left(I - \sum_{k=1}^p A_k \right)^{-1} A_0$, $\Phi_0 = I$ and $\Phi_i = \sum_{k=1}^{\min(i,p)} \Phi_{i-k} A_k$. As the conditional quantile itself is an autoregressive process, $Q_{u_{it}}(\tau | \mathcal{F}_{t-1})$ in (2) has a representation in terms of past innovations replacing \mathbf{y}_{t-1} with $\mu + \sum_{k=0}^{\infty} \Phi_k \mathbf{u}_{t-1-k}$. The evolution of the conditional distribution of \mathbf{u}_t is described based on its past history $\{\mathbf{u}_k\}_{k=-\infty}^{t-1}$.

Let $\mathbf{Q}_{\mathbf{u}_t}(\tau | \mathcal{F}_{t-1}) = [Q_{u_{1t}}(\tau | \mathcal{F}_{t-1}) \ Q_{u_{2t}}(\tau | \mathcal{F}_{t-1}) \ \dots \ Q_{u_{nt}}(\tau | \mathcal{F}_{t-1})]^\top$. Define an $n \times 1$ matrix $c_\tau = [c_{1,\tau} \ c_{2,\tau} \ \dots \ c_{n,\tau}]^\top$ and $n \times n$ matrices

$$A_\tau = \begin{bmatrix} \mathbf{a}_{1,\tau}^\top \\ \mathbf{a}_{2,\tau}^\top \\ \vdots \\ \mathbf{a}_{n,\tau}^\top \end{bmatrix}, \quad B_{k,\tau} = \begin{bmatrix} b_{k,1,\tau} & 0 & \dots & 0 \\ 0 & b_{k,2,\tau} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_{k,n,\tau} \end{bmatrix}, \quad D_{k,\tau} = \begin{bmatrix} d_{k,1,\tau} & 0 & \dots & 0 \\ 0 & d_{k,2,\tau} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{k,n,\tau} \end{bmatrix}.$$

With the matrix notation, a multi-equation system of (2) can be concisely expressed as

$$\begin{aligned} \mathbf{Q}_{\mathbf{u}_t}(\tau | \mathcal{F}_{t-1}) &= c_\tau + A_\tau \mathbf{y}_{t-1} \\ &+ \sum_{k=1}^l [B_{k,\tau} (\mathbf{Q}_{\mathbf{u}_{t-k}}(\tau_U | \mathcal{F}_{t-k-1}) - \mathbf{Q}_{\mathbf{u}_{t-k}}(\tau_L | \mathcal{F}_{t-k-1})) + D_{k,\tau} \mathbf{Q}_{\mathbf{u}_{t-k}}(\tau | \mathcal{F}_{t-k-1})]. \end{aligned} \quad (3)$$

Remark 2.2. Some QR models study the quantile dependence across variables using high frequency financial time series data. See, e.g., WKM, Li et al. (2015), and Han et al. (2016). However, our model does not allow interactions between quantiles of different variables. The first main reason is that the frequency of macroeconomic data is relatively low, such as quarterly or monthly observations. If we allow non-zero off-diagonal entries in $B_{k,\tau}$ and $D_{k,\tau}$, the estimation becomes infeasible due to the small number of observations relative to the number of parameters. Compared to related macroeconomic literature, the diagonal assumption on $B_{k,\tau}$ and $D_{k,\tau}$ is not too restrictive. In the literature of VAR models with time-varying volatility, volatility dynamics are usually modeled as geometric random walks of its own such as $\log \sigma_{i,t} = \log \sigma_{i,t-1} + \eta_{i,t}$ where $\sigma_{i,t}$ and $\eta_{i,t}$ represent the volatility of structural shock to variable i and the stochastic error, respectively. In those literature, volatility interactions across variables are not allowed either. Secondly, the macroeconomic variables show relatively long-term fluctuation and co-dependence, so the degree of tail dependence across variables is lower than high frequency financial data. Thus, $B_{k,\tau}$ and $D_{k,\tau}$ in (3) are assumed to be diagonal matrices in our model.⁹

Under models (1) and (3), there is a one-to-one relation between $\mathbf{Q}_{\mathbf{y}_t}(\tau | \mathcal{F}_{t-1})$ and $\mathbf{Q}_{\mathbf{u}_t}(\tau | \mathcal{F}_{t-1})$ as the distribution of \mathbf{y}_t is governed by \mathbf{u}_t . Thus, the conditional quantile of \mathbf{y}_t has the following functional form:

$$\begin{aligned} \mathbf{Q}_{\mathbf{y}_t}(\tau | \mathcal{F}_{t-1}) &= A_0 + \sum_{k=1}^p A_k \mathbf{y}_{t-k} + c_\tau + A_\tau \mathbf{y}_{t-1} \\ &+ \sum_{k=1}^l [B_{k,\tau} (\mathbf{Q}_{\mathbf{u}_{t-k}}(\tau_U | \mathcal{F}_{t-k-1}) - \mathbf{Q}_{\mathbf{u}_{t-k}}(\tau_L | \mathcal{F}_{t-k-1})) + D_{k,\tau} \mathbf{Q}_{\mathbf{u}_{t-k}}(\tau | \mathcal{F}_{t-k-1})]. \end{aligned}$$

3. QUANTILE IMPULSE RESPONSE FUNCTION

In this section, we construct QIRFs to investigate how the conditional quantile of an outcome variable responds to a shock over time. If the shock affects only the location of an outcome distribution, quantile responses will be homogeneous across quantiles. If the shock changes the shape of its distribution, however, QIRFs will show heterogeneous impacts on each quantile. The complete picture of the QIRF mechanism, therefore, complements the conventional IRF.

3.1. Definition and Construction of QIRF

Since there has not been an agreement on the definition of the quantile response, we propose an alternative definition using the structure of the QR model in Section 2. We measure the quantile response by comparing the conditional quantiles from the following two dynamic paths:

$$\begin{aligned} & \{\dots, \mathbf{y}_{t-2}, \mathbf{y}_{t-1}, \mathbf{y}_t, \mu_{t+1}(\mathbf{y}_t), \mu_{t+2}(\mathbf{y}_t), \mu_{t+3}(\mathbf{y}_t), \dots\}, \\ & \{\dots, \mathbf{y}_{t-2}, \mathbf{y}_{t-1}, \tilde{\mathbf{y}}_t, \mu_{t+1}(\tilde{\mathbf{y}}_t), \mu_{t+2}(\tilde{\mathbf{y}}_t), \mu_{t+3}(\tilde{\mathbf{y}}_t), \dots\}, \end{aligned} \quad (4)$$

where $\tilde{\mathbf{y}}_t = \mathbf{y}_t + \mathbf{u}_t = \mathbf{y}_t + \Theta_0 \epsilon_t$ and

$$\mu_{t+s}(\mathbf{y}_t) = \begin{cases} E[\mathbf{y}_{t+1} | \mathbf{y}_t; \mathcal{F}_{t-1}], & \text{for } s = 1, \\ E[\mathbf{y}_{t+s} | \mu_{t+s-1}(\mathbf{y}_t), \dots, \mu_{t+1}(\mathbf{y}_t), \mathbf{y}_t; \mathcal{F}_{t-1}], & \text{for } s \geq 2. \end{cases}$$

The two paths are identical up to time $t-1$. At time t , one path is hit by a shock ϵ_t , but the other is not. After time t , realizations of $\{\mathbf{y}_{t+s}\}_{s \geq 1}$ are assumed to be the conditional mean based on their own history in each time path. [Koop et al. \(1996\)](#) illustrate that the conventional mean IRF can be defined as the difference between the conditional means from the two time paths:

$$IRF^{(s)} := \begin{cases} E[\mathbf{y}_{t+1} | \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}] - E[\mathbf{y}_{t+1} | \mathbf{y}_t; \mathcal{F}_{t-1}], & \text{for } s = 1, \\ E[\mathbf{y}_{t+s} | \mu_{t+s-1}(\tilde{\mathbf{y}}_t), \dots, \mu_{t+1}(\tilde{\mathbf{y}}_t), \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}] \\ \quad - E[\mathbf{y}_{t+s} | \mu_{t+s-1}(\mathbf{y}_t), \dots, \mu_{t+1}(\mathbf{y}_t), \mathbf{y}_t; \mathcal{F}_{t-1}], & \text{for } s \geq 2. \end{cases}$$

They propose a generalized impulse response function applying a similar impulse response concept to nonlinear models.

Following their intuition, we define the QIRF as the difference between the conditional quantiles from the two time paths (4).

Definition 3.1

$$QIRF_{\tau}^{(s)} := \begin{cases} Q_{\mathbf{y}_{t+1}}(\tau | \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}) - Q_{\mathbf{y}_{t+1}}(\tau | \mathbf{y}_t; \mathcal{F}_{t-1}), & \text{for } s = 1, \\ Q_{\mathbf{y}_{t+s}}(\tau | \mu_{t+s-1}(\tilde{\mathbf{y}}_t), \dots, \mu_{t+1}(\tilde{\mathbf{y}}_t), \tilde{\mathbf{y}}_t; \mathcal{F}_{t-1}) \\ \quad - Q_{\mathbf{y}_{t+s}}(\tau | \mu_{t+s-1}(\mathbf{y}_t), \dots, \mu_{t+1}(\mathbf{y}_t), \mathbf{y}_t; \mathcal{F}_{t-1}), & \text{for } s \geq 2. \end{cases}$$

Definition 3.1 can be interpreted as a quantile version of the IRF.¹⁰ Under this definition, QIRFs are recursively expressed as

$$\begin{aligned} QIRF_{\tau}^{(s)} &= IRF^{(s)} + A_{\tau} IRF^{(s-1)} \\ &+ \sum_{k=1}^r [B_{k,\tau}(QIRF_{\tau_U}^{(s-k)} - QIRF_{\tau_L}^{(s-k)}) + D_{k,\tau}(QIRF_{\tau}^{(s-k)} - IRF^{(s-k)})], \end{aligned} \quad (5)$$

where $r = \min\{s-1, l\}$ and

$$IRF^{(s)} = \begin{cases} \Theta_0 \epsilon_t, & \text{for } s = 0, \\ \sum_{k=1}^{\min\{s,p\}} A_k IRF^{(s-k)}, & \text{for } s \geq 1. \end{cases}$$

Using the QIRF, heterogeneous dynamics across quantiles can be closely examined. When tail risks are more responsive to volatility, then these dynamics are described by the QIRF at tail quantiles as $B_{k,\tau}$ captures such heterogeneous responses. When a shock has a more persistent impact on specific parts of the distribution, their QIRFs account for this effect as $D_{k,\tau}$ measures the degree of persistence.

3.2. Comparison to Recent QIRF Studies

Recently, there have been a few attempts to investigate quantile dynamics based on QR models. In this subsection, we compare our QIRF with other approaches in the related literature.

Some studies estimate quantile responses applying local projections by [Jordà \(2005\)](#), though econometric theories have not been thoroughly examined yet. [Adrian et al. \(2018\)](#) apply the local projection method to a standard QR model, whereas [Han et al. \(2022\)](#) apply the method to the autoregressive conditional quantile model of WKM. [Loria et al. \(2019\)](#), on the other hand, use local projections indirectly. They first estimate quantiles of a dependent variable using a standard QR model, then construct quantile responses applying the local projection method to an ordinary least squares (OLS) regression in which the response variable is the estimated quantile. Instead of using the local projection, we take a different approach to the construction of quantile responses: the QIRF is constructed based on a multi-equation describing the evolution of the system. As theoretical results for the application of local projections to the quantile framework have not been developed yet, our approach can complement their local projection methods. In particular, our model can be effective when unobservable latent information, which are not easily controlled for in local projections, plays nontrivial role in the evolution of quantiles.

As the attention to downside and upside risks to economic variables increase, a growing number of researchers are constructing QIRFs in a VAR framework. However, there is not yet an agreement about how to define the quantile response, and each study has defined it in different ways. Similar to our chapter, WKM and [Montes-Rojas \(2019\)](#) define their QIRFs as the difference between the conditional quantiles from two time paths: one path is affected by a shock, but the other path (as the benchmark) is not. However, their formulation of the time paths is different from ours. For the pseudo-QIRF of WKM, the two time paths are identical except at the time when a shock hits the system. This scenario does not account for the effect of the shock on subsequent conditional distributions. As a result, their pseudo-QIRF underestimates the magnitude of a shock on quantile responses.¹¹ [Montes-Rojas \(2019\)](#), on the other hand, assumes persistent realizations of the lower (or upper) quantile for the time paths compared in his

QIRF construction.¹² Thus, his QIRF describes the cumulative impact of shocks as if the economy is under continuous distress. Moreover, the quantile responses are not directly comparable across quantiles because the response at each quantile is against a series of different shocks.

The QIRF of Lee et al. (2021) is conceptually the same as the QIRF of this chapter. Their QIRF measures the expected change in the conditional quantile due to a shock. But, their underlying QR model is different from ours: their model specifies the conditional quantile of y , whereas ours specifies that of \mathbf{u}_t . As a result, their QIRF is constructed in a different way from ours. Chavleishvili and Manganelli (2019) take a quite different approach to quantile responses. Their QIRF is derived applying what they call *the law of iterated quantiles*, and it measures the effect of a shock on the quantiles of future quantiles.

4. ESTIMATION AND STATISTICAL INFERENCE

In this section, we provide inferential tools for the QIRF estimation. We discuss inferential methods based on asymptotics and the residual-based MBB. This chapter mainly employs the statistical inference of Brüggemann et al.(2016). While their inferential methods are for the mean impulse response in VAR models with conditional heteroskedasticity, this chapter applies the methods for the quantile impulse response.

4.1. Assumption

First, we adopt Assumption 2.1 of Brüggemann et al.(2016) who examine stable VAR models with conditional heteroskedasticity.

Assumption 4.1. (1) Let $A(L) = I - \sum_{k=1}^p A_k L^k$. $\det(A(z)) \neq 0$ for all $|z| \leq 1$. (2) The white noise process $\{\mathbf{u}_t\}$ is strictly stationary and strong mixing. (3) $\Sigma_{\mathbf{u}} = \mathbb{E}[\mathbf{u}_t \mathbf{u}_t^\top]$ is positive definite.

Assumption 4.2. (1) Let $\alpha_{\mathbf{u}}(k) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_k^\infty} |P(A \cap B) - P(A)P(B)|$ for $k \in \mathbb{N}$ denote the α -mixing coefficients of the process $\{\mathbf{u}_t\}$. For some $\delta > 0$, $\sum_{k=1}^{\infty} (\alpha_{\mathbf{u}}(k))^{\delta/(2+\delta)} < 0$ and $\mathbb{E}|\mathbf{u}_t|_{4+2\delta}^{4+2\delta}$ is bounded where $|A|_p = \left(\sum_{i,j} |a_{ij}|^p \right)^{1/p}$ for $A = (a_{ij})$. (2) For $a, b, c \in \mathbb{Z}$, define an $(n^2 \times n^2)$ matrix $\kappa_{a,b,c} := \mathbb{E}\left(\text{vec}(\mathbf{u}_t \mathbf{u}_{t-a}^\top) \text{vec}(\mathbf{u}_{t-b} \mathbf{u}_{t-c}^\top)^\top\right)$ and denote $\tilde{n} = n(n+1)/2$. For $m \in \mathbb{N}$, there exists an $(n^2 m + \tilde{n}) \times (n^2 m + \tilde{n})$ positive definite matrix Σ_m defined as

$$\Sigma_m := \begin{bmatrix} \Sigma_m^{(1,1)} & \Sigma_m^{(2,1)\top} \\ \Sigma_m^{(2,1)} & \Sigma^{(2,2)} \end{bmatrix},$$

where

$$\begin{aligned}\Sigma_m^{(1,1)} &:= \left(\sum_{h=-\infty}^{\infty} \kappa_{i,h,h+j} \right)_{i,j=1,2,\dots,m}, \\ \Sigma_m^{(2,1)} &:= \sum_{h=-\infty}^{\infty} L_n(\kappa_{0,h,h+1}, \dots, \kappa_{0,h,h+m}), \\ \Sigma^{(2,2)} &:= \sum_{h=-\infty}^{\infty} L_n(\kappa_{0,h,h} - \text{vec}(\Sigma_u) \text{vec}(\Sigma_u)^\top) L_n^\top,\end{aligned}$$

and L_n is the $(\tilde{n} \times n^2)$ elimination matrix which is defined such that $\text{vech}(A) = L_n \text{vec}(A)$ holds for any $(n \times n)$ matrix A .

For the innovation process, we assume it satisfies the mixing condition of Assumption 4.2 (1) instead of the *iid* assumption, and this allows for conditional heteroskedasticity. Their dynamics are described using the conditional quantile model of (3). Given Assumption 4.2 (2), the asymptotic covariance matrix of the VAR model is positive definite.

We also adopt the modeling assumptions in Section 2 and Appendix of WKM. As our QR model is not the same as their model, we adjust their assumptions.¹³ With Assumptions A.1–A.5, which are provided in the online supplement, the asymptotic distribution of the QR estimator is derived.

4.2. Estimation

Estimation of our QR model is not trivial because of the autoregressive conditional quantile specifications for multi quantiles. We use the following three-step estimation procedure.¹⁴

Step 1

Estimate the following VAR model using OLS:

$$\mathbf{y}_t = A_1 \mathbf{y}_{t-1} + \dots + A_p \mathbf{y}_{t-p} + \mathbf{u}_t, \quad \mathbf{u}_t \sim (0, \Sigma_u).$$

Let the estimator be $\{\hat{A}_k\}_{k=1}^p$. We denote $\hat{\mathbf{y}}_t = \sum_{k=1}^p \hat{A}_k \mathbf{y}_{t-k}$ and $\hat{\mathbf{u}}_t = [\hat{u}_{1t} \ \hat{u}_{2t} \ \dots \ \hat{u}_{mt}]^\top = \mathbf{y}_t - \hat{\mathbf{y}}_t$.

Step 2

Define a $(1+n+2l) \times 1$ vector $\theta_{i,\tau} := [c_{i,\tau} \ \mathbf{a}_{i,\tau}^\top \ b_{1,i,\tau} \ \dots \ b_{l,i,\tau} \ d_{1,i,\tau} \ \dots \ d_{l,i,\tau}]^\top$ and estimate coefficients at the *shoulder quantiles*, τ_U and τ_L , by solving the following minimization problem:

$$\min_{\theta_{i,\tau_U}, \theta_{i,\tau_L}} \frac{1}{T} \sum_{t=1}^T \left[\rho_{\tau_U} \left(\hat{u}_t - q_{i,t,\tau_U}(\theta_{i,\tau_U}, \theta_{i,\tau_L}) \right) + \rho_{\tau_L} \left(\hat{u}_t - q_{i,t,\tau_L}(\theta_{i,\tau_U}, \theta_{i,\tau_L}) \right) \right], \quad (6)$$

where

$$\begin{aligned} q_{i,t,\tau}(\theta_{i,\tau_U}, \theta_{i,\tau_L}) &= c_{i,\tau} + \mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1} + \sum_{k=1}^l [b_{k,i,\tau}(q_{i,t-k,\tau_U}(\theta_{i,\tau_U}, \theta_{i,\tau_L}) - q_{i,t-k,\tau_L}(\theta_{i,\tau_U}, \theta_{i,\tau_L})) \\ &\quad + d_{k,i,\tau} q_{i,t-k,\tau}(\theta_{i,\tau_U}, \theta_{i,\tau_L})] \end{aligned}$$

for $\tau = \tau_U, \tau_L$, and $\rho_\tau(u) = u(\tau - \mathbb{I}[u < 0])$.

Step 3

Based on the estimator $\hat{\theta}_{i,\tau_U}$ and $\hat{\theta}_{i,\tau_L}$ from Step 2, estimate the τ -coefficient $\theta_{i,\tau}$ for $\tau \neq \tau_U, \tau_L$, which solves the minimization problem below:

$$\min_{\theta_{i,\tau}} \frac{1}{T} \sum_{t=1}^T \rho_\tau(\hat{u}_t - q_{i,t,\tau}(\theta_{i,\tau} | \hat{\theta}_{i,\tau_U}, \hat{\theta}_{i,\tau_L})),$$

where

$$\begin{aligned} q_{i,t,\tau}(\theta_{i,\tau} | \theta_{i,\tau_U}, \theta_{i,\tau_L}) &= c_{i,\tau} + \mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1} \\ &\quad + \sum_{k=1}^l [b_{k,i,\tau}(q_{i,t-k,\tau_U}(\theta_{i,\tau_U}, \theta_{i,\tau_L}) - q_{i,t-k,\tau_L}(\theta_{i,\tau_U}, \theta_{i,\tau_L})) \\ &\quad + d_{k,i,\tau} q_{i,t-k,\tau}(\theta_{i,\tau} | \theta_{i,\tau_U}, \theta_{i,\tau_L})] \end{aligned}$$

for $\tau \neq \tau_U, \tau_L$.

Remark 4.1. Instead of Steps 2 and 3, $\theta_{i,\tau_U}, \theta_{i,\tau_L}$ and $\theta_{i,\tau}$ could be estimated simultaneously. The simultaneous estimation solves the following minimization problem:

$$\begin{aligned} \min_{\theta_{i,\tau_U}, \theta_{i,\tau_L}, \theta_{i,\tau}} \frac{1}{T} \sum_{t=1}^T & [\rho_{\tau_U}(\hat{u}_t - q_{i,t,\tau_U}(\theta_{i,\tau_U}, \theta_{i,\tau_L})) \\ & + \rho_{\tau_L}(\hat{u}_t - q_{i,t,\tau_L}(\theta_{i,\tau_U}, \theta_{i,\tau_L})) + \rho_\tau(\hat{u}_t - q_{i,t,\tau}(\theta_{i,\tau_U}, \theta_{i,\tau_L}, \theta_{i,\tau}))], \end{aligned}$$

$$\begin{aligned} \text{where } q_{i,t,\tau}(\theta_{i,\tau_U}, \theta_{i,\tau_L}, \theta_{i,\tau}) &= c_{i,\tau} + \mathbf{a}_{i,\tau}^\top \mathbf{y}_{t-1} \\ &\quad + \sum_{k=1}^l [b_{k,i,\tau}(q_{i,t-k,\tau_U}(\theta_{i,\tau_U}, \theta_{i,\tau_L}) - q_{i,t-k,\tau_L}(\theta_{i,\tau_U}, \theta_{i,\tau_L})) \\ &\quad + d_{k,i,\tau} q_{i,t-k,\tau}(\theta_{i,\tau_U}, \theta_{i,\tau_L}, \theta_{i,\tau})] \end{aligned}$$

for $\tau \neq \tau_U, \tau_L$. Under the simultaneous estimation, a different choice of τ leads to different estimates for θ_{i,τ_U} and θ_{i,τ_L} though the differences are not substantial. Since the distance between the τ_U and τ_L -quantiles serves as a measure of volatility in the QR model, the coefficients θ_{i,τ_U} and θ_{i,τ_L} play an important role in the construction of the QIRF. Hence, we adopt the three-step estimation strategy which yields robust estimates of the coefficients (thus the QIRF). We show the QR estimators are consistent in the following section.

In this chapter, we assume a structural shock is identified by the Cholesky restriction: $\Sigma_{\mathbf{u}} = \Theta_0 \Theta_0^\top$ and $\hat{\Sigma}_{\mathbf{u}} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t^\top = \hat{\Theta}_0 \hat{\Theta}_0^\top$ where Θ_0 and $\hat{\Theta}_0$ are lower triangular.¹⁵ Let $IRF_i^{(s)}$ and $QIRF_{i,\tau}^{(s)}$ denote the i -th element of $IRF^{(s)}$ and $QIRF_\tau^{(s)}$, respectively. From (5), the estimator for $QIRF_{i,\tau}^{(s)}$ is recursively constructed using the OLS estimator ($\{\hat{A}_k\}_{k=1}^p$ and $\hat{\Theta}_0$) and QR estimator ($\hat{\theta}_{i,\tau_U}, \hat{\theta}_{i,\tau_L}$ and $\hat{\theta}_{i,\tau}$):

$$\begin{aligned} \widehat{QIRF}_{i,\tau}^{(s)} &= \widehat{IRF}_i^{(s)} + \hat{\mathbf{a}}_{i,\tau}^\top \widehat{IRF}^{(s-1)} \\ &+ \sum_{k=1}^r \left[\hat{b}_{k,i,\tau} \left(\widehat{QIRF}_{i,\tau_U}^{(s-k)} - \widehat{QIRF}_{i,\tau_L}^{(s-k)} \right) + \hat{d}_{k,i,\tau} \left(\widehat{QIRF}_{i,\tau}^{(s-k)} - \widehat{IRF}_i^{(s-k)} \right) \right], \end{aligned} \quad (7)$$

where $r = \min\{s-1, l\}$ and

$$\widehat{IRF}^{(s)} = \begin{cases} \hat{\Theta}_0 \epsilon_t, & \text{for } s = 0, \\ \sum_{k=1}^{\min\{s,p\}} \hat{A}_k \widehat{IRF}^{(s-k)} & \text{for } s \geq 1. \end{cases}$$

4.3. Asymptotic Inference

Let $\beta = \text{vec}(A_1, \dots, A_p)$, $\sigma = \text{vech}(\Sigma_{\mathbf{u}})$, $\hat{\beta} = \text{vec}(\hat{A}_1, \dots, \hat{A}_p)$ and $\hat{\sigma} = \text{vech}(\hat{\Sigma}_{\mathbf{u}})$. Using the moving-average representation of $\mathbf{y}_t = \sum_{k=0}^{\infty} \Phi_k \mathbf{u}_{t-k}$, let $\Phi_0 = I$ and $\Phi_i = \sum_{k=1}^{\min(i,p)} \Phi_{i-k} A_k$ for $i \in \mathbb{N}$. Define an $np \times n$ matrix $C_i = (\Phi_{i-1}^\top, \dots, \Phi_{i-p}^\top)^\top$ and an $np \times np$ matrix $\Gamma = \sum_{k=1}^{\infty} C_k \Sigma_{\mathbf{u}} C_k^\top$.

Lemma 4.1 follows from Theorem 2.1 of [Brüggemann et al. \(2016\)](#).

Lemma 4.1. *Under Assumptions 4.1 and 4.2,*

$$\sqrt{T} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\sigma} - \sigma \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} V_{\hat{\beta}} & V_{\hat{\sigma}, \hat{\beta}}^\top \\ V_{\hat{\sigma}, \hat{\beta}} & V_{\hat{\sigma}} \end{bmatrix}\right), \quad (8)$$

where

$$\begin{aligned} V_{\hat{\beta}} &= (\Gamma^{-1} \otimes I) \left(\sum_{i,j=1}^{\infty} (C_i \otimes I) \sum_{h=-\infty}^{\infty} \kappa_{i,h,h+j} (C_j \otimes I)^\top \right) (\Gamma^{-1} \otimes I)^\top, \\ V_{\hat{\sigma}, \hat{\beta}} &= L_n \left(\sum_{j=1}^{\infty} \sum_{h=-\infty}^{\infty} \kappa_{0,h,h+j} (C_j \otimes I)^\top \right) (\Gamma^{-1} \otimes I)^\top, \\ V_{\hat{\sigma}} &= L_n \left(\sum_{h=-\infty}^{\infty} (\kappa_{0,h,h} - \text{vec}(\Sigma_{\mathbf{u}}) \text{vec}(\Sigma_{\mathbf{u}})^\top) \right) L_n^\top. \end{aligned}$$

$$V_{\hat{\sigma}, \hat{\beta}} = L_n \left(\sum_{j=1}^{\infty} \sum_{h=-\infty}^{\infty} \kappa_{0,h,h+j} (C_j \otimes I)^{\top} \right) (\Gamma^{-1} \otimes I)^{\top}.$$

As $IRF^{(s)}$ is continuously differentiable functions of β and σ , the asymptotic distribution of the IRF estimator is obtained applying the Delta method to Lemma 4.1. Lemma 4.2 follows from Corollary 5.1 of Brüggemann et al. (2016).

Lemma 4.2. *Under Assumptions 4.1 and 4.2,*

$$\sqrt{T} \left(\widehat{IRF}_i^{(s)} - IRF_i^{(s)} \right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, V_{\widehat{IRF}_i^{(s)}}\right), \quad (9)$$

where $V_{\widehat{IRF}_i^{(s)}} = C_{i,\beta}^{(s)} V_{\hat{\beta}} C_{i,\beta}^{(s)\top} + C_{i,\sigma}^{(s)} V_{\hat{\sigma}} C_{i,\sigma}^{(s)\top} + C_{i,\beta}^{(s)} V_{\hat{\sigma}, \hat{\beta}} C_{i,\sigma}^{(s)\top} + C_{i,\sigma}^{(s)} V_{\hat{\sigma}, \hat{\beta}} C_{i,\beta}^{(s)\top}$, $C_{i,\beta}^{(s)} = \frac{\partial IRF_i^{(s)}}{\partial \beta^{\top}}|_{\hat{\beta}}$

and $C_{i,\sigma}^{(s)} = \frac{\partial IRF_i^{(s)}}{\partial \sigma^{\top}}|_{\hat{\sigma}}$.

Define $\Theta_{i,\tau} := [\theta_{i,\tau_U}^{\top} \quad \theta_{i,\tau_L}^{\top} \quad \theta_{i,\tau}^{\top}]^{\top}$ and its estimator $\hat{\Theta}_{i,\tau} := [\hat{\theta}_{i,\tau_U}^{\top} \quad \hat{\theta}_{i,\tau_L}^{\top} \quad \hat{\theta}_{i,\tau}^{\top}]^{\top}$

which is estimated using the consistent estimator $\hat{\beta}$.¹⁶ Consistency and asymptotic normality of the QR estimator follow from the asymptotic theories of Engle and Manganelli (2004) and WKM.

Lemma 4.3. *Under Assumptions 4.1, 4.2, and A.1–A.5,*

$$\sqrt{T} \left(\hat{\Theta}_{i,\tau} - \Theta_{i,\tau} \right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{Q}_{i,\tau}^{-1} \mathbf{V}_{i,\tau} \mathbf{Q}_{i,\tau}^{-1}\right), \quad (10)$$

where $\mathbf{V}_{i,\tau}$ and $\mathbf{Q}_{i,\tau}$ are defined in the online supplement.

While the above lemmas derive the asymptotic distributions of the OLS, IRF, and QR estimators, it is challenging to derive the asymptotic distribution of the QIRF estimator. Since the estimator is a function of both the OLS and QR estimators ($\hat{\beta}, \hat{\sigma}$ and $\hat{\Theta}_{i,\tau}$) as in (7), its asymptotic distribution could be derived applying the Delta method to the joint asymptotic distribution of $\hat{\beta}, \hat{\sigma}$ and $\hat{\Theta}_{i,\tau}$. However, the derivation is not easy because the QR estimator does not have an explicit expression.

Moreover, it is difficult to estimate the asymptotic covariance matrix of the QR estimator accurately because of a nuisance parameter.¹⁷ The asymptotic inference is less satisfactory particularly at tail quantiles due to the small number of relevant observations. Accordingly, the performance of the QIRF estimator based on asymptotics might not be ideal. Thus, we provide inferential tools for the QIRF based on the residual-based MBB.

4.4. Residual-Based MBB

This section describes the residual-based MBB procedure and provides the bootstrap consistency for the QIRF estimator. The procedure mainly follows the bootstrap algorithm in Brüggemann et al. (2016).¹⁸ We propose the following bootstrap procedure for the inference of the QIRF.

Step 1

Choose a block length $l_b < T$ and let $N = \lceil T/l_b \rceil$ be the number of blocks needed such that $l_b N \geq T$. Draw i_1, \dots, i_N from a random variable uniformly distributed on the set $\{1, 2, \dots, T - l_b + 1\}$. Define $(n \times l_b)$ -dimensional blocks $B_{i,l_b} = (\hat{\mathbf{u}}_i, \hat{\mathbf{u}}_{i+1}, \dots, \hat{\mathbf{u}}_{i+l_b-1})$ where $\hat{\mathbf{u}}_i$ is defined in Section 4.2. Bootstrap residuals $\{\mathbf{u}_t^*\}_{t=1}^T$ are obtained laying blocks $B_{i_1,l_b}, B_{i_2,l_b}, \dots, B_{i_N,l_b}$ end-to-end together with the last $Nl_b - T$ values discarded.

Step 2

Define centered bootstrap residuals $\{\mathbf{u}_t^*\}_{t=1}^T$:

$$\mathbf{u}_{il_b+j}^* := \mathbf{u}_{il_b+j}^{(*)} - \frac{1}{T-l_b+1} \sum_{k=0}^{T-l_b} \hat{\mathbf{u}}_{j+k},$$

for $j = 1, 2, \dots, l_b$ and $i = 0, 1, \dots, N-1$. Set bootstrap pre-sample values $\{\mathbf{y}_t^*\}_{t=-p+1}^0 = 0$ and generate the bootstrap sample $\{\mathbf{y}_t^*\}_{t=1}^T$ according to

$$\mathbf{y}_t^* = \sum_{k=1}^p \hat{A}_k \mathbf{y}_{t-k}^* + \mathbf{u}_t^*.$$

Step 3

Compute the bootstrap OLS estimator $\hat{\beta}^* = \text{vec}(\hat{A}_1^*, \dots, \hat{A}_p^*)$ based on $\{\mathbf{y}_t^*\}_{t=-p+1}^T$.

Denote the bootstrap residual from the VAR model $\hat{\mathbf{u}}_t^* = \mathbf{y}_t^* - \sum_{k=1}^p \hat{A}_k^* \mathbf{y}_{t-k}^*$, and define $\hat{\Sigma}_{\mathbf{u}}^* = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{u}}_t^* \hat{\mathbf{u}}_t^{*\top}$ and $\hat{\sigma}^* = \text{vech}(\hat{\Sigma}_{\mathbf{u}}^*)$.

Step 4

Based on $\{\hat{\mathbf{u}}_t^*\}_{t=1}^T$ and $\{\mathbf{y}_t^*\}_{t=-p+1}^0$, estimate QR estimator $\hat{\Theta}_{i,\tau}^* = \begin{bmatrix} \hat{\theta}_{i,\tau_U}^{*\top} & \hat{\theta}_{i,\tau_L}^{*\top} & \hat{\theta}_{i,\tau}^{*\top} \end{bmatrix}^\top$ following Steps 2 and 3 of Section 4.2.

Step 5

Using $\hat{\beta}^*$, $\hat{\sigma}^*$ and $\hat{\Theta}_{i,\tau}^*$, compute $\widehat{IRF}^{(s)*}$ and $\widehat{QIRF}_{\tau}^{(s)*}$, the bootstrap version of $IRF^{(s)}$ and $QIRF_{\tau}^{(s)}$.

Step 6

Let the number of bootstrap be B which is large. Repeating Steps 1 through 5 B times, empirical distributions of $\widehat{IRF}^{(s)*}$ and $\widehat{QIRF}_{\tau}^{(s)*}$ based on the repetition provide consistent approximation of the distributions of the IRF and QIRF estimators, respectively. The lower and upper bounds of the $100 \cdot (1-\alpha)\%$ confidence interval for $\widehat{IRF}^{(s)}$ is constructed using $100 \cdot \left(1 - \frac{\alpha}{2}\right)$ and $100 \cdot \left(\frac{\alpha}{2}\right)$ empirical quantiles of $\widehat{IRF}^{(s)*}$. The confidence interval for $\widehat{QIRF}_{\tau}^{(s)}$ is obtained in the same way.

As for the choice of the block length in our empirical applications, we use the rule $l_b = \kappa T^{1/4}$ following [Jentsch and Lunsford \(2019\)](#). The following assumption ensures the bootstrap consistency. The assumption is implied by the existent of all moments up to order eight of $\{\mathbf{u}_t\}$ and $\sum_{k=1}^{\infty} k^6 (\alpha_u(k))^{\delta/(14+\delta)} < 0$. See Remark A.1 of [Künsch \(1989\)](#).

Assumption 4.3. *The innovation process $\{\mathbf{u}_t\}$ has absolutely summable cumulants up to order eight. That is, for all $j = 2, \dots, 8$ and $a_1, \dots, a_j \in \{1, \dots, n\}$,*

$$\sum_{h_2, \dots, h_j=-\infty}^{\infty} |\text{cum}_{(a_1, \dots, a_j)}(0, h_2, \dots, h_j)| < \infty$$

where $\text{cum}_{(a_1, \dots, a_j)}(0, h_2, \dots, h_j)$ denotes the j -th order joint cumulant of $(u_{a_1, 0}, u_{a_2, h_2}, \dots, u_{a_j, h_j})$.

Lemma 4.4 follows from Theorem 4.1 and Corollary 5.2 of [Brüggemann et al. \(2016\)](#), and it shows the validity of the residual-based MBB for the OLS and the IRF estimators.

Lemma 4.4. *Suppose Assumptions 4.1–4.3 hold. If $l_b \rightarrow \infty$ such that $l_b^3 / T \rightarrow 0$ as $T \rightarrow \infty$, then*

$$\sup_{x \in \mathbb{R}^N} \left| Pr^* \left(\sqrt{T} \begin{bmatrix} \hat{\beta}^* - \hat{\beta} \\ \hat{\sigma}^* - \hat{\sigma} \end{bmatrix} \leq x \right) - Pr \left(\sqrt{T} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\sigma} - \sigma \end{bmatrix} \leq x \right) \right| \rightarrow 0$$

and

$$\sup_{x \in \mathbb{R}} \left| Pr^* \left(\sqrt{T} \left(\widehat{IRF}_i^{(s)*} - \widehat{IRF}_i^{(s)} \right) \leq x \right) - Pr \left(\sqrt{T} \left(\widehat{IRF}_i^{(s)} - IRF_i^{(s)} \right) \leq x \right) \right| \rightarrow 0$$

in probability, where Pr^* is the probability measure induced by the residual-based MBB and $\tilde{N} = pn^2 + \hat{n}$.

The following theorem provides the validity of the residual-based MBB for the QR estimator.

Theorem 4.1. *Suppose Assumptions 4.1–4.3 and A.1–A.5 hold. If $l_b \rightarrow \infty$ such that $l_b^3 / T \rightarrow 0$ as $T \rightarrow \infty$, then*

$$\sup_{x \in \mathbb{R}^{\tilde{M}}} \left| Pr^* \left(\sqrt{T} \left(\hat{\Theta}_{i,\tau}^* - \hat{\Theta}_{i,\tau} \right) \leq x \right) - Pr \left(\sqrt{T} \left(\hat{\Theta}_{i,\tau} - \Theta_{i,\tau} \right) \leq x \right) \right| \rightarrow 0$$

in probability, where $\tilde{M} = 2(1+n+2l)$.

As $QIRF_{i,\tau}^{(s)}$ is continuously differentiable functions of β, σ and $\Theta_{i,\tau}$, the asymptotic validity of the bootstrap extends to the QIRF estimator from Lemma 4.4 and Theorem 4.1. The following corollary summarizes the result.

Corollary 4.1. Suppose Assumptions 4.1–4.3 and A.1–A.5 hold. If $l_b \rightarrow \infty$ such that $l_b^3 / T \rightarrow 0$ as $T \rightarrow \infty$, then

$$\sup_{x \in \mathbb{R}} \left| Pr^* \left(\sqrt{T} \left(\widehat{QIRF}_{i,\tau}^{(s)*} - \widehat{QIRF}_{i,\tau}^{(s)} \right) \leq x \right) - Pr \left(\sqrt{T} \left(\widehat{QIRF}_{i,\tau}^{(s)} - QIRF_{i,\tau}^{(s)} \right) \leq x \right) \right| \rightarrow 0$$

in probability.

5. QUANTILE IMPULSE RESPONSE ANALYSIS OF THE US ECONOMY

Monetary policy has been one of the most heavily studied topics in macroeconomics, and the IRF is the main tool for evaluating its policy implications. Certain financial conditions indices have received much attention recently for explaining economic fluctuations.¹⁹ The impact of financial shocks on the whole economy is now considered dominant after the 2007–2009 financial crisis.

In this section, we apply the QIRF to US macroeconomic and financial data and investigate their dynamic quantile responses to monetary policy and financial shocks. In particular, we provide the dynamic responses of Growth-at-Risk (5% quantile of CFNAI) using the QIRF.²⁰ We also examine the quantile responses of macroeconomic variables in a distress scenario of financial instability.

5.1. Data

The variables under study are the CFNAI, the inflation rate (CPI), the federal funds rate (FFR), and the NFCI. The CFNAI is a monthly index for US economic activity, released by the Federal Reserve Bank of Chicago. The index is a weighted average of 85 indicators of national economic activity and captures movements of the GDP growth well. For our sample period (1971Q1–2019Q4), the correlation between GDP growth rate and CFNAI is 0.73. The NFCI is a weekly index describing US financial conditions in the money market, debt and equity markets, and traditional and shadow banking systems. The index, also released by the Federal Reserve Bank of Chicago, is a weighted average of 105 indicators of national financial activity.²¹

For the sample period from January 1971 to December 2019, we use monthly data of the four variables. We measure inflation rate as the log difference of CPI, multiplied by 100. For the FFR between 2009 and 2015 during which it reached the zero lower bound, we use the shadow FFR estimated by Wu and Xia (2016).²² For the NFCI, we use its monthly average. All data are from Federal Reserve Economic Data (FRED).

We estimate models (1) and (3) with the four variables. Following the Bayesian information criterion (BIC), we first estimate a VAR(3) model. The VAR model is stable since the largest eigenvalue of the companion matrix is strictly less than one. Then, we estimate the QR model of lag order 1 for lagged conditional quantile terms (i.e., $l = 1$ in (2)). For the residual-based MBB, we use a block length of $l_b = 25$ following Jentsch and Lunsford (2019).

5.2. Estimated Conditional Quantiles

Prior to investigating QIRFs, we examine how the conditional quantiles of the four variables evolved over the sample period. Fig. 1 illustrates the estimated conditional 5% and 95% quantiles over 2001–2015. While the two quantiles co-move in each variable, they do not fluctuate in the same way. Due to their heterogeneous movements, the distance between the lower and upper quantiles (the dispersion of distribution) changes over time. For instance, the conditional distribution was more dispersed during the GFC compared to other periods, in all variables.

However, the degree of heterogeneity in the quantile movements varies among the variables. In particular, dissimilar dynamics of the downside and upside risks are pronounced in the CFNAI. Between February 2007 and 2009, for example, the conditional 5% quantile decreased by 4.2, but the 95% quantile decreased by 3.5, which illustrates heterogeneous tail risk dynamics in economic activity. For other variables, the movements of the tail quantiles are not heterogeneous as much as in the CFNAI.

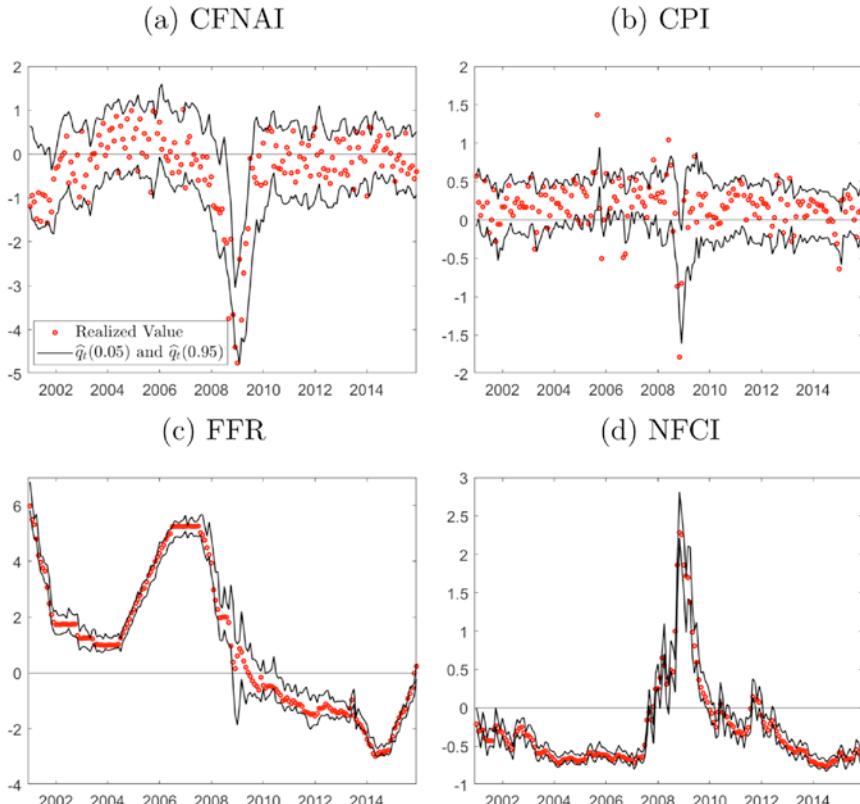


Fig. 1. The Estimated Conditional 5% and 95% Quantiles over 2001–2015.

The summary statics for the estimated conditional 5% and 95% quantiles in [Table 1](#) also highlight that dynamics of the downside and upside risks are the most disparate in the CFNAI. The correlation coefficient between the two quantiles is relatively smaller for the CFNAI. Moreover, its left tail shows much greater time variation than its right tail. Measured by standard deviation, the variation of the 5% quantile is $1.4 \left(= \frac{0.97}{0.69} \right)$ times as large as that of the 95% quantile. These results suggest substantial heterogeneity in the quantile response of the CFNAI.

For the CPI, the correlation coefficient between its 5% and 95% quantiles is smaller too: 0.63. But, their time variations are not as substantially different as in the CFNAI. The variation of the 95% quantile is $1.1 \left(= \frac{0.26}{0.23} \right)$ times that of the 5% quantile. Accordingly, a certain degree of heterogeneity is expected in the quantile response of CPI, but not as much as that of the CFNAI.

On the contrary, the quantile response of the financial variables (FFR and NFCI) is expected to be much less heterogeneous. As shown in [Fig. 1\(c\)](#) and [1\(d\)](#), the correlation between their conditional 5% and 95% quantiles is very strong: their correlation coefficients are close to one. Their time variations of the left and right tails are not substantially different as in the CFNAI. The standard deviation of the 95% quantile is 1.2 times as large as the 5% quantile standard deviation in the variables.

5.3. Quantile Impulse Response Analysis

We now construct QIRFs based on the model estimates as in [\(7\)](#). Since our QR model is a reduced form, we use a mean-based VAR model with Cholesky restrictions to identify a structural shock. Under the recursive identification, a variable is affected by the contemporaneous shocks to other variables if the variable is ordered after them, but not affected if ordered before them. Thus, slowly moving variables are ordered before fast-responding variables. The ordering of our variables (CFNAI, CPI, FFR, and NFCI, standard in the literature) implies that the NFCI instantly responds to all structural shocks. But economic activity (CFNAI) does not contemporaneously respond to shocks other than a shock to itself.

5.3.1. QIRF to a Monetary Policy Shock

First, we present how the conditional quantile of the variables responds to a monetary policy shock. [Fig. 2](#) is the QIRF to an expansionary monetary policy shock (-25bp) at five quantiles (5%, 16%, 50%, 84%, and 95%) as well as the IRF.²³

Against the expansionary monetary policy shock, the median and mean of CFNAI show similar dynamics. However, the QIRF clearly illustrates that its tail responses are highly heterogeneous. The monetary policy shock significantly increases the 5% quantile (i.e., the shock effectively reduces downside risk to growth). On the contrary, the response of the 95% quantile is close to zero implying upside risks are much less affected. These results highlight the monetary policy shock not only shifts the economic activity distribution but also significantly changes its shape, which cannot be learned from the conventional IRF.

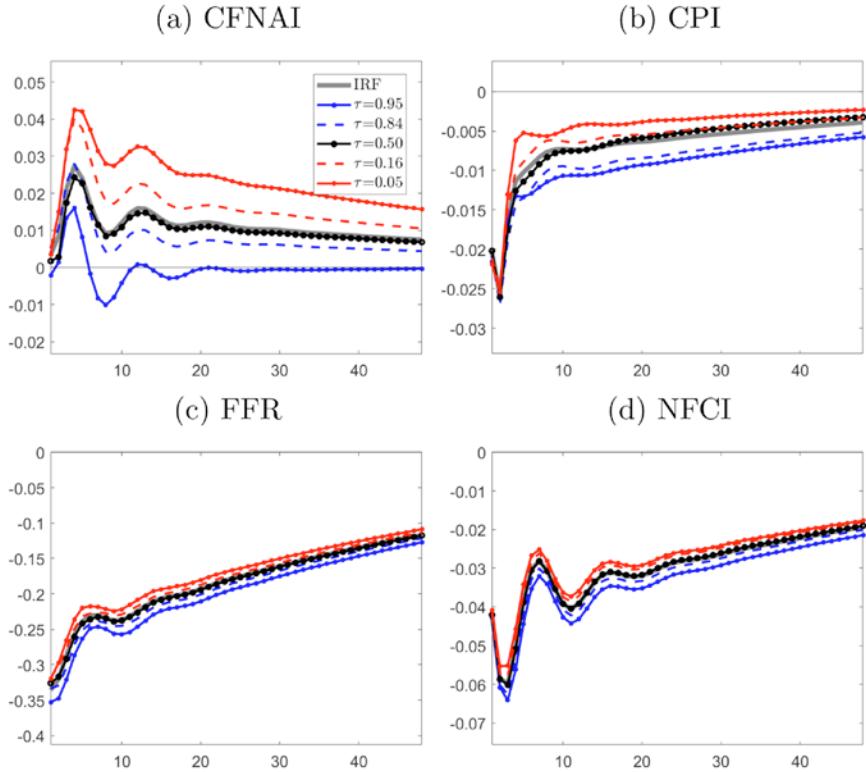


Fig. 2. QIRF to an Expansionary Monetary Policy Shock (-25bp).

Since quantiles at the left tail are more responsive, the volatility of economic activity decreases while its location shifts to the right. Loria et al. (2019) also find that a monetary policy shock affects lower quantiles of GDP growth more than its upper quantiles using local projections. Unlike their result, however, our QIRF does not show the reverse effect of the shock at longer horizons.²⁴

In addition, the 5% quantile of economic activity displays a more persistent response to a monetary policy shock than the mean IRF. It takes 26 months for the 5% quantile response of CFNAI to dissipate by half from its peak, while it takes only 11 months for the mean response. These dynamics illustrate that the effect of a monetary policy shock is stronger on Growth-at-Risk in terms of persistence as well as magnitude.

For CPI, the quantile response, as in the mean response, exhibits the so-called price puzzle: a decrease (increase) in the price level in response to an expansionary (contractionary) monetary policy shock.²⁵ Though the QIRF reveals certain amount of heterogeneity across quantiles, the difference between the 5% and 95% quantiles is small (less than 0.1 percentage point in annualized rate).

An expansionary monetary policy shock leads to looser financial conditions (a decrease in the NFCI) and has persistent effects on the FFR. The shock reduces volatility of the FFR and the NFCI, as the upper quantiles decline slightly more than the lower quantiles. However, the financial variables show much more homogeneous quantile responses, especially compared to economic activity. We interpret their QIRFs as relatively stable responses of the Fed and financial markets: their response to a one-off monetary shock does not significantly increase tail risks to the FFR and financial conditions.

5.3.2. QIRF to a Financial Shock

[Fig. 3](#) plots the QIRF and IRF to a financial shock. As in the case of a monetary policy shock, the response of the CFNAI is highly heterogeneous across quantiles. An adverse financial shock shifts the conditional distribution of economic activity to the left, but the left tail quantiles decrease substantially more than the right tail. This empirical result is in line with [Adrian et al. \(2019\)](#): economic growth is vulnerable to deteriorating financial conditions. They argue that downside

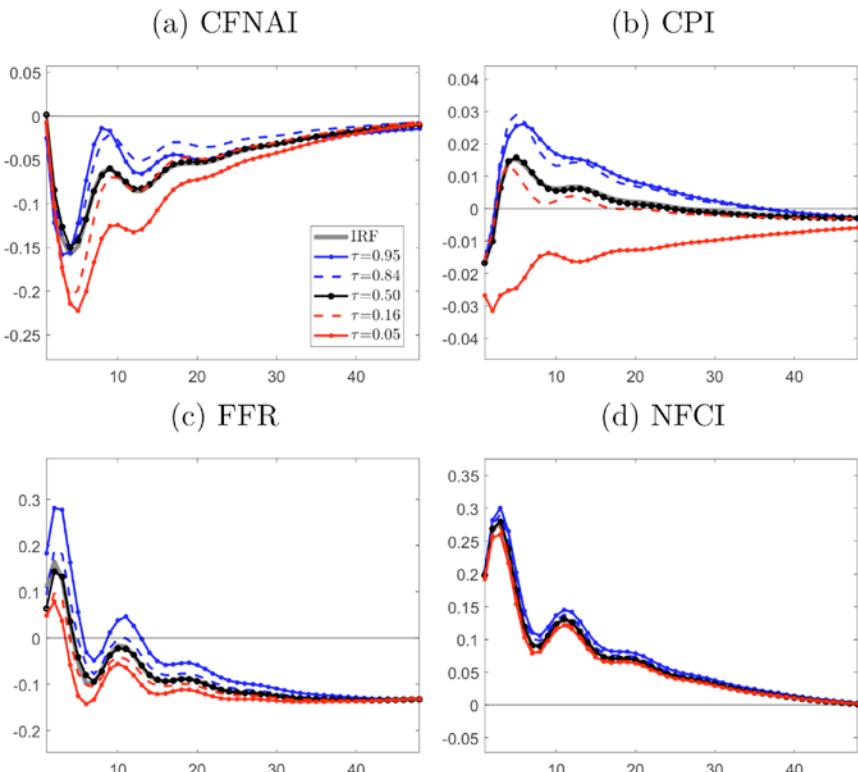


Fig. 3. QIRF to a Financial Shock (One Standard Deviation Shock).

Table 1. Summary Statistics for the Estimated Conditional 5% and 95% Quantiles.

	CFNAI	CPI	FFR	NFCI
Correlation coefficient between $\hat{q}_{y_t}(0.05)$ and $\hat{q}_{y_t}(0.95)$	0.71	0.63	0.99	0.998
Standard deviation of $\hat{q}_{y_t}(0.05)$	0.97	0.23	3.96	0.92
Standard deviation of $\hat{q}_{y_t}(0.95)$	0.69	0.26	4.57	1.10

GDP vulnerability can be explained by amplification mechanisms in the financial sector, such as the feedback loops mechanism by [Brunnermeier and Sannikov \(2014\)](#). While [Adrian et al. \(2019\)](#) investigate the impact of financial conditions on the following period's Growth-at-Risk, our approach goes further describing the evolution of the quantile response over time.

Meanwhile, the impact of a financial shock on Growth-at-Risk is not as persistent as that of a monetary policy shock. The magnitudes of the IRF and 5% QIRF decay by half from its trough in 11 months and 10 months, respectively. This suggests a financial shock has a relatively acute effect on downside risks to economic activity.

The CPI also shows a certain degree of heterogeneity in its quantile response to a financial shock: its 95% quantiles increases more than the mean response, but its 5% quantile decreases. While the shock increases upside and downside risks to inflation, the magnitude of the quantile response is not so significant (between -0.4% and $+0.3\%$ in annualized rate) as in the case of a monetary policy shock.

For the FFR, the 95% quantile increases much more than other quantiles in response to a financial shock. The 95% quantile increases by up to 28 basis points and stays positive for 5 months, and the IRF also shows positive response for 4 months after the shock. Such responses are counterintuitive as accommodative monetary policies are expected against a financial shock. This upside risk to the FFR seems to be driven by observations in the 1970s. With a sample period from January 1981 to December 2019, the initial increases in both the 95% quantile and mean responses dramatically disappear. These empirical results may attribute to the change in monetary policy practice or the change in how the other variables respond to shocks.²⁶

Fig. 3(d) suggests that financial conditions deteriorate (increases in the NFCI) further for a few months after the initial financial shock. For all of the five quantiles considered, the response of NFCI is quite homogeneous along the mean response. That is, a financial shock shifts the distribution of the NFCI with little change in its volatility.

5.4. Growth-at-Risk Dynamics During the Global Financial Crisis

In this section, we examine how severely the conditional quantile of economic activity was affected by a series of shocks during a particular historical episode: the 2007–2009 GFC. We pay close attention to the conditional 5% quantile of the CFNAI, considered as *Growth-at-Risk* in this chapter. The dynamic response of the quantile to a one-off shock is described by the QIRF as in Section 5.3. We now investigate the cumulative effects of a set of shocks on the quantile using QIRFs.

We provide an answer to the following question: what are the impacts of financial and monetary policy shocks concerning the GFC on the downside and upside risks to growth? As seen in the previous section, the lower quantiles of economic activity are much more responsive than the median or upper quantiles. Thus, quantitative assessment of the downside risks during the recession period is of great importance. In the following exercises, we first study the impacts of financial shocks on the risks during the financial distress period of August 2007–June 2009. We then examine the effects of ensuing unconventional monetary policy measures using the identified monetary policy shocks from July 2009 to December 2015.

The assessment is carried out in a similar way to historical decomposition in the VAR analysis.²⁷ Suppose, for example, we want to quantify the effects of structural shocks from time 1 to T , $\{\epsilon_t\}_{t=1}^T$, on the quantile response. Let $QIRF_{i,\tau}^{(s)} | \epsilon_i$ denote the τ -quantile response of the i -th variable to ϵ_i at horizon s . The QIRF estimates quantile dynamics against a shock at a specific time. To construct cumulative impacts of $\{\epsilon_t\}_{t=1}^T$ on quantiles, we aggregate quantile responses of each of the shocks accounting for their dynamics. That is, the cumulative impacts of $\{\epsilon_t\}_{t=1}^T$ on the τ -quantile of the variable i at time s is calculated as

$$\begin{cases} \sum_{t=1}^{s-1} QIRF_{i,\tau}^{(s-t)} | \epsilon_i, & \text{for } s \leq T, \\ \sum_{t=1}^T QIRF_{i,\tau}^{(s-t)} | \epsilon_i, & \text{for } s > T. \end{cases}$$

For estimation, we replace $QIRF_{i,\tau}^{(s-t)}$ and ϵ_i with their respective estimators, $\widehat{QIRF}_{i,\tau}^{(s-t)}$ and $\widehat{\epsilon}_i$.

The cumulative impacts of financial shocks from August 2007 to June 2009 on the conditional quantile of the CFNAI are illustrated in Fig. 4(a).²⁸ Financial shocks during the period substantially increased the downside risk, but the upside risk was mildly affected. Over 2008–2010, the financial shocks decreased the 5% quantile of the CFNAI by 1.5 on average. However, the impacts of those shocks on the 95% quantile were much less; the conditional 95% quantile was lowered by 0.8 on average over the same period.

We then answer the following counterfactual question: what would have happened if the financial shocks were shut down? Fig. 4(b) and 4(c) describe the counterfactual paths of the conditional 5% and 95% quantiles based on the cumulative impacts. The figures highlight that the increase in downside risks during the GFC was mainly attributable to financial shocks. The 5% quantile is much lower than the counterfactual 5% quantile. In the absence of the financial shocks, the downside risks would have been moderate. The average of the counterfactual 5% quantile over 2008–2009, −1.3, is a little less than the average of the 5% quantile over 1971–2007, −1.0. Without the financial shocks, the 95% quantile would have been higher. But the difference between the 95% quantile and its counterfactual is narrower compared to the case of the 5% quantile, suggesting asymmetric impact of the financial shocks on the economic activity distribution.

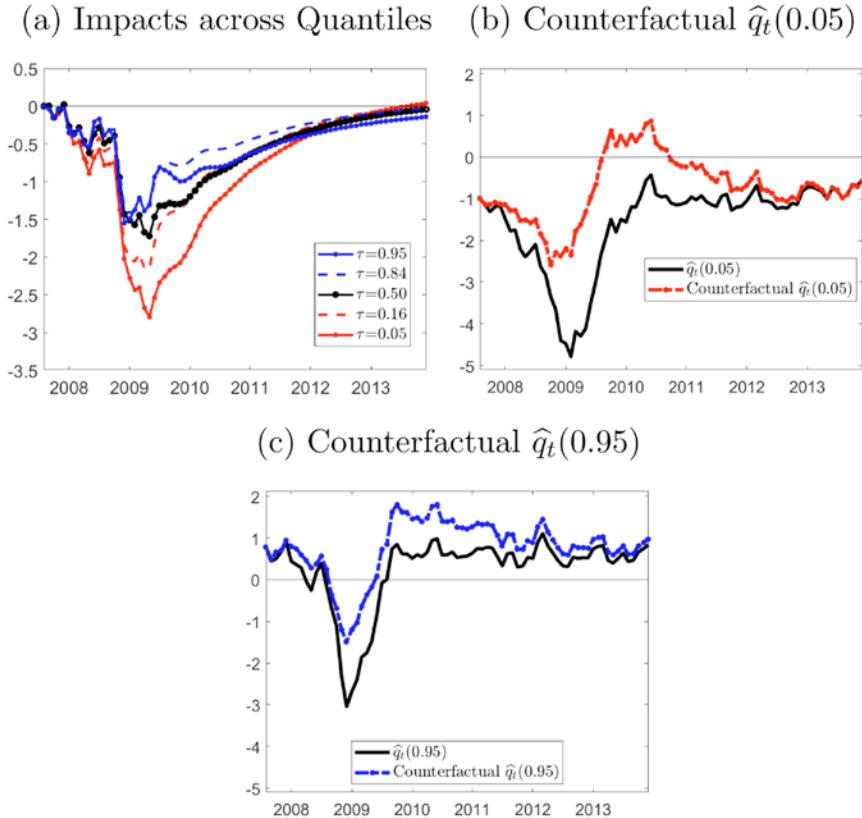


Fig. 4. The Impacts of Financial Shocks (August 2007–June 2009) on CFNAI.
Note: The counterfactual path describes the quantile path of CFNAI if the financial shocks were shut down.

We perform the same exercise to quantify the effects of unconventional monetary policy measures implemented by the Fed in response to the financial crisis. Following Wu and Xia (2016), structural shocks to the FFR from July 2009 to December 2015 are used for assessment of the monetary policy.²⁹ Fig. 5(a) demonstrates the cumulative impacts of those monetary policy shocks on the quantile of the CFNAI. The figure emphasizes the effectiveness of unconventional monetary policy for reducing the downside risks to growth. Over 2010–2015, the monetary policy increased its 5% quantile by 0.4, on average, but its 95% quantile was hardly affected.

The counterfactual paths in Fig. 5(b) and 5(c) describe the heterogeneous effects of the unconventional measures on economic growth. Without the measures, downside risks would have been higher: the averages of the 5% quantile and its counterfactual over 2011–2015 are −0.9 and −1.3, respectively. The unconventional

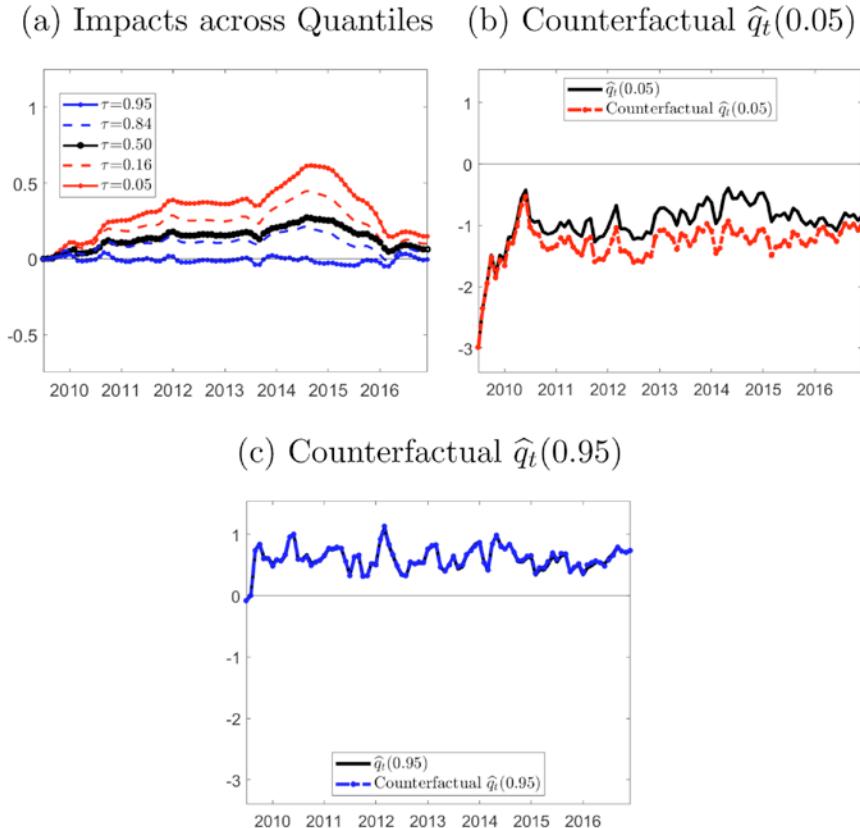


Fig. 5. The Effects of Unconventional Monetary Policy (July 2009–December 2015) on CFNAI. Note: The counterfactual path describes the quantile path of CFNAI if the monetary policy shocks were shut down.

monetary policy consistently reduced the downside risks over the period. In contrast, the 95% quantile and its counterfactual are almost indistinguishable.

5.5. Quantile Responses Under a Hypothetical Distress Scenario

Since extreme shocks leave long-lasting scars on the economy, policy makers require effective tools for testing macroeconomic resilience in a stress scenario. In this section, we conduct a hypothetical analysis to examine the quantile response in a distress scenario where a series of unfavorable tail events follow an initial shock.

The QIRF defined in Section 3.1 measures the impact of a shock assuming realizations of endogenous variables after the shock are their conditional mean. Here, we replace the conditional mean-time path with a stress scenario in which unfavorable tail events for endogenous variables materialize, then construct the

quantile responses for a stress testing. This exercise is comparable to the stress testing of Chavleishvili and Manganelli (2019) and the QIRFs of Montes-Rojas (2019) in that the impacts of consecutive realizations of tail events are studied.

In this exercise, we examine how the economy responds to a distress scenario initiated by a financial shock. In the scenario, an initial financial shock (one standard deviation shock) is followed by realizations of the NFCI at its conditional 95% quantile for six months. After that, realizations of its conditional mean are assumed for the NFCI. For variables other than the NFCI, realizations of the conditional mean are assumed after the initial shock. The distress scenario describes a rapid deterioration of financial conditions. As in the QIRF, the quantile response under this scenario can be expressed in a recursive manner, and its derivation is relegated to Online Appendix Section C.

Fig. 6 describes the mean and quantile responses under the hypothetical scenario. First of all, NFCI displays non-stationary (locally explosive) behavior

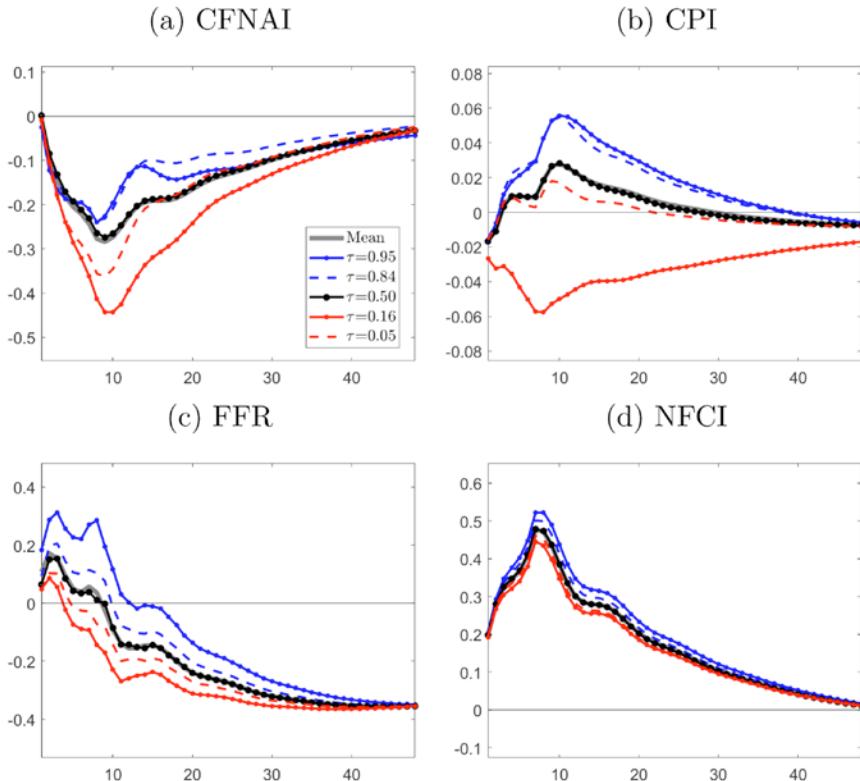


Fig. 6. Responses of Variables Under the Distress Scenario. Notes: After a shock of one standard deviation to the NFCI, realizations of the conditional mean are assumed for variables other than the NFCI. For the NFCI, realizations of the conditional 95% follow for the first 6 months after the shock. Afterwards, realizations of the conditional mean are assumed for the variable.

when it stays at its conditional 95% quantile. The distress scenario substantially shifts the distribution of NFCI to the right; financial conditions deteriorate rapidly. The mean of NFCI increases by up to 0.48. While a one-time financial shock increases its mean by up to 0.28 as in Fig. 3(d), the consecutive tail events lead to acute financial distress.

As a result of the financial instability, the economy suffers a severe downturn with substantial downside risk. The economic activity distribution shifts to the left greatly, and its left tail decreases much more than the median or upper quantiles. The 5% quantile of CFNAI decreases by up to 0.44, whereas its mean and 95% quantile decline by 0.28 and 0.24, respectively. These results may suggest that a financial crisis can develop in a short period of time.

Under the scenario, the initial response of the FFR displays certain amount of heterogeneity across quantiles. But, its response becomes more homogeneous at longer horizons; the considered five quantiles of FFR decrease by more than 30 basis points three years after the initial shock.

6. CONCLUSION

This chapter studies QIRF theory and its applications in macroeconomics and finance. Our QR model provides a multi-equation system with autoregressive specifications accounting for important dynamics of distributional evolution. The QIRF complements the conventional IRF providing a more complete shock-response mechanism.

The comprehensive QIRF analysis of the US economy provides evidence of a strong heterogeneity in the responses of economic activity across its distribution. Against monetary policy and financial shocks, the downside risks to growth are more responsive than the median or upside risks. We also quantitatively assessed the evolution of macroeconomic tail risks during the 2007–2009 Global Financial Crisis and in a hypothetical scenario where financial conditions rapidly deteriorate. Considering tremendous implications of extreme events, such as market booms and crashes, our QR model and QIRF provide useful tools for dynamic risk management and policy analysis.

NOTES

1. CFNAI is a monthly index for US economic activity. Section 5.1 explains the index in detail.

2. NFCI is an index for US financial conditions measuring the tightness of financial markets. A high value of NFCI represents tighter financial conditions. See Section 5.1 for more details about the index.

3. See also [Xiao \(2012\)](#) and [Linton and Xiao \(2018\)](#) for recent advances in time series QR models and their applications.

4. The QR methods are also used to study heterogeneous dynamics for various macroeconomic and financial variables. [Chevapatrakul et al. \(2009\)](#) evaluate the quantile response of interest rates to inflation and the output gap. [Galvao et al. \(2013\)](#) study the effects of income and interest rates on UK house price returns across quantiles. [Mumtaz and Surico \(2015\)](#) investigate the dynamic relationship between interest rates and the conditional quantile of consumption.

5. They derive the quantile response employing *the law of iterated quantiles* (p. 15).

6. In the literature of VAR models with SV, volatility dynamics are usually modeled as geometric random walks such as $\log \sigma_t = \log \sigma_{t-1} + \eta_t$, where η_t is a mean-zero stochastic error term. See, e.g., [Stock and Watson \(2002\)](#) and [Primiceri \(2005\)](#).

7. For the intuition behind the QR model, let us consider the following linear autoregressive conditional heteroscedasticity (ARCH) volatility dynamics: $\sigma_t = \gamma + \alpha y_{t-1} + \beta \sigma_{t-1}$. Then, the quantile dynamics belongs to the specification in (2):

$$Q_{y_t}(\tau | \mathcal{F}_{t-1}) = c_\tau + a_\tau y_{t-1} + b_\tau (Q_{y_{t-1}}(\tau_U | \mathcal{F}_{t-2}) - Q_{y_{t-1}}(\tau_L | \mathcal{F}_{t-2}))$$

where $c_\tau = F_\epsilon^{-1}(\tau)\gamma$, $a_\tau = F_\epsilon^{-1}(\tau)\alpha$, and $b_\tau = F_\epsilon^{-1}(\tau)\beta / (F_\epsilon^{-1}(\tau_H) - F_\epsilon^{-1}(\tau_L))$. In addition, our QR model accounts for quantile persistence arises from asymmetric dynamics (or skewness dynamics). Asymmetric dynamics have not been theoretically examined as much as volatility dynamics, thus it is not easy to provide such an example in explicit form. However, a growing number of literature are suggesting empirical evidence of asymmetric dynamics. Hence, we incorporate quantile persistence effect, $d_{k,i,\tau} Q_{u_{i,k}}(\tau | \mathcal{F}_{t-k-1})$, in our QR model.

8. Those quantiles are less volatile than tail quantiles and far enough apart from each other for a volatility measure. Moreover, the [16%, 84%] interval is commonly provided for posterior probability bands in Bayesian inference, and the interval covers approximately two standard deviations around the mean in the case of a Normal distribution.

9. Although quantile interactions across variables are not accounted for in the model, the cross-sectional tail dependence can be examined indirectly. [Adrian and Brunnermeier \(2016\)](#) measure such tail co-dependence using CoVaR defined as changes in tail risk conditional on another tail event relative to the median state. In a similar spirit, conditional on one variable in particular tail events, responses of another variable's tail risk are investigated in Section 5.5.

10. As explained in Section 2, $Q_{u_i}(\tau | \mathcal{F}_{t-1})$ can be expressed as a linear function of the past innovations. Accordingly, the QIRF could be defined as $\frac{\partial Q_{u_i}(\tau | \mathcal{F}_{t-1})}{\partial \epsilon'_i}$ for a structural shock ϵ_i , in the same way as the IRF is defined as $\frac{\partial y_{t+s}}{\partial \epsilon'_i}$. Under such a definition, the resulting QIRF still has the same representation as (5). We use Definition 3.1 since it provides a more intuitive interpretation of the QIRF.

11. WKM acknowledge that the pseudo-QIRF ignores the dynamic evolution of distribution. Han et al. (2022) discuss and evaluate the performance of the pseudo-QIRF.

12. In his empirical application, he estimates QIRFs with three variables: the output gap, inflation, and the Fed Funds rate. For the construction of the quantile response of output gap at $\tau = 0.1$, for example, he considers the following time path. After a shock, realizations of the output gap are assumed to be at its conditional 10% quantile continuously, but realizations of inflation and the Fed Funds rate. are assumed to be their conditional median.

13. There are two main differences between the QR model of WKM and ours. First, we decompose $\{y_t\}$ into its conditional mean and the innovations, then the QR is used to explain the conditional quantile of the latter. However, their QR model describes the conditional quantile of $\{y_t\}$ directly without such decomposition. Second, as explained in Remark 2.2, our QR model does not allow the quantile dependence across variables. On the contrary, their model incorporates such codependence across variables.

14. In this section, we assume the intercept of the VAR model is zero ($A_0 = 0$) for notational simplicity. The QIRF in (5) does not depend on A_0 .

15. Instead of the Cholesky restriction, an alternative identification strategy can be used with the identification restriction $C_{\Theta_0^{-1}} \text{vec}(\Theta_0^{-1}) = c_{\Theta_0^{-1}}$ where $C_{\Theta_0^{-1}}$ is an $\tilde{n} \times n^2$ selection matrix and $c_{\Theta_0^{-1}}$ is a suitable $\tilde{n} \times 1$ fixed vector. For details, see of [Lütkepohl \(2005\)](#).

16. Recall that $\hat{\mathbf{u}}_t = \mathbf{y}_t - \sum_{k=1}^p \hat{A}_k \mathbf{y}_{t-k}$ is used in Steps 2 and 3 of the estimation procedure.
17. $\mathbf{Q}_{i,t,\tau}$ depends on the density function $f_{i,t,\tau}(0)$. See, e.g., Koenker (1994, 2005) for details about the asymptotic inference in QRs.
18. The residual-based MBB applies the block bootstrap to residuals after fitting a model to capture a weak dependence structure in time series data. See, e.g., Paparoditis and Politis (2003), Ioannidis (2005), and Jentsch et al. (2015).
19. See, e.g., Brave and Butters (2011), Matheson (2012), and Koop and Korobilis (2014).
20. In the spirit of Value-at-Risk in finance, IMF (2017) introduced Growth-at-Risk to measure macrofinancial risks to economic activity.
21. More details about the two indices are available at <https://www.chicagofed.org/publications/cfnai/index> (CFNAI) and <https://www.chicagofed.org/publications/nfci/index> (NFCI).
22. The data are available at <https://sites.google.com/view/jingcynthiawu/shadow-rates>.
23. In the Appendix section, bootstrap confidence intervals of the QIRF to monetary policy and financial shocks are provided in Figs. A1 and A2, respectively.
24. The estimates of Loria et al. (2019) suggest the effect of a monetary policy shock reverse in the medium term. For example, they claim that a contractionary monetary policy shock has a positive effect on the GDP growth after 10–15 quarters.
25. This price puzzle is attributable to identification of structural shocks. See, e.g., Sims (1992), Christiano et al. (1999), and Hanson (2004) for further discussion of the price puzzle.
26. See, e.g., Primiceri (2005) and Sims and Zha (2006) for a discussion of the change in monetary policy rules. We leave more rigorous investigation of these causal explanations to future research.
27. See Kilian and Lütkepohl (2017) for details about historical decomposition in the VAR model.
28. In August 2007, BNP Paribas halted redemptions on three investment funds because it could not value their holdings, which marked the start of the financial crisis. The National Bureau of Economic Research (NBER) identified June 2009 as the end of the recession associated with the financial crisis.
29. While Wu and Xia (2016) study the effects of unconventional policy measures using monetary policy shocks from July 2009 to December 2013, we extend the period to December 2015, the end of the zero lower bound period.

REFERENCES

- Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109, 1263–1289.
- Adrian, T., & Brunnermeier, M. K. (2016). CoVaR. *The American Economic Review*, 106, 1705–1741.
- Adrian, T., Grinberg, F., Liang, N., & Malik, S. (2018). The term structure of Growth-at-Risk [IMF Working Paper].
- Brave, S. A., & Butters, R. A. (2011). Monitoring financial stability: A financial conditions index approach. *Economic Perspectives, Federal Reserve Bank of Chicago*, 35(1), 22–43.
- Brüggemann, R., Jentsch, C., & Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, 191, 69–85.
- Brunnermeier, M. K., & Sannikov, Y. (2014). A macroeconomic model with a financial sector. *American Economic Review*, 104, 379–421.
- Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics*, 214(1), 274–294.
- Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, 192(1), 152–167.
- Chang, Y., Kim, S., & Park, J. Y. (2021a). *Effects of macroeconomic shocks on income distribution* [Working paper].

- Chang, Y., Miller, Z., & Park, J. Y. (2021b). *What drives temperature anomalies? An econometric analysis using functional autoregression* [Working paper].
- Chavleishvili, S., & Manganelli, S. (2019). *Forecasting and stress testing with quantile vector autoregression* [ECB Working Paper Series No. 2330].
- Chevapatrakul, T., Kim, T.-H., & Mizen, P. (2009). The Taylor principle and monetary policy approaching a zero bound on nominal rates: Quantile regression results for the United States and Japan. *Journal of Money, Credit and Banking*, 41, 1705–1723.
- Christiano, L. J., Eichenbaum, M., & Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In J. B. Taylor & M. Woodford (Eds.), *Handbook of macroeconomics* (Vol. 1, pp. 65–148). Elsevier.
- Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22, 367–381.
- Galvao, A. F., Montes-Rojas, G., & Park, S. Y. (2013). Quantile autoregressive distributed lag model with an application to house price returns. *Oxford Bulletin of Economics and Statistics*, 75, 307–321.
- Han, H., Jung, W., & Lee, J. H. (2022). Estimation and inference of quantile impulse response functions by local projections: With applications to VaR dynamics. *Journal of Financial Econometrics* (forthcoming). doi.org/10.1093/jjfinec/nbac026
- Han, H., Linton, O., Oka, T., & Whang, Y.-J. (2016). The cross-quantilogram: Measuring quantile dependence and testing directional predictability between time series. *Journal of Econometrics*, 193, 251–270.
- Hanson, M. S. (2004). The “price puzzle” reconsidered. *Journal of Monetary Economics*, 51, 1385–1413.
- IMF. (2017). Financial conditions and growth at risk. In *Global financial stability report (October)* (Chapter 3). International Monetary Fund.
- Ioannidis, E. E. (2005). Residual-based block bootstrap unit root testing in the presence of trend breaks. *The Econometrics Journal*, 8, 323–351.
- Jentsch, C., & Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Comment. *American Economic Review*, 109(7), 2655–2678.
- Jentsch, C., Politis, D. N., & Paparoditis, E. (2015). Block bootstrap theory for multivariate integrated and cointegrated processes. *Journal of Time Series Analysis*, 36, 416–441.
- Jordà, Ò. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review*, 95, 161–182.
- Jung, W., & Lee, J. H. (2022). Online Supplement to “Quantile Impulse Response Analysis with Applications in Macroeconomics and Finance”. Retrieved from <https://sites.google.com/view/whayoung/research>
- Kilian, L., & Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.
- Koenker, R. (1994). Confidence intervals for regression quantiles. In P. Mandl & M. Hušková (Eds.), *Asymptotic statistics* (pp. 349–359). Physica.
- Koenker, R. (2005). *Quantile regression*. Cambridge University Press.
- Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33–50.
- Koenker, R., & Xiao, Z. (2006). Quantile autoregression. *Journal of the American Statistical Association*, 101, 980–990.
- Koenker, R., & Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory*, 12, 793–813.
- Koop, G., & Korobilis, D. (2014). A new index of financial conditions. *European Economic Review*, 71, 101–116.
- Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74, 119–147.
- Künsch, H. R. (1989) The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17, 1217–1241.
- Lee, D. J., Kim, T.-H., & Mizen, P. (2021). Impulse response analysis in conditional quantile models with an application to monetary policy. *Journal of Economic Dynamics and Control*, 127, 104102.

- Li, G., Li, Y., & Tsai, C.-L. (2015). Quantile correlations and quantile autoregressive modeling. *Journal of the American Statistical Association*, 110, 246–261.
- Linton, O. & Xiao, Z. (2018). Quantile regression applications in finance. In R. Koenker, V. Chernozhukov, X. He, & L. Peng (Eds.), *Handbook of quantile regression* (Chapter 20, pp. 381–407). Chapman and Hall/CRC.
- Loria, F., Matthes, C., & Zhang, D. (2019). Assessing macroeconomic tail risk [Working Paper No. 19-10]. Federal Reserve Bank of Richmond.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag.
- Matheson, T. D. (2012). Financial conditions indexes for the United States and euro area. *Economics Letters*, 115, 441–446.
- Montes-Rojas, G. (2019). Multivariate quantile impulse response functions. *Journal of Time Series Analysis*, 40, 739–752.
- Mumtaz, H., & Surico, P. (2015). The transmission mechanism in good and bad times. *International Economic Review*, 56, 1237–1259.
- Paparoditis, E., & Politis, D. N. (2003). Residual-based block bootstrap for unit root testing. *Econometrica*, 71, 813–855.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72, 821–852.
- Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36, 975–1000.
- Sims, C. A., & Zha, T. (2006). Were there regime switches in US monetary policy? *American Economic Review*, 96, 54–81.
- Stock, J. H., & Watson, M. W. (2002). Has the business cycle changed and why? *NBER Macroeconomics Annual*, 17, 159–218.
- White, H., Kim, T.-H., & Manganelli, S. (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187, 169–188.
- Wu, J. C., & Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48, 253–291.
- Xiao, Z. (2009). Quantile cointegrating regression. *Journal of Econometrics*, 150, 248–260.
- Xiao, Z. (2012). Time series quantile regressions. In T. S. Rao, S. S. Rao, & C. R. Rao (Eds.), *Handbook of statistics: Time series analysis: Methods and applications* (Chapter 9, pp. 213–257). Elsevier.
- Xiao, Z., & Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, 104, 1696–1712.

APPENDIX

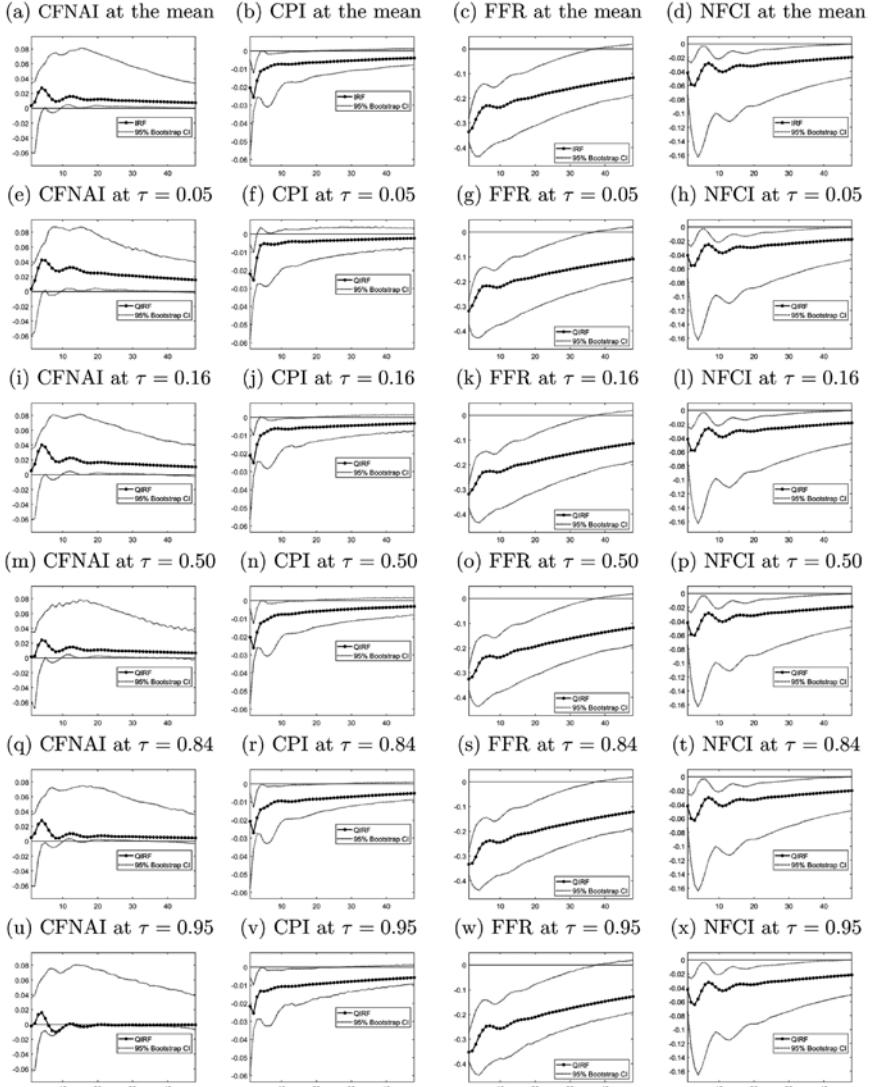


Fig. A1. QIRF and IRF to a Monetary Policy Shock (-25bp) with 95% Bootstrap Confidence Interval Using Residual-Based MBB, the Number of Bootstrap Draws Is 1,000 and the Block Length Is 25.

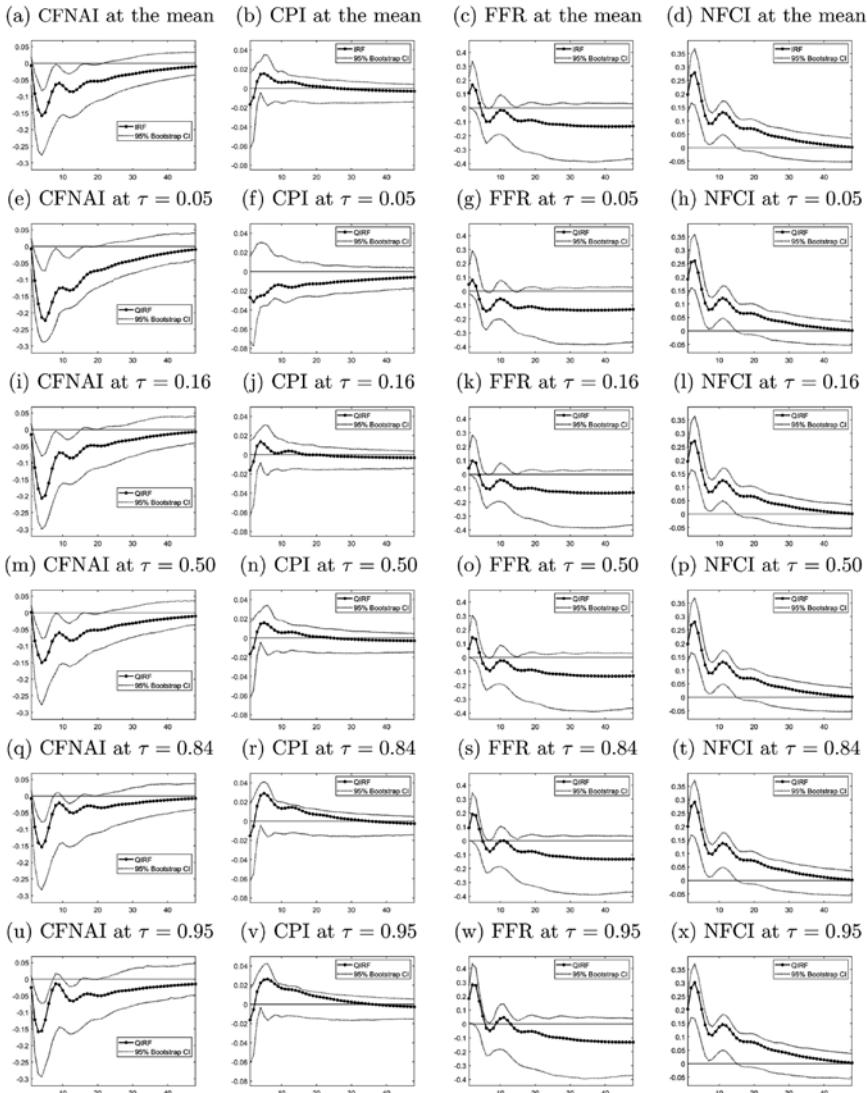


Fig. A2. QIRF to a Financial Shock (One Standard Deviation Shock) with 95% Bootstrap Confidence Interval Using Residual-Based MBB, the Number of Bootstrap Draws Is 1,000 and the Block Length Is 25.

This page intentionally left blank

CHAPTER 5

RISK NEUTRAL DENSITY ESTIMATION WITH A FUNCTIONAL LINEAR MODEL

Marine Carrasco and Idriss Tsafack

Economics Department, University of Montreal, Montreal, Quebec, Canada

ABSTRACT

This chapter proposes a nonparametric estimator of the risk neutral density (RND) based on cross-sectional European option prices. The authors recast the arbitrage-free equation for option pricing as a functional linear regression model where the regressor is a curve and the independent variable is a scalar corresponding to the option price. Then, the authors show that the RND can be viewed as the solution of an ill-posed integral equation. To estimate the RND, the authors use an iterative method called Landweber-Fridman (LF). Then, the authors establish the consistency and asymptotic normality of the estimated RND. These results can be used to construct a confidence interval around the curve. Finally, some Monte Carlo simulations and application to the S&P 500 options show that this method performs well compared to alternative methods.

Keywords Risk neutral density; option pricing; regularization; functional regression; Landweber-Fridman; nonparametric

1. INTRODUCTION

Estimating the RND has been an important topic for financial market participants and monetary policymakers. This tool is used for derivatives pricing, hedging and

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 133–157

Copyright © 2023 by Marine Carrasco and Idriss Tsafack

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:[10.1108/S0731-90532023000045B005](https://doi.org/10.1108/S0731-90532023000045B005)

market sentiment analysis. Additionally, it is used to analyze the trader's reaction to a potential shock in the financial market and predict the extreme shocks probabilities. For the policymaker, this tool is used to evaluate the effectiveness of monetary policies through direct observation of changes in investor's expectations and beliefs to future maturities (see [Souissi, 2017](#)).

This concept is also fundamental in the arbitrage-free asset pricing theory (see the textbooks by [Campbell et al., 1997](#) and [Cochrane, 2005](#)). Indeed, the RND is the density measure under which the price of each security in the market is equal to the expected value of its future payoff discounted back to the present given a risk-free interest rate. This means that for most of the securities in the market, the number of states of the economy could be very large, which in turn leads to situations where the number of potential future payoffs is very dense. Then the future payoffs can be considered as a continuous function of the potential states of the economy. On the same line, since the set of different states of the economy is very large, the RND can be viewed as a continuous function of the future payoffs which form is unknown. Therefore, these properties should be taken into account in the estimation procedure of the RND.

To address the estimation of the functional form of the RND, two main approaches have been suggested in the literature. The first approach relies on parametric modeling (see [Bahra, 1997](#); [Black & Scholes, 1973](#); [Figlewski, 2010](#)), which focus on considering a specific form for the RND and then estimate the related parameters. The most used distribution in this context is the log-normal density ([Jarrow & Rudd, 1982](#)) or a linear mixture of the log-normal distributions ([Bahra, 1997](#)). Unfortunately, this approach fails to capture all the features of the data (see [Ait-Sahalia & Lo, 2000](#)).

The second approach is the nonparametric technique. Several papers use the fact that the RND is the rescaled second derivative of the put (or call) price. Indeed, for a put option with strike price κ and time to maturity $T - t$, the payoff is $Z(s) = \max(\kappa - s, 0)$ and the put option price P satisfies

$$P(\kappa) = e^{-r(T-t)} \int_0^\kappa (\kappa - s) f(s) ds \quad (1)$$

where f is the RND and r is the risk-free rate. So differentiating (1) twice with respect to κ yields

$$f(S_T) = e^{r(T-t)} \frac{\partial^2 P}{\partial \kappa^2} \Big|_{\kappa=S_T} .$$

To recover the RND, one needs to estimate the second derivative of the put (or call) price. [Ait-Sahalia and Lo \(2000\)](#) suggest to estimate the function P by nonparametric kernel regression and then differentiate it twice to recover the RND. However, the resulting estimator is very volatile. Hence, the authors use a semiparametric approach based on Black and Scholes formula and realized volatility. To be fully nonparametric, [Ait-Sahalia and Duarte \(2003\)](#) propose to use a constrained locally polynomial kernel smoothing (see also [Grith et al., 2012](#)) which has the advantage to reduce the variance. Additionally, [Jackwerth and Rubinstein](#)

(1996) propose a method to extract RND by minimizing a criterion which penalizes unsmoothness. Garcia and Gençay (2000) propose to fit the option pricing function using neural network. While our chapter focuses on the estimation of the RND at a specific time t , Panigirtzoglou and Skiadopoulos (2004) and Bliss and Panigirtzoglou (2004) tackle the estimation of the dynamics of RND over time.

Bondarenko (2003) suggests to estimate the density with a nonparametric method called the positive convolution approximation (PosConv). Let $D^{-2}f$ be the second integral of f . For a cross-section of put prices P_i and assuming that the risk free rate r is 0, the density is selected to satisfy the following criterion

$$\min_{\hat{f}} \sum_{i=1}^n (P_i - D^{-2} \hat{f}(\kappa_i))^2$$

subject to $\hat{f}(\kappa) = \sum_j a_j \phi_h(\kappa - z_j)$ for a fine grid of values z_j and $\phi_h(x) = \frac{e^{-(x^2/h^2)}}{\sqrt{2\pi}h}$, $a_j \geq 0$, $\sum_j a_j = 1$. In other words, the density is approximated by a weighted sum of normal densities, i.e. a linear mixture of normal densities. One of the challenges is to select the number of normal densities to be included and the bandwidth h driving the variance of the normal distributions. The author uses a two-step data driven method to select these tuning parameters.

An alternative method consists in projecting on a polynomial basis. This idea was suggested by Shimko et al. (1993) to estimate the RND and predict the observed implied volatility. Rosenberg (1998) suggests to use a sigma-shaped polynomial technique for the same problem. Also, Yatchew and Härdle (2006) use a nonparametric least-squares with smoothness penalty and are able to impose various shape restrictions. Fengler (2009) uses smoothing spline. More recently, Vogt (2014) propose to approximate the RND with the squared of a series expansion based on Hermite polynomials and Kundu et al. (2018) use Bernstein polynomial basis. For an overview of the parametric and nonparametric methods used to estimate RND, we refer the reader to the book by Jondeau et al. (2010) and the recent review by Figlewski (2018).

To avoid imposing as much restrictions as possible, this chapter proposes to estimate nonparametrically the RND for European option pricing by exploiting the functional data analysis framework. The advantage of this approach is to connect the fundamental theory of asset pricing and the functional feature of the RND while realizing a good fitting performance. Additionally, the estimation does not rely on any latent form of the distribution for approximation. We deal with a functional linear regression model where the predictor is a function representing the future payoffs at the maturity and the response is a scalar representing the call and put price of the considered security. Then, the call and put prices are treated as a weighted sum of all the potential payoffs of the considered option at the time-to-maturity, with the weights represented by the density values.

The contribution of this chapter is to use the functional data analysis framework in order to estimate the RND estimation. This approach has not been explored in the extant literature. The estimation of the density function in this context leads to an ill-posed inverse problem. Taking the naive inverse would lead to an estimator which would not be consistent. To overcome this issue, we propose to use a regularization technique called the LF method. The LF technique is an iterative method used to solve an equation related to option pricing problem, in such a way that the estimation is made without inverting any operator in the procedure. The advantage of this method is that it relies neither on a basis projection nor on a kernel smoothing and helps to reduce the variability of the estimated density coming from the ill-posed problem. Another advantage is the possibility to derive directly the asymptotic normality results and confidence sets for the estimated density and the predictions of option prices. Compared to functional principal components commonly used in functional regression (see [Bosq, 2000](#)), LF method requires weaker conditions on the eigenvalues of the covariance operator. In particular, they do not need to be disjoint and multiple eigenvalues are allowed. As the resulting estimator is not a density, we then apply a density correction procedure in order to obtain a nonnegative function which integrates to one. We establish the consistency and the asymptotic normality of the estimated RND. Finally, we analyze its performance using some Monte Carlo simulations and real data of S&P 500. Based on our empirical analysis, we find that the proposed estimation method yields better out-of-sample results compared to the approach by [Bondarenko \(2003\)](#).

This chapter is related to the literature on functional data analysis. For a general treatment, see the books by [Bosq \(2000\)](#) and [Ferraty and Vieu \(2006\)](#). Theoretical results on the model where the predictor is a function and the response is a scalar are developed by [Cardot et al. \(1999\)](#), [Cai and Hall \(2006\)](#), [Delaigle and Hall \(2012\)](#), [Tsafack \(2020\)](#) among others. The more general case where both predictor and response variables are functions is discussed by [Kargin and Onatski \(2008\)](#), [Cai et al. \(2009\)](#), [Kokoszka and Zhang \(2010\)](#), [Park and Qian \(2012\)](#), [Chang et al. \(2016\)](#), [Benatia et al. \(2017\)](#) among others. [Park and Qian \(2012\)](#) consider a functional regression model where the predictor and response variables are densities. In our chapter, the density is the functional parameter to be estimated. Another difference is that [Park and Qian \(2012\)](#) tackle the ill-posed problem in their model by estimating the operator via the functional principal components (FPCA) while we use LF method. Compared to FPCA, LF has the advantage to require weaker assumptions on the eigenvalues of the covariance operator of the predictor function. In a follow-up paper, [Chang et al. \(2016\)](#) analyze the nonstationarity in the time-series of densities via a unit root test.

The rest of the chapter is organized as follows. Section 2 introduces the theoretical model and the estimation method. Section 3 establishes the rate of convergence and the asymptotic normality results of the estimator. It also suggests a data-driven method to select the optimal tuning parameter. Section 4 presents the results of the simulations. Section 5 is dedicated to the empirical analysis. Section 6 concludes. An online appendix <https://www.dropbox.com/>

s/bk13cppnqnqz4j0/RND_Appendix.pdf?dl=0 describes the implementation and contains all the proofs.

2. ESTIMATION OF RISK-NEUTRAL DENSITY

This section shows that the risk-neutral density can be viewed as the solution to an integral equation and explains how to estimate it using LF estimation technique.

2.1. Option Pricing Formula

In the intertemporal equilibrium models, the current price of a security can be expressed as the expected net present value of its future payoffs discounted back to the present. The expectation is obtained with respect to the risk-neutral density (Cox & Ross, 1976), also called state-price density or equivalent-martingale measure (see Harrison & Kreps, 1979). More specifically, in the derivative market, an option is defined as a contract giving the right (and not the obligation) to buy or sell a risky asset with price s at a predetermined value called strike price κ at (or within) a given maturity date of the contract. There exist many kinds of options in the market. The main ones are the American and the European options. In this chapter, we focus on European options characterized by the fact that the exercise of the contract is possible only at the given maturity date.

Then, in the context of a complete market, the price of a European put option P_t with a maturity $T - t$, an underlying price at maturity S_T and a strike price κ , is equal to the expected pay-offs $Z(S_T)$ discounted back to the present. In other words, it is given by

$$P_t = e^{-r_{t,T}(T-t)} \int_0^\infty Z(S_T) f(S_T | (T-t)) dS_T \quad (2)$$

where $f(S_T | (T-t))$ is the unobserved risk-neutral density (RND) of S_T conditional on the maturity T , $r_{t,T}$ is the riskless interest rate between date t and T , and $Z(S_T) = \max(\kappa - S_T, 0)$ is the pay-off. To simplify notation, we will denote $\tau = T - t$ and $r_{t,T} = r$. It is important to mention that we use a cross-section of option prices all observed at the same time t .

The previous equation holds when it is assumed that the market is complete, this means that market participants have all the information about the risky assets. Because of illiquidity in the market, transaction cost, taxes, measurement errors, the market is usually incomplete (see Gourieroux & Jasiak, 2001). Then, it may exist an error term capturing all those uncontrolled information and this uncertainty may vary according to the time to maturity of the option. The longer the time-to-maturity is, the bigger the variability (see Ait-Sahalia et al., 2018; Driessen et al., 2009). Then, for each

option i (corresponding to a strike price κ_i) at the same fixed time t and the same maturity τ , we have the following equation:

$$Y_i = \int_0^\infty Z_i(s) f(s | \tau) ds + \varepsilon_i \quad (3)$$

where $Y_i = e^{r\tau} P_i$, ε_i is a conditionally zero-mean, homoskedastic error term. For the sake of this model, we assume that there is an infinite possibilities of pay-offs at the maturity date, which means that the set of potential payoffs is very dense and the conditional density is a function taking its values on the real line \mathbb{R} . This leads to a functional linear regression with the functional predictor represented by the future payoff $Z_i(s) = \max(\kappa_i - s, 0)$ and a scalar response (Y_i) . The call options are also considered by using the appropriate payoff $Z_i(s) = \max(s - \kappa_i, 0)$.

2.2. RND as the Solution of an Integral Equation

Let us define $\mathbb{H} = L^2([0, +\infty))$ the space of square integrable functions mapping from the interval $[0, +\infty)$ to the set of real numbers \mathbb{R} . \mathbb{H} is a Hilbert-space endowed with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\|\cdot\|$, which are respectively defined as follows: $\langle f, g \rangle = \int_0^{+\infty} f(t)g(t) dt$ and $\|f\| = (\int_0^{+\infty} f^2(t) dt)^{1/2}$.

Let us consider the sample $((\kappa_1, Y_1), \dots, (\kappa_n, Y_n))$ of independent pairs following the same distribution as the population version (κ, Y) . We consider the functional linear model where $(Z_i)_{i=1\dots n}$ is the sample of functional predictor variables of the regression representing the set of possible pay-offs at maturity for each option and $(Y_i)_{i=1\dots n}$ is the scalar response.

For each $i \in \{1, \dots, n\}$, $Z_i(s) = \max(\kappa_i - s, 0)$, hence Z_i is random only through κ_i . Additionally, the predictor function $(Z_i(s))_{i=1\dots n}$ is such that $Z_i(s) \geq 0$, this means that $E(Z_i(s)) \geq 0$ for each $s \in [0, +\infty)$ and $E(Y_i) > 0$. We assume that $E[\kappa_i^3] < +\infty$. Indeed, this assumption guarantees that $\int_0^{+\infty} E(Z_i^2(s)) ds < \infty$,

which means that the predictor function is square integrable (see Lemma 1 in the Appendix). Then, for each time t the model is a cross-sectional regression presented as follows:

$$Y_i = \int_0^{+\infty} Z_i(s) f(s) ds + \varepsilon_i \quad (4)$$

where $(f(s))_{s \in [0, +\infty)}$ is a function that belongs to the space \mathbb{H} and ε_i , $i = 1, \dots, n$, are independent and homoskedastic¹ such that $\mathbb{E}(\varepsilon_i | \kappa_1, \dots, \kappa_n) = 0$ and $\mathbb{E}(\varepsilon_i^2 | \kappa_1, \dots, \kappa_n) = \sigma^2 < \infty$ for each $i \in \{1, \dots, n\}$.

By premultiplying both sides of Equation (4) by $Z_i(u)$ and taking the expectation, we obtain

$$\mathbb{E}[Z_i(u)Y_i] = \int_0^{+\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds + \mathbb{E}[Z_i(u)\varepsilon_i].$$

Since $\mathbb{E}[Z_i(u)\varepsilon_i] = 0$, then

$$\mathbb{E}[Z_i(u)Y_i] = \int_0^{+\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds$$

In a compact form we can write

$$C_{zy} = Kf, \quad (5)$$

where $C_{zy}(u) = \mathbb{E}[Z_i(u)Y_i]$ is the cross-covariance function between the predictor variable Z and the response variable Y and K is the covariance operator from \mathbb{H} to \mathbb{H} defined as

$$Kf = \int_0^{\infty} \mathbb{E}[Z_i(u)Z_i(s)]f(s)ds.$$

Our main goal is to estimate the function f solution of (5). Equation (5) is a Fredholm equation of the first kind. Solving this equation is an ill-posed problem as K is a bounded operator mapping from an infinite dimensional space \mathbb{H} to \mathbb{H} . This means that the direct inverse of K is not continuous and K is not invertible in \mathbb{H} but only on a subset of \mathbb{H} . If the operator K were invertible, we could estimate f using $f(s) = K^{-1}C_{zy}(s)$ for each $s \in [0, +\infty)$. However, in our context, estimating f by $\hat{K}^{-1}\hat{C}_{zy}$ would lead to an unstable estimator of the functional parameter, which would not be consistent (see [Carrasco et al., 2007](#)). To overcome this issue, we propose to use a regularization technique called LF method. This method will stabilize the inverse of K and permits to obtain a consistent estimator of f . In the next section, we will present the LF method.

2.3. The Landweber-Fridman Method

Recall that we want to estimate f solution of the equation $C_{zy} = Kf$. The main idea of the LF method is to approach the solution to this equation by an iterative algorithm similar to the fixed point procedure with the goal of minimizing $\|C_{zy} - Kf\|$. Instead of iterating all the way to convergence, the algorithm stops after a finite number of iterations. The early termination stabilizes the solution. It is a regularization technique used in the inverse problem literature (see [Carrasco et al., 2007; Engl et al., 1996](#)). The algorithm is presented below. Let ω be a constant such that $0 \leq \omega \leq 1/\lambda_1$, where λ_1 is the largest eigenvalue of K .

- At the first iteration, take $f_0(s) = \omega C_{zy}(s)$.

- For $h = 1, \dots, \frac{1}{\alpha} - 1$, calculate

$$f_h(s) = f_{h-1}(s) + \omega(C_{zy}(s) - Kf_{h-1}(s)) \quad (6)$$

where α is a regularization parameter chosen so that $\frac{1}{\alpha} - 1$ is an integer, hence $0 < \alpha < 1$.

- For convenience, the resulting estimator f_h is denoted f_α with $f_\alpha(s) = K_\alpha^{-1}C_{zy}(s)$, for each $s \in [0, +\infty)$ and K_α^{-1} denotes the regularized inverse of K defined below.

After $\frac{1}{\alpha} - 1$ iterations, the regularized inverse of K is given by

$$K_\alpha^{-1}\phi(s) = \omega \sum_{l=0}^{\frac{1}{\alpha}-1} (I - \omega K)^l \phi(s)$$

where $s \in [0, +\infty)$. The regularization parameter α will be chosen via a data driven method described later. Let us denote $(\lambda_j, v_j)_{j \geq 1}$ the eigensystem of the covariance operator K , then we can also write K_α^{-1} in terms of the eigensystem of K as follows

$$K_\alpha^{-1}\phi = \sum_{j=1}^{\infty} \frac{q(\alpha, \lambda_j)}{\lambda_j} \langle v_j, \phi \rangle v_j$$

for each function ϕ and $q(\alpha, \lambda_j) = \left[1 - (1 - \omega \lambda_j)^{1/\alpha} \right]$.

The true operators K_α^{-1} , K and C_{zy} are unobservable. In practice, they are replaced by their empirical counterparts. Then, the estimated density function is given by $\hat{f}_\alpha(s) = \hat{K}_\alpha^{-1}\hat{C}_{zy}(s)$. In other words, we have

$$\hat{f}_\alpha(s) = \omega \sum_{l=0}^{\frac{1}{\alpha}-1} (I - \omega \hat{K})^l \hat{C}_{zy}(s) \quad (7)$$

where \hat{K} is the integral operator from \mathbb{H} to \mathbb{H} with kernel

$$\hat{k}(u, s) = \frac{1}{n} \sum_{i=1}^n Z_i(u) Z_i(s)$$

and

$$\hat{C}_{zy}(s) = \frac{1}{n} \sum_{i=1}^n Z_i(s) Y_i.$$

In Section 2 of the online appendix, we explain how to compute \hat{f}_α using only products of matrices instead of operators. This is how we implement the method in the simulations and application.

Other regularization techniques could be used to solve (5). For instance, Tikhonov regularization is a popular method which has been used by [Benatia et al. \(2017\)](#) for the estimation of a functional regression with functional response. It is easy to implement but suffers from saturation (its rate of convergence cannot improve beyond a certain level) while LF does not suffer of saturation. Another popular regularization technique is principal components, but analyzing its properties requires some assumptions on the decay rate of the eigenvalues λ_j which are not needed here, see for instance [Hall and Horowitz \(2007\)](#) for an application of principal components to a functional regression with scalar response.

2.4. Density Correction

Our estimator of f , \hat{f}_α , is not necessarily positive and does not integrate to one.

As the true function f is a density, we propose to transform our estimator \hat{f}_α into a density using the methods proposed by [Glad et al. \(2003\)](#). The correction is different depending on whether $\int_0^\infty \max\{0, \hat{f}_\alpha(s)\} ds \geq 1$ or $\int_0^\infty \max\{0, \hat{f}_\alpha(s)\} ds < 1$.

Case 1: Case where $\int_0^\infty \max\{0, \hat{f}_\alpha(s)\} ds \geq 1$.

The corrected estimator is given by

$$\tilde{f}_\alpha(s) = \max\{0, \hat{f}_\alpha(s) - \xi\}$$

where ξ is a positive constant chosen so that $\int_0^\infty \tilde{f}_\alpha(s) ds = 1$.

Note that such a ξ necessarily exists because $\int_0^\infty \tilde{f}_\alpha(s) ds \geq 1$ for $\xi = 0$ and $\int_0^\infty \tilde{f}_\alpha(s) ds = 0$ for $\xi = \infty$.

Case 2: Case where $\int_0^\infty \max\{0, \hat{f}_\alpha(s)\} ds < 1$.

The corrected estimator \check{f}_α is computed as follows

$$\check{f}_\alpha(s) = \begin{cases} \max\{0, \hat{f}_\alpha(s)\} + \eta_M & \text{for } |s| \leq M, \\ \max\{0, \hat{f}_\alpha(s)\} & \text{for } |s| > M, \end{cases}$$

where

$$\eta_M = \frac{1}{2M} \left[1 - \int_0^\infty \max\{0, \hat{f}_\alpha(s)\} ds \right].$$

Remarks.

1. In Case 1, Theorem 1 of [Glad et al. \(2003\)](#) shows that \tilde{f}_α is always better than $\hat{f}_\alpha(s)$ in the sense that $\|\tilde{f}_\alpha - f\|^2 \leq \|\hat{f}_\alpha - f\|^2$ for all n , almost surely, hence $E\|\tilde{f}_\alpha - f\|^2 \leq E\|\hat{f}_\alpha - f\|^2$ so that the mean integrated squared error (MISE) of \tilde{f}_α is always smaller than that of \hat{f}_α .
2. In Case 2, Theorem 2 of [Glad et al. \(2003\)](#) establishes that

$$E\left\|\tilde{f}_\alpha - f\right\|^2 \leq E\left\|\hat{f}_\alpha - f\right\|^2 + \frac{3}{2M}.$$

Hence, one can make the MISE of \tilde{f}_α arbitrary close to that of \hat{f}_α by choosing M dependent of n and large, for instance $M = n$.

3. An algorithm to select ξ in practice is presented in Section 1 of the online appendix.
4. An alternative correction of \hat{f}_α could have relied on a normalization

$$\frac{\max\{0, \hat{f}_\alpha\}}{\int_0^\infty \max\{0, \hat{f}_\alpha(s)\} ds}. \quad (8)$$

However, there is no guarantee that this normalization improves the accuracy of the estimator. The MISE of the normalized estimator may actually be worse than that of the original estimator as discussed in [Glad et al. \(2003\)](#).

3. ASYMPTOTIC PROPERTIES OF THE LF ESTIMATOR

3.1. Convergence Rate

In this section, we derive the convergence rate of the conditional mean square error (MSE) of \hat{f}_α . For this purpose, the following assumptions are imposed.

Assumption A1. (κ_i, Y_i) are i.i.d. κ_i has a continuous density on \mathbb{R}^+ with $E[\kappa_i^6] < \infty$.

Assumption A2. $\int_0^{+\infty} f^2(t) dt < \infty$, $E[\varepsilon_i | \kappa_1, \dots, \kappa_n] = 0$, $E[\varepsilon_i^2 | \kappa_1, \dots, \kappa_n] = \sigma^2 < +\infty$, $E[\varepsilon_i^4 | \kappa_1, \dots, \kappa_n] < \infty$.

Assumption A3. The eigenvalues of the covariance operator K are positive and ordered in decreasing order, that is $\lambda_1 \geq \lambda_2 \geq \dots > 0$. Similarly, the eigenvalues of \hat{K} are ordered in decreasing order, $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n$.

Assumption A4. We assume that for some $\mu \geq 0$, f satisfies

$$\sum_{j=1}^{\infty} \frac{\langle f, v_j \rangle^2}{\lambda_j^\mu} < \infty.$$

Assumption **A1** imposes that $(Z_i, Y_i)_{i=1,\dots,n}$ are independent, identically distributed as (Z, Y) . It is useful in order to derive the consistency of the covariance operators \hat{K} and \hat{C}_{zy} , and to prove the central limit properties of the estimated functions. The fact that $E[\kappa_i^3] < +\infty$ guarantees that the predictor functions $(Z_i)_{i=1,\dots,n}$ are square integrable. Moreover, it also guarantees that the covariance operator K is nuclear, which in turn implies that it is Hilbert-Schmidt² (see Lemma 1 in the online Appendix). However, we need a stronger condition ($E[\kappa_i^6] < +\infty$) to guarantee that $E(\|Z_i\|^4) < \infty$ and hence show the consistency of \hat{K} to K (see Lemma 1 in Appendix and comment below).

Assumption **A2** imposes that the error term ε_i is homoskedastic and Z_i is exogenous. The homoskedasticity assumption is reasonable because we use cross-sectional data. It could be relaxed at the cost of a more complicated expression for the asymptotic variance. **A1** and **A2** are sufficient conditions to ensure that $\|\hat{K} - K\|_{HS}^2 = O_p\left(\frac{1}{n}\right)$, see the proof of Theorem 4 of Carrasco and Florens (2000),

where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm of operators. Moreover, Assumptions **A1** and **A2** imply that the fourth moment of Y_i exists, this rules out possible fat-tails of the distribution of Y_i . To relax this assumption, one would need to use appropriate statistical tools (see Ibragimov et al., 2015) In that case, we expect that the rate of convergence and asymptotic distribution would be different.

Assumption **A3** ensures that the null space of K reduces to 0, $\mathcal{N}(K) = \{0\}$. Hence, f is the unique solution of $C_{xy} = Kf$ and therefore it is identified. Note that the eigenvalues do not have to be distinct and multiple eigenfunctions may be associated with the same eigenvalue. This is quite a bit more general than assumptions usually imposed for principal component method where the eigenvalues need to be distinct and sufficiently spaced from each other (see for instance Hall & Horowitz, 2007).

Assumption **A4** is a source condition important to derive how the bias and estimation error terms behave. To satisfy this condition, the Fourier coefficients of f , $\langle f, v_j \rangle$, need to decline to zero fast compared to the eigenvalues of K . The parameter μ characterizes the severity of the ill-posed problem. Larger values of μ are associated with less severe ill-posed problem. This source condition is discussed in Engl et al. (1996) and Carrasco et al. (2007) and was used in econometric papers by Chen and Reiss (2011), Darolles et al. (2011), and Gagliardini and Scaillet (2012) among others.

Proposition 1. Let \hat{f}_α be the estimated density function corrected either with Case 1 or 2. Under Assumptions A1–A4, if $\alpha^2 n \rightarrow \infty$, then

$$\mathbb{E}\left[\left\|\hat{f}_\alpha - f\right\|^2 | \kappa_1, \dots, \kappa_n\right] = O_p(\alpha^\mu) + O_p\left(\frac{1}{\alpha^2 n}\right) \quad (9)$$

where μ is the nonnegative constant defined in Assumption A4. Hence the conditional MISE converges to zero as the sample size increases.

Remarks.

- The first term of the conditional MISE is the squared bias and the second term is the estimation error.
- As α goes to zero, the squared bias term goes to zero, while the estimation error term increases. So, we have the usual trade-off between bias and variance. The optimal parameter α is selected in such a way that the squared bias and the variance term are of the same order.
- If $\alpha \sim n^{-1/(2+\mu)}$, then $MISE \sim n^{-\frac{\mu}{2+\mu}}$.

3.2. Asymptotic Normality

Carrasco et al. (2014) derive results on the asymptotic normality of $\hat{f}_\alpha - f_\alpha$ for a fixed α . While this approach can be useful to perform hypothesis testing, it cannot be used to construct confidence intervals of f . Here, we are going to study the pointwise asymptotic normality assuming α goes to zero instead. For related results, see Horowitz (2007) in the context of nonparametric instrumental variables and Carrasco and Florens (2011) who study the asymptotic normality of a deconvolution estimator.

Replacing \hat{f}_α and f_α by their expressions, we obtain

$$\hat{f}_\alpha - f_\alpha = K_\alpha^{-1} \hat{C}_{ze} + (\hat{K}_\alpha^{-1} - K_\alpha^{-1}) \hat{C}_{ze} + (\hat{K}_\alpha^{-1} \hat{K} - K_\alpha^{-1} K) f.$$

The distribution of $\hat{f}_\alpha - f_\alpha$ will be driven by that of $K_\alpha^{-1} \hat{C}_{ze}(s)$. As α depends on n and $K_\alpha^{-1} \hat{C}_{ze}(s) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i(K_\alpha^{-1} Z_i)(s)$ with $E(\varepsilon_i(K_\alpha^{-1} Z_i)(s)) = 0$, it follows that $K_\alpha^{-1} \hat{C}_{ze}(s)$ is a triangular array. A sufficient condition for

$$\frac{(K_\alpha^{-1} \hat{C}_{ze})(s)}{\sqrt{V((K_\alpha^{-1} \hat{C}_{ze})(s))}} \xrightarrow{d} \mathcal{N}(0,1)$$

is that the Lyapunov's condition holds (Billingsley, 1995, Theorem 27.3), i.e., for some $\delta > 0$,

$$\frac{E\left[\left|\varepsilon_i(K_\alpha^{-1} Z_i)(s)\right|^{2+\delta}\right]}{n^{\delta/2} \left[E\left(\left(\varepsilon_i(K_\alpha^{-1} Z_i)(s)\right)^2\right)\right]^{1+\delta/2}} \rightarrow 0. \quad (10)$$

A sufficient condition for (10) is given in the next assumption.

$$\text{Assumption A5. } \frac{E\left[\left|\varepsilon_i(K_\alpha^{-1}Z_i)(s)\right|^3\right]}{n^{1/2}\left[E\left(\left(\varepsilon_i(K_\alpha^{-1}Z_i)(s)\right)^2\right)\right]^{3/2}} \rightarrow 0.$$

To obtain

$$\frac{\hat{f}_\alpha(s) - f_\alpha(s)}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i(K_\alpha^{-1}Z_i)(s)\right)^2\right)}} \xrightarrow{d} \mathcal{N}(0,1), \quad (11)$$

it is sufficient that the following condition is satisfied.

Assumption A6.

$$\frac{\left|(\hat{K}_\alpha^{-1} - K_\alpha^{-1})\hat{C}_{z\varepsilon}(s) + (\hat{K}_\alpha^{-1}\hat{K} - K_\alpha^{-1}K)f(s)\right|}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i(K_\alpha^{-1}Z_i)(s)\right)^2\right)}} \xrightarrow{P} 0.$$

Finally, to be able to replace f_α by f in (11), we need an extra condition guaranteeing that the bias is negligible.

Assumption A7.

$$\frac{\left|\sum_{j=1}^{\infty} (q(\alpha, \lambda_j) - 1) \langle v_j, f \rangle v_j(s)\right|}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i(K_\alpha^{-1}Z_i)(s)\right)^2\right)}} \xrightarrow{P} 0.$$

Remark that Assumptions A5 to A7 restrict the rate of convergence of α and possibly the admissible range of values of s . As a result, the rate of convergence of \hat{f}_α depends on s .

An extra assumption is needed to be able to consistently estimate the variance.

$$\text{Assumption A8. } E(\varepsilon_i^4) < \infty, \frac{1}{n} E\left[\left((K_\alpha^{-1}Z_i)(s)\right)^4\right] \rightarrow 0, \text{ and } n\alpha^2 \rightarrow \infty.$$

Proposition 2. Under Assumptions A1 to A7,

$$\frac{\hat{f}_\alpha(s) - f(s)}{\sqrt{\frac{1}{n} E\left(\left(\varepsilon_i(K_\alpha^{-1}Z_i)(s)\right)^2\right)}} \xrightarrow{d} \mathcal{N}(0,1).$$

Moreover, under Assumptions A1 to A8,

$$\frac{\hat{f}_\alpha(s) - f(s)}{\sqrt{\hat{V}_n(s)}} \xrightarrow{d} \mathcal{N}(0,1)$$

where $\hat{V}_n(s) = \hat{\sigma}^2 \frac{1}{n^2} \sum_{i=1}^n \left((\hat{K}_\alpha^{-1} Z_i)(s) \right)^2$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \int z_i(s) \hat{f}_\alpha(s) ds \right)^2$

It is important to notice that Proposition 2 does not imply that $\hat{f}_\alpha(s)$ converges at a \sqrt{n} rate of convergence because the term $E\left(\left(\varepsilon_i(K_\alpha^{-1} Z_i)(s)\right)^2\right)$ usually diverges so that the rate is slower. The previous results can be used to construct an asymptotic confidence interval for $f(s)$.

Corollary 1. *Under Assumptions A1 to A8, the asymptotic $1-a$ confidence interval for $f(s)$ is given by*

$$\hat{f}_\alpha(s) - z_{a/2} \hat{V}_n(s)^{1/2} \leq f(s) \leq \hat{f}_\alpha(s) + z_{a/2} \hat{V}_n(s)^{1/2}$$

where $z_{a/2}$ is the $1-a/2$ quantile of the standard normal distribution.

3.3. Data-driven Selection of the Tuning Parameter

According to the consistency results, it can be noticed that the estimation of the RND depends on the tuning parameter α . Then, this parameter should be selected optimally. Since the main goal is to estimate the RND and therefore predict the call and put prices, we define a prediction criterion to select the optimal parameter α . Then, we choose the regularization parameter in such a way that the mean squared prediction error (MSPE) is minimized. We use the K-fold cross-validation for the selection procedure. Let us split the initial sample into M subsamples denoted I_1, \dots, I_M .

$$\alpha_{op} = \operatorname{argmin}_{\alpha \in \mathcal{I}_\alpha} \frac{1}{M} \sum_{\ell=1}^M \frac{1}{\operatorname{card}(I_\ell)} \sum_{j \in I_\ell} (Y_j - \hat{Y}_j)^2. \quad (12)$$

For $\ell \in \{I_1, \dots, I_M\}$, we estimate the parameter f in the sample $\mathcal{I}_{-\ell}$ representing all the observations not in \mathcal{I}_ℓ . Then, we predict the response variable in \mathcal{I}_ℓ considered as the hold-out sample. \hat{Y}_j is the prediction of the j^{th} observation in \mathcal{I}_ℓ . Hence, we calculate the MSPE for each candidate α . \mathcal{I}_α is the set of candidate α .

An alternative approach suggested by [Engl et al. \(1996\)](#) is to choose the parameter α such that the following objective function is minimized.

$$\alpha_{op} = \operatorname{argmin}_{\alpha \in \mathcal{I}_\alpha} \left\| \hat{f}_\alpha \right\|^2 \left\| \hat{C}_{zy} - \hat{K}(\hat{f}_\alpha) \right\|^2. \quad (13)$$

4. SIMULATIONS

In this section, the Monte-Carlo simulations are used to evaluate the proposed estimation method. For this purpose, we consider a variety of data generating

processes. We compare the proposed estimation method with the positive convolution approach (PosConv) suggested by [Bondarenko \(2003\)](#) considered as a benchmark. This simulation follows the same idea as the one by [Bondarenko \(2003\)](#) with a change on the underlying discretization, the error term distribution and the dates. We consider the functional regression model.

$$Y_i = \int_0^{+\infty} Z_i(s) f(s) ds + \varepsilon_i \quad (14)$$

where $(Y_i, Z_i)_{i=1,\dots,n}$ is the sample of generated data. The response variable $Y_i = \exp(r\tau) P_i$ where P_i are call option prices and r is set equal to 0. The predictors are the pay-offs functions $Z_i(s) = \max\{s - \kappa_i, 0\}$. To characterize the incompleteness of the market and allow for measurement errors, an error term ε_i is added to the option price with $\mathbb{E}[\varepsilon_i] = 0$. This error term is assumed to follow a uniform distribution, $\varepsilon_i \sim \mathcal{U}[-0.5, 0.5]$.

The true RND $f(s)$ is specified as a linear mixture of lognormal distributions presented as follows

$$\begin{aligned} f(s) &= \pi_1 \mathcal{LN}(s | \eta_1, \sigma_1) + \pi_2 \mathcal{LN}(s | \eta_2, \sigma_2) + \pi_3 \mathcal{LN}(s | \eta_3, \sigma_3), \\ \pi_1 + \pi_2 + \pi_3 &= 1, \end{aligned}$$

and $\mathcal{LN}(s | \eta_j, \sigma_j)$, $j = 1, 2, 3$ is a lognormal distribution

$$\mathcal{LN}(s | \eta_j, \sigma_j) = \frac{1}{\sqrt{2\pi}\sigma_j s} e^{-\frac{\left(\ln\left(\frac{s}{\eta_j}\right) - \frac{1}{2}\sigma_j^2\right)^2}{2\sigma_j^2}}.$$

Four different models are considered. Models 1 and 2 correspond to option data with maturity between one to two months. Model 3 considers options with maturity between three and six months and Model 4 considers options with maturity exceeding six months.

The set of strike prices is defined as follows $\mathcal{K} = [1500, 1505, 1510, \dots, 3000]$, which means that the sample size is $n = 301$ and the underlying follows the aforementioned lognormal mixture distribution. This parametrization is used to match a typical cross-section of the S&P 500 index options traded at the Chicago Board Options Exchange (CBOE) on June 25, 2017 for Model 1, August 04, 2017 for Model 2 and 4, and June 05, 2017 for Model 3. The parameters of the RND are presented in [Table 1](#).

We simulate the data and evaluate the performance of the estimation method with 2 criteria:

- The Root Mean Squared Prediction Error (RMSPE) between the estimated put prices \hat{Y}_α and the theoretical one Y .

Table 1. Parameters of the Lognormal Mixture Density.

Parameters	Model 1	Model 2	Model 3	Model 4
π_1	0.0812	0.0823	0.0562	0.5934
π_2	0.0914	0.8115	0.3347	0.2894
π_3	0.8274	0.1062	0.5791	0.1172
η_1	7.8020	7.7669	7.5628	7.8594
η_2	7.7023	7.8176	7.7690	7.7779
η_3	7.8052	7.8165	7.8340	7.5688
σ_1	0.0285	0.0543	0.2032	0.0369
σ_2	0.0987	0.0214	0.0599	0.0696
σ_3	0.0245	0.0247	0.0323	0.2090
Time to maturity (days)	53	35	148	224
Risk free rate (r in %)	0.95	1.06	0.95	1.06
Current index price (S_0)	2436.10	2476.83	2436.10	2476.83
Trading date	2017-06-05	2017-08-04	2017-06-05	2017-08-04

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_{i,\alpha} - Y_i)^2}.$$

- The Root Integrated Squared Error (RISE) between the estimated density $\hat{f}_{\alpha,\delta}$ and the theoretical one f .

$$RISE = \frac{1}{\|f\|} \sqrt{\int_0^\infty (\hat{f}_\alpha(s) - f(s))^2 ds}.$$

Simulations are performed with the following procedure:

1. Select the parameters of the model and fix the RND $f(s)$.
2. Generate the sample of strike prices and the predictor variable Z_i .
3. Generate the n random variables ε_i from the uniform distribution.
4. Compute the values of the option prices $Y_i = \int_0^\infty Z_i(s) f(s) ds + \varepsilon_i$.
5. Center the variables Y_i and Z_i . Then, apply the LF method on centered data.
6. Run the 10-fold cross-validation procedure in order to select the optimal number of iterations. The whole sample is randomly divided into 10 groups. Nine groups are used as the training sample and the last one is used as the validation sample. This operation is repeated 10 times as we have 10 different groups in the cross-validation procedure. Then, we average to get an estimate of the mean squared prediction error. This quantity is minimized with respect to α .
7. Estimate the RND using the optimal tuning parameter α selected in Step 6 and calculate the predicted option price.
8. Calculate the RISE and RMSPE by 10-fold cross-validation.
9. Repeat the steps 2 to 8 with 100 iterations and calculate the average RISE and RMSPE over the simulations.

All the numerical integrations are performed with the trapezoidal rule. It is also possible to use other integration rules such as the Newton-Cotes or adaptive

quadrature. For LF, the tuning parameter ω is chosen equal to $1/\sum_j \lambda_j$, where λ_j are the eigenvalues of \hat{K} . To correct the density, we use the formula (8), which is faster than the procedure proposed by Glad et al. (2003). To implement PosConv, we approximate f by $\hat{f}(x) = \sum a_j \phi_h(x - z_j)$ with $a_j \geq 0$ and $\sum a_j = 1$ using an equispaced grid of 70 points for z_j . The a_j are selected by minimizing³ $\sum_i \left(Y_i - \int_0^\infty Z_i(s) f(s) ds \right)^2$. Moreover, the bandwidth h is selected by 10-fold cross-validation.

Table 2 shows the results from the simulations. Comparing the LF method with PosConv, we can observe that when considering the option pricing with a time-to-maturity of less than 6 months, PosConv method tends to outperform LF in terms of RMSPE (see results from Model 1). In contrast, for options with a maturity of more than 6 months, LF method tends to outperform PosConv in both RISE and RMSPE (Models 3 and 4). Across all models, LF always displays a smaller RISE than PosConv. Also, we can see that as the time to maturity increases, the RISE improves but the RMSPE increases for both methods. This result can be confirmed by observing Figs. 1–3 which respectively display the RND, the cumulative distribution, and the predicted option prices. These figures show that the adjustment of the true RND improves with the maturity, but PosConv tends to adjust better thin tails, while LF tends to fit better fat tails distributions. As PosConv approximates the unknown density with a mixture of normal distributions which has relatively thin tails, it makes sense that this method will perform best when the true density has thin tails. Our method on the other hand lets the data speak by themselves.

5. APPLICATION TO S&P500 OPTIONS

This section focuses on a real data example. For this purpose, we consider the S&P 500 index (SPX) as the data of interest to derive the underlying and the strikes. The S&P 500 option is one of the most liquid and tradable options in

Table 2. Comparison of the Estimation Methods.

Models	Maturity (days)	Criteria	LF	PosConv
Model 1	53	RMSPE	1.513	1.125
		RISE	0.132	0.140
Model 2	35	RMSPE	2.585	2.640
		RISE	0.184	0.340
Model 3	148	RMSPE	4.021	5.058
		RISE	0.032	0.072
Model 4	224	RMSPE	4.850	7.31
		RISE	0.030	0.034

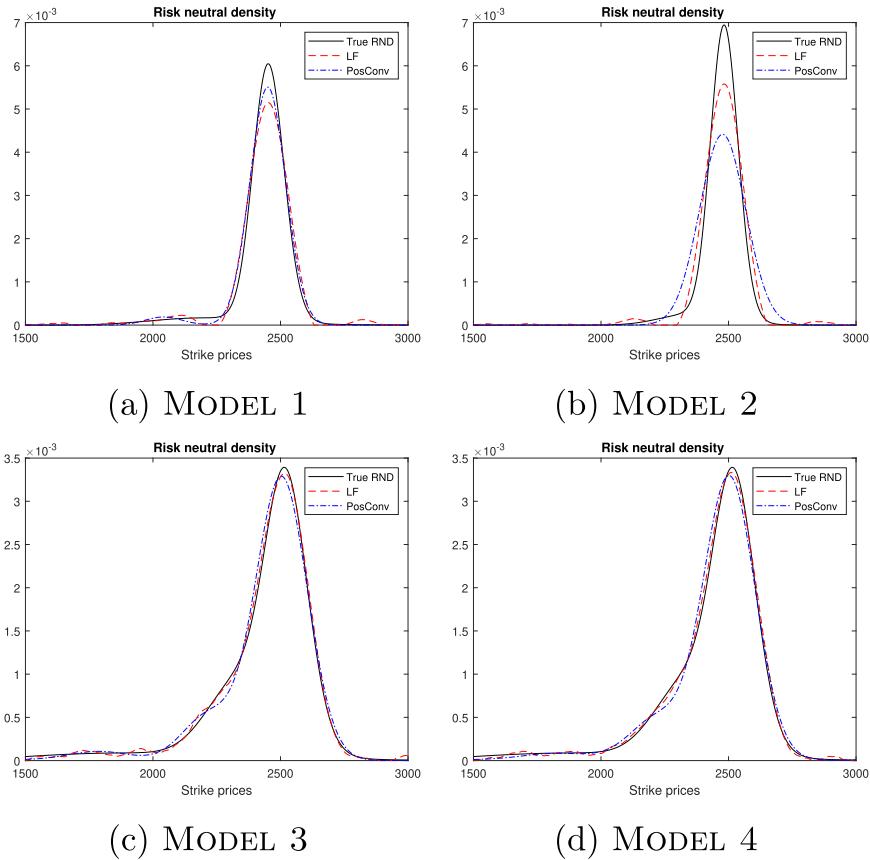


Fig. 1. Estimated Risk Neutral Density.

the market. It represents the aggregated capitalization of the 500 most important corporations in the United States. It is also used as a benchmark to see how well the most important companies are behaving. The data are taken from [Barratt et al. \(2020\)](#). They have been extracted from OptionMetric Ivy database and have been made public on the github account of the authors on https://github.com/cvxgrp/cvx_opt_risk_neutral.

From the database, the best bid and ask prices of all S&P 500 European options are collected for the date of June 3, 2019 with a maturity of 25 days. The price of the index at the end of the same day is also collected. Indeed, it is equal to 2744.45 dollars. The range of values for s is between 1500 and 3999.50 dollars. Put and call options are both considered in the sample. The option prices are not directly observed. We follow the literature by taking the average of the bid and ask prices as proxy for the actual price. The put and call prices are then stacked together. The resulting sample includes 157 observations (49 calls and 108 puts). Next, we

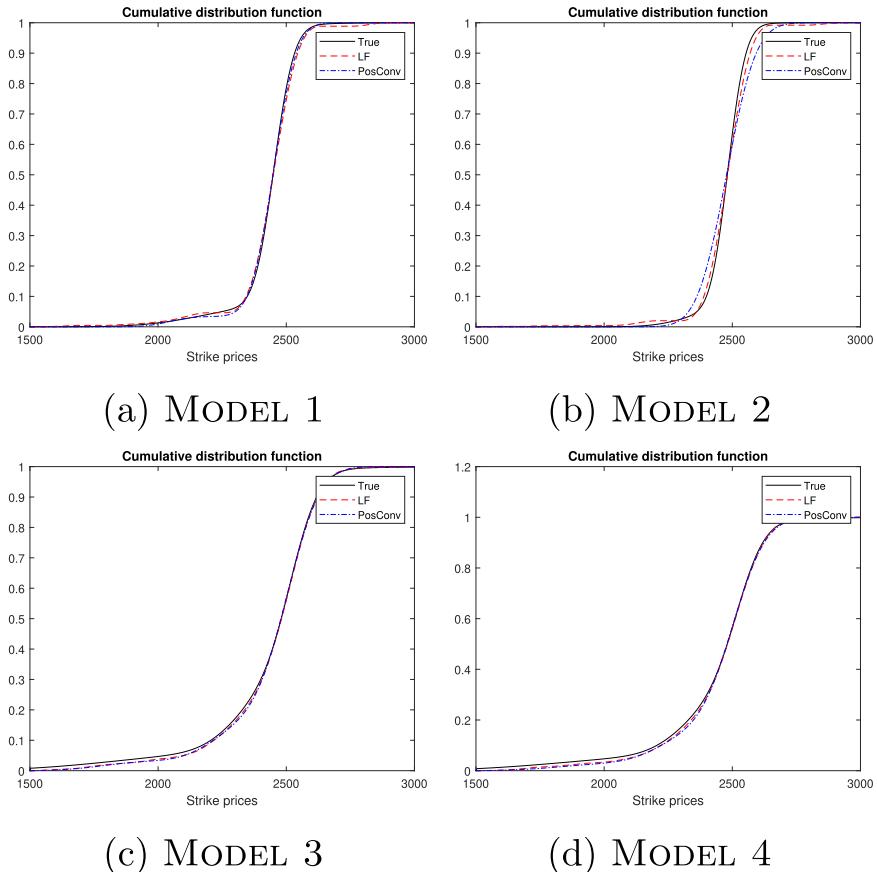


Fig. 2. Estimated Cumulative Distribution Function.

estimate Model (14) where Y_i is the option price and $Z_i(s) = \max(\kappa_i - s, 0)$ for put options and $Z_i(s) = \max(s - \kappa_i, 0)$ for call options.

As in the simulations, the optimal tuning parameter α and optimal bandwidth h are selected by minimizing the RMSPE evaluated by 10-fold cross-validation. Once these tuning parameters are chosen, one can estimate the RND and therefore predict the option prices.

The results are presented in Fig. 4. The estimated distribution from the 25-days maturity options displays a bell-shaped curve centered around the value 2,750 dollars. This density also presents long tails. Fig. 5 presents the predicted call and put prices with their respective 95% (in-sample) confidence interval. The confidence interval is based on LF estimator. It can be noticed that the option prices are well predicted via the suggested estimation approach.

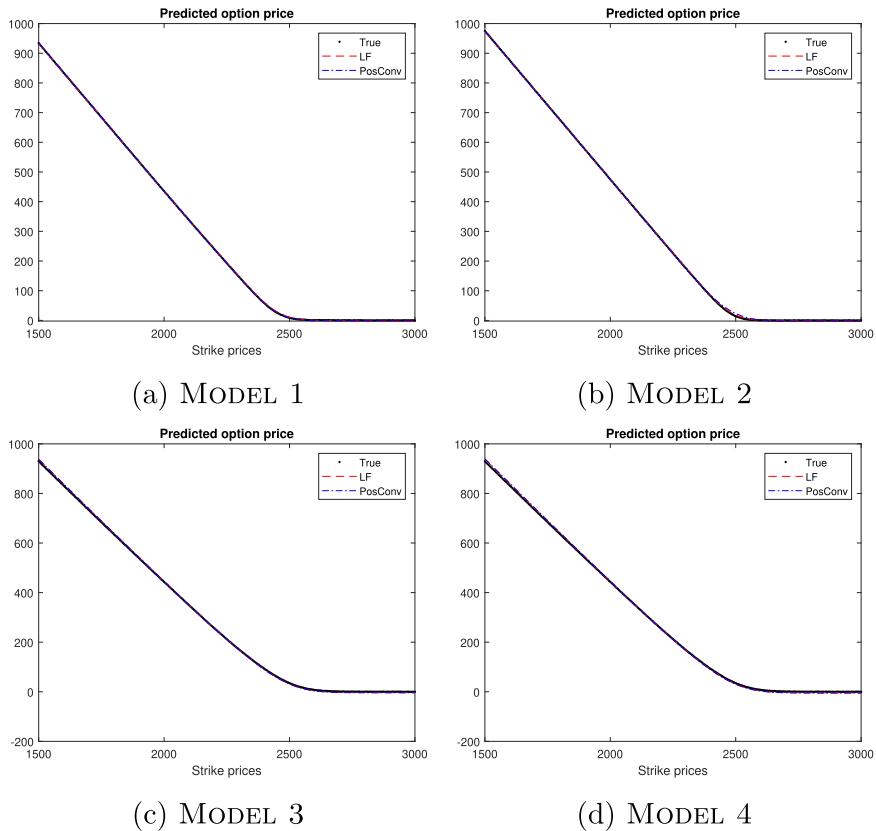


Fig. 3. Predicted Option Prices.

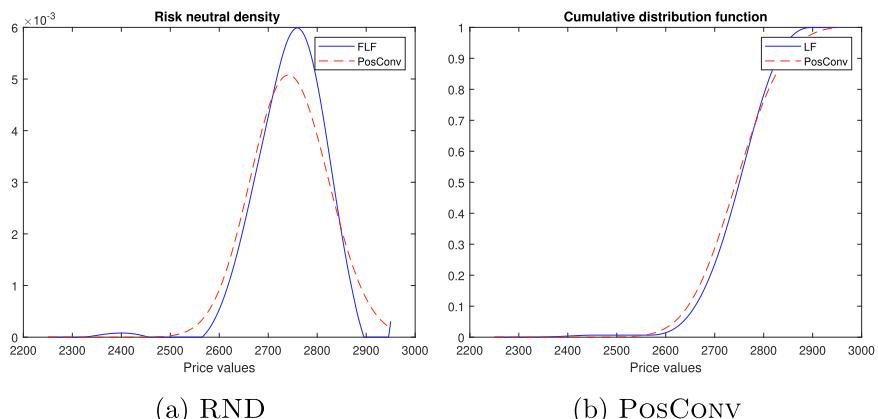


Fig. 4. Estimated RND and CDF with LF and PosConv Methods.

Table 3 reports the LF and PosConv RMSPE evaluated by 10-fold cross-validation. We can see that LF outperforms PosConv method as it displays a smaller RMSPE than PosConv. Additionally, **Fig. 5** shows that LF tends to predict better than PosConv.

One of the potential reasons of the weaker performance for PosConv is that PosConv uses to estimate f a mixture of normal distributions which have thin tails. In the real data, the RND may have fat tails and it was observed, in the simulations, that LF outperforms PosConv for RND with fat tails.

6. CONCLUSION

This chapter proposes to estimate the RND for option pricing models with the functional data analysis framework. Indeed, we consider that a European option price of an asset is evaluated as a weighted average of all possible payoffs of the asset, where the weights represents here the RND of a market participant. To use the functional data analysis framework, we assume that for each asset, one can have an infinity of possible price values at the maturity. This means that at the maturity date, a market participant is exposed to an infinite possibility of payoffs. The set of potential payoffs for each option price is then very dense and is considered as a function. On the same line, the RND is also considered as a collection of values observed on a very fine grid. Therefore, the model setting considered is a functional linear model where the predictor functions are represented by the set of potential payoffs and the response variable is the price of the option, which is a scalar. The estimation method

Table 3. Comparison of the Estimation Methods on Real Data.

Criterion	LF	PosConv
RMSPE	1.065	2.801

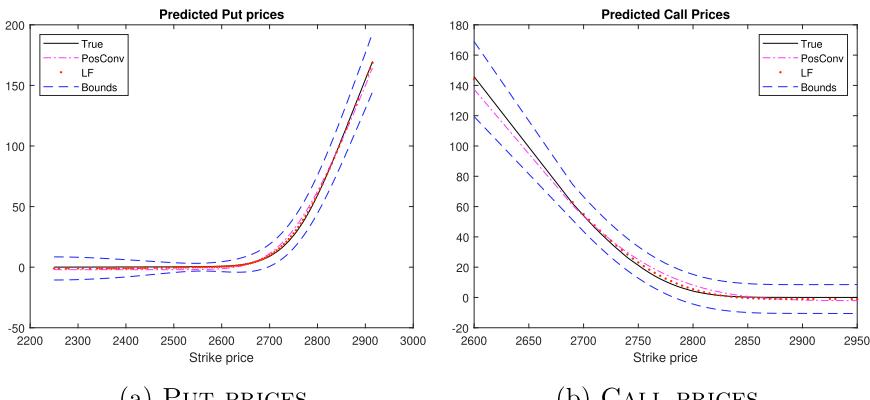


Fig. 5. SPX Predicted Call and Put Prices with 95% Confidence Bounds Using the LF Method.

proposed in this chapter is free of any parametric or semi-parametric assumptions and also takes into account the arbitrage-free theory for option pricing in the model.

One main issue of this model is the high dimensionality problem as the inverse of the covariance operator of the predictor variable is not continuous. This problem leads to unstable estimated function. To overcome this issue, we propose to use a regularization technique called LF. We also control for the positivity and the integration to one of the density by applying a density correction. We derive the consistency and asymptotic normality of the estimated density function. Additionally, we provide confidence intervals for the RND.

Our simulations show that LF estimator captures well the bell-shaped form of the true RND. Comparing the results of LF method with those obtained by [Bondarenko \(2003\)](#), we observed that the PosConv method of [Bondarenko \(2003\)](#) outperforms our LF method when the tails for the RND are thin and our method outperforms when the tails of the RND are fat. The results from real data on S&P 500 options show that LF outperforms PosConv in terms of RMSPE. Therefore, the proposed approach in this chapter can be considered as a promising alternative to the existing ones.

Below, we discuss a few possible extensions.

1. **RND and implied volatility:** one important application of the RND estimation is the usage of the inversion formula to get the implied volatility. As our model is nonparametric and we derive the estimation results directly on option prices and underlying, this inversion formula cannot be obtained. To estimate implied volatility, one should consider a cross-sectional model of options prices with different maturities for the same asset, which is now a heteroskedastic model. An extension of our results to heteroskedastic errors would be needed to obtain the pricing function for put and call options. The next step would be to calculate the realized volatility, which is the variance of the estimated put and call prices. Then, the implied volatility is nothing else but the mean of the realized volatilities for each time t considered in the data.
2. **RND and positivity of the stochastic discount factor:** the stochastic discount factor is another variable of interest in asset pricing models. Based on our estimation approach, one could worry about the positivity of the discount factor. To address this question, one could estimate the option prices via our method, then consider a utility function and estimate its parameters using the recursive utility framework proposed by [Garcia et al. \(2003\)](#). The last step would be to check if the stochastic discount factor is positive. This is left for future research.
3. **Fat tails:** Many studies have empirically documented the presence of fat-tails in financial returns (see [Embrechts et al., 1997](#) among others). This feature should translate into fat tails of the RND and has motivated some work on RND models allowing for semi-fat and fat tails (see [Figlewski, 2010](#); [Hartwig et al., 2001](#)). Our simulations suggest that our method may work well when the tails of the RND are fat. However, a thorough investigation would be needed to confirm this conjecture and is left for future research.
4. **American options:** This chapter focuses on European options only. However, most traded options are American options. Because American options can be exercised before maturity, one cannot express the price of American options

as a linear function of the RND. Some authors have tried to exploit American options to estimate the RND. [Melick and Thomas \(1997\)](#) provide bounds using American options. On the other hand, [Tian \(2011\)](#) develops a method to extract European option prices from American option prices and then use the usual method to estimate the RND. Also, [Flamouris and Giamouridis \(2002\)](#) use the Edgeworth series expansion (ESE) technique to estimate the RND and [Borovkova et al. \(2012\)](#) exploit the partial differential equations approach to derive the option prices. An in-depth analysis of the American options is beyond the scope of the present chapter.

NOTES

1. Heteroskedasticity could be introduced if one considers options with different time-to-maturity. The same estimation procedure could be applied but the standard errors would need to be adjusted to account for heteroskedasticity.
2. A covariance operator is nuclear if and only if its eigenvalues are summable, i.e., $\sum_{j=1}^{\infty} \lambda_j < \infty$. It is Hilbert-Schmidt if and only if $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$.
3. We do not use the formula $\sum \left(Y_i - D^{-2} \hat{f}(x_i) \right)^2$ suggested by [Bondarenko \(2003\)](#) because Bondarenko's formula to compute $D^{-2} f$ seems to ignore the fact that f has support $[0, \infty)$ instead of \mathbb{R} . Moreover, the correct expression for $D^{-2} f$ is different depending on whether the option is a put or a call.

ACKNOWLEDGMENT

The authors thank the editor and referees for their useful comments. Carrasco thanks NSERC for partial financial support.

REFERENCES

- Ait-Sahalia, Y., & Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1–2), 9–47.
- Ait-Sahalia, Y., Karaman, M., & Mancini, L. (2018). *The term structure of variance swaps and risk premia* [Swiss Finance Institute Research Paper 18–37].
- Ait-Sahalia, Y., & Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1–2), 9–51.
- Bahra, B. (1997). *Implied risk-neutral probability density functions from option prices: Theory and application* [Working paper]. Bank of England.
- Barratt, S., Tuck, J., & Boyd, S. (2020). Convex optimization over risk-neutral probabilities. *arXiv preprint arXiv:2003.02878*.
- Benatia, D., Carrasco, M., & Florens, J.-P. (2017). Functional linear regression with functional response. *Journal of Econometrics*, 201(2), 269–291.
- Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Bliss, R. R., & Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The Journal of Finance*, 59(1), 407–446.
- Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, 116(1–2), 85–112.

- Borovkova, S., Permana, F., & Van Der Weide, J. (2012). American basket and spread option pricing by a simple binomial tree. *The Journal of Derivatives*, 19(4), 29–38.
- Bosq, D. (2000). *Linear processes in function spaces: Theory and applications, volume 149 of lecture notes in statistics*. Springer-Verlag New York Inc.
- Cai, T. T., & Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34(5), 2159–2179.
- Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2), 101–113.
- Campbell, J., Lo, A., & MacKinlay, A. (1997). *The econometrics of financial markets*. Princeton University Press.
- Cardot, H., Ferraty, F., & Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1), 11–22.
- Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, 16, 797–834.
- Carrasco, M., & Florens, J.-P. (2011). A spectral method for deconvolving a density. *Econometric Theory*, 27, 546–581.
- Carrasco, M., Florens, J.-P., & Renault, E. (2007). Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics*, 6, 5633–5751.
- Carrasco, M., Florens, J.-P., & Renault, E. (2014). Asymptotic normal inference in linear inverse problems. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 73(74), 140.
- Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, 192(1), 152–167.
- Chen, X., & Reiss, M. (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27, 497–521.
- Cochrane, J. (2005). *Asset pricing: Revised edition*. Princeton University Press.
- Cox, J. C., & Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3(1–2), 145–166.
- Darolles, S., Fan, Y., Florens, J.-P., & Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79(5), 1541–1565.
- Delaigle, A., & Hall, P. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 40(1), 322–352.
- Driessens, J., Maenhout, P. J., & Vilkov, G. (2009). The price of correlation risk: Evidence from equity options. *The Journal of Finance*, 64(3), 1377–1406.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer.
- Engl, H. W., Hanke, M., & Neubauer, A. (1996). *Regularization of inverse problems* (Vol. 375). Springer Science & Business Media.
- Fengler, M. R. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, 9(4), 417–428.
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: Theory and practice*. Springer Science & Business Media.
- Figlewski, S. (2010). Estimating the implied risk neutral density. In T. Bollerslev, J. Russell & M. Watson (Eds.), *Volatility and time series econometrics: Essay in honor of Robert F. Engle*. (pp. 323–353). Oxford University Press.
- Figlewski, S. (2018). Risk neutral densities: A review. *Annual Review of Financial Economics*, 10, 329–359.
- Flamouris, D., & Giamouridis, D. (2002). Estimating implied pdfs from American options on futures: A new semiparametric approach. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 22(1), 1–30.
- Gagliardini, P., & Scaillet, O. (2012). Nonparametric instrumental variable estimation of structural quantile effects. *Econometrica*, 80, 1533–1562.
- Garcia, R., & Gençay, R. (2000). Pricing and hedging derivative securities with neural networks and a homogeneity hint. *Journal of Econometrics*, 94(1–2), 93–115.

- Garcia, R., Luger, R., & Renault, E. (2003). Empirical assessment of an intertemporal option pricing model with latent variables. *Journal of Econometrics*, 116(1–2), 49–83.
- Glad, I., Hjort, N. L., & Ushakov, N. (2003). Correction of density estimators that are not densities. *Scandinavian Journal of Statistics*, 30, 415–427.
- Gourieroux, C., & Jasiak, J. (2001). *Financial econometrics: Problems, models, and methods*. Princeton University Press.
- Grith, M., Härdle, W., & Schienle, M. (2012). Nonparametric estimation of risk-neutral densities. In J.-C. Duan et al. (Eds.), *Handbook of computational finance* (pp. 277–304). Springer-Verlag.
- Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics*, 35(1), 70–91.
- Harrison, J. M., & Kreps, D. (1979). Martingales and arbitrage in multiperiod security markets. *Journal of Economic Theory*, 20, 381–408.
- Hartvig, N., Jensen, J., & Pedersen, J. (2001). A class of risk neutral densities with heavy tails. *Finance and Stochastics*, 5, 115–128.
- Horowitz, J. L. (2007). Asymptotic normality of a nonparametric instrumental variables estimator. *International Economic Review*, 48(4), 1329–1349.
- Ibragimov, M., Ibragimov, R., & Walden, J. (2015). *Heavy-tailed distributions and robustness in economics and finance* (Vol. 214). Springer.
- Jackwerth, J. C., & Rubinstein, M. (1996). Recovering probability distributions from option prices. *The Journal of Finance*, 51(5), 1611–1631.
- Jarrow, R., & Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3), 347–369.
- Jondeau, E., Poon, S.-H., & Rockinger, M. (2010). *Financial modeling under non-gaussian distributions*. Springer.
- Kargin, V., & Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99(10), 2508–2526.
- Kokoszka, P., & Zhang, X. (2010). *Improved estimation of the kernel of the functional autoregressive process* [Tech. Rep.]. Utah State University.
- Kundu, A., Kumas, S., & Tomar, N. (2018). Option implied risk-neutral density estimation: A robust and flexible method. *Computational Economics* 54, 705–728.
- Melick, W., & Thomas, C. (1997). Recovering an asset's implied pdf from option prices: An application to crude oil during the gulf crisis. *Journal of Financial and Quantitative Analysis*, 32, 91–115.
- Panigirtzoglou, N., & Skiadopoulos, G. (2004). A new approach to modeling the dynamics of implied distributions: Theory and evidence from The S&P 500 options. *Journal of Banking & Finance*, 28(7), 1499–1520.
- Park, J. Y., & Qian, J. (2012). Functional regression of continuous state distributions. *Journal of Econometrics*, 167(2), 397–412.
- Rosenberg, J. V. (1998). Pricing multivariate contingent claims using estimated risk-neutral density functions. *Journal of International Money and Finance*, 17(2), 229–247.
- Shimko, D., Teijima, N., & Van Deventer, D. R. (1993). The pricing of risky debt when interest rates are stochastic. *Journal of Fixed Income*, 3(2), 58–65.
- Souissi, N. (2017). The implied risk neutral density dynamics: Evidence from the S&P TSX 60 index. *Journal of Applied Mathematics*, 2017.
- Tian, Y. (2011). Extracting risk-neutral density and its moments from American option prices. *Journal of Derivatives*, 18, 17–34.
- Tsafack, I. (2020). *Intraday stock market forecasting via functional time series* [Working paper].
- Vogt, E. (2014). Option-implied term structures. Available at SSRN 2541954.
- Yatchew, A., & Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2), 579–599.

This page intentionally left blank

CHAPTER 6

ESTIMATING DIFFUSION MODELS OF INTEREST RATES AT THE ZERO LOWER BOUND: FROM THE GREAT DEPRESSION TO THE GREAT RECESSION AND BEYOND

Lealand Morin^a

^a*Department of Economics, University of Central Florida, United States*

ABSTRACT

The time series of the federal funds rate has recently been extended back to 1928, now including several episodes during which interest rates remained near the lower bound of zero. This series is analyzed, using the method of indirect inference, by applying recent research on bounded time series to estimate a set of bounded parametric diffusion models. This combination uncouples the specification of the bounds from the law of motion. Although Louis Bachelier was the first to use arithmetic Brownian motion to model financial time series, he has often been criticized for this proposal, since the process can take on negative values. Most researchers favor processes such as geometric Brownian motion (GBM), which remains positive. Under this framework, Bachelier's proposal remains valid when specified with bounds and is shown to compare favorably when modeling the federal funds rate.

Keywords: Federal funds target rate; interest rate; zero lower bound; diffusion processes; regulated Brownian motion; bounded processes

1. INTRODUCTION

Recent research has uncovered the historical series of the federal funds rate, back to 1928 – the earliest date that this series was recorded in newspapers. The extension of this data series affords the opportunity to study interest rates during another chapter of history in which the economy dipped into a severe recession. Both the Great Depression and, more recently, the Great Recession feature extended periods of time during which the federal funds rate remained near the zero lower bound. The current state of the economy during the coronavirus pandemic has again brought interest rates to the lower bound. This suggests that, although these events are rare, the benchmark interest rate can remain near zero for extended periods of time. Properly accounting for these episodes is especially important when discounting cash flows over a period of decades, for example, to evaluate securities or long-lived investment projects.

This research follows a long history of models for interest rates using a variety of diffusion models. The tendency of these series to remain near the zero lower bound suggests that a model should incorporate this characteristic. When used for this purpose, diffusion models typically impose a specification of variance that vanishes as the series approaches zero – to enforce the boundary – such as with GBM. In contrast, the historical record shows non-zero variance that characterizes the changes that, by constraint, are skewed upward, when the interest rate is near zero. In this chapter, I combine research on time series in the presence of bounds with estimation methods for diffusion models for the federal funds rate. This combination produces a model that remains above the bound of zero, without placing constraints directly on the specification of the drift and diffusion terms, in a way that adequately characterizes the movement in the federal funds rate.

The remainder of the chapter proceeds as follows. In the next section, I describe the historical series of the federal funds rate. Then, I outline a set of diffusion models appropriate for the features of the series. I also augment this set of models with the explicit specification of the zero lower bound, without imposing constraints on the parameters in the law of motion of the series. In the next section, I describe the empirical methodology of indirect inference, a method of simulated minimum distance (SMD), applied to this estimation problem. I present the empirical results in the following section and then draw conclusions.

2. FEDERAL RESERVE DATA

The sample of 24,121 daily observations of the federal funds rate spans the period April 4, 1928 to September 15, 2020. The data were obtained from two sources. For the period beginning July 1, 1954, the sample was drawn from a single series in the FRED database at the Federal Reserve Bank of St. Louis. Over this period, the series was collected by the Federal Reserve Bank of New York. Until recently, the earlier values of this series were not listed in any digital form.

Through the extensive data collection efforts of [Anbil et al. \(2020\)](#), the series has been extended back to 1928 by transcribing the rates published in printed copies of the *New York Herald Times* and the *Wall Street Journal*. From these reports, four series are available from the FRED database, specifically, the high and low reported federal funds rates from each of the two newspapers. These series were combined to extend the history of the federal funds rate to cover the nearly century-long period 1928 to 2020. The process of combining and cleaning the data is described in the Appendix and is further described in an appendix to [Anbil et al. \(2020\)](#), which is the documentation for the data source.

The entire series of the Federal Reserve rate is depicted in panel (a) of [Fig. 1](#). During the Great Depression and also the Great Recession, the prolonged sojourns at the lower bound of zero are clearly visible. In the histogram shown in panel (b), it is clear that the rate has been near the zero lower bound for several thousand days throughout the sample, which represents a considerable fraction of the observations. The distribution is skewed to the right, with a short visit to double-digit interest rates during the period of high inflation in the late 1970s and the early 1980s.

[Fig. 2](#) depicts level curves of kernel-smoothed density plots of the daily changes in the federal funds rate against the levels of the federal funds rate. The mode of the joint distribution appears at an interest rate just above 5 per cent, with daily changes lower than 10 basis points, centered just below zero. To illustrate the variability of changes in the series, panel (b) of [Fig. 2](#) depicts the kernel-smoothed density of absolute changes in the federal funds rate. When the federal funds

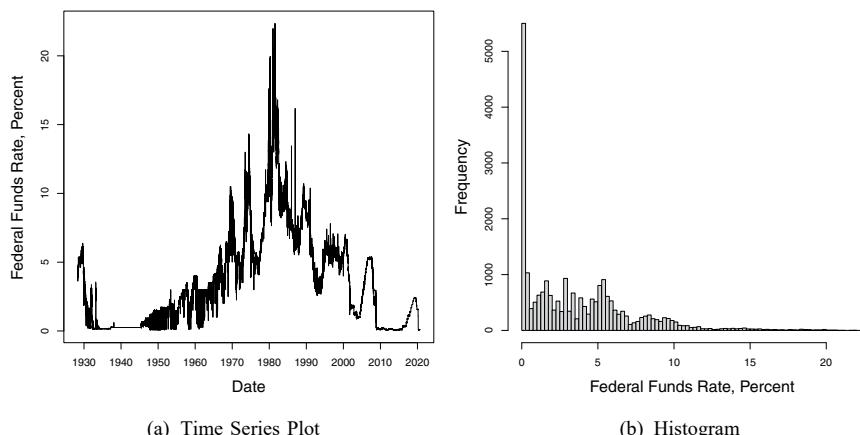


Fig. 1. Daily Series of the Federal Funds Rate, 1928–2020. (a) Time Series Plot and (b) Histogram. Notes: Periods in which the federal funds rate was near zero are clearly visible in panel (a) after the Great Depression and the Great Recession and,

more recently, during the coronavirus pandemic. These near-zero observations represent the mode of the distribution in the histogram in panel (b).

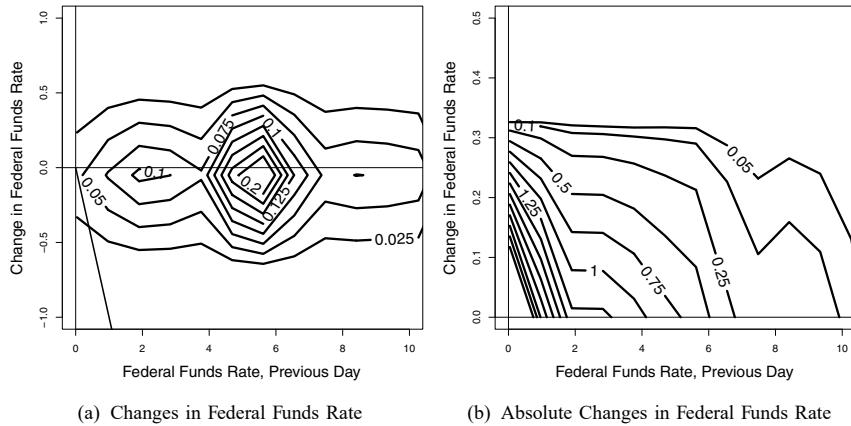


Fig. 2. Density Plots of Daily Changes in the Federal Funds Rate, 1928–2020.

(a) Changes in Federal Funds Rate and (b) Absolute Changes in Federal Funds Rate. *Notes:* Panel (a) depicts the level curves of the kernel-smoothed joint density of the daily changes in the federal funds rate against the level of the federal funds rate the previous day. Panel (b) depicts the level curves of the joint density of the absolute value of the daily changes against the level of the federal funds rate the previous day.

rate is lower, the variability tends to be concentrated around smaller changes. The changes in the series are smaller near the zero lower bound, with the mode of the joint distribution near the origin. Nevertheless, the series exhibits daily variation of up to 10 basis points when the rate is near zero, which is a large percentage change of the interest rate.

Fig. 3 depicts two histograms of the changes in the federal funds rate divided into two subsamples. Panel (a) of Fig. 3 depicts the histogram from the sample that includes only the days in which the federal funds rate was 25 basis points or less the previous day. The histogram in panel (b) covers the remainder of the sample. An important feature of the series, with regards to the specification of an econometric model, is that the series is highly skewed to the right when the series is near zero. With the constraint of the zero lower bound there is little else that can happen. In contrast, in the histogram in panel (b), the distribution appears symmetric. The distribution also appears leptokurtic, with long tails and a sharp peak at zero. This is partly an artifact of the large rate changes that occurred when interest rates were very high, during the 1970s and 1980s. These findings suggest that the variability of changes in the series should be modeled as an increasing function of the current rate. In addition, the econometric model should be specified with a non-negligible degree of variability at low interest rates as well.

Fig. 4 presents a rudimentary attempt at modeling the changes in the interest rate as a function of the level. The daily changes in the rate are shown in

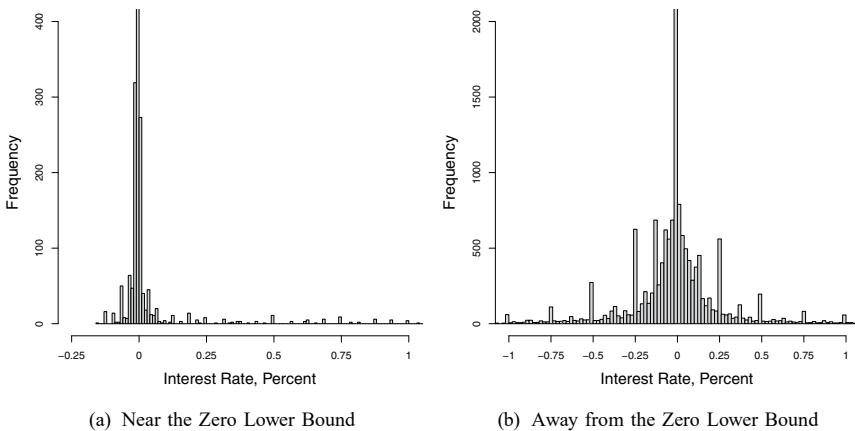


Fig. 3. Histograms of Daily Changes in the Federal Funds Rate, 1928–2020.

(a) Near the Zero Lower Bound and (b) Away from the Zero Lower Bound.

Notes: The sample was divided into two parts based on whether the federal funds rate was above or below 25 basis points the day before each observation. On the days after the federal funds rate was greater than 25 basis points, in panel (b), the distribution appears symmetric, with some point masses on changes that are multiples of 25 basis points. In the rest of the sample, in panel (a), when the series was within 25 basis points of the zero lower bound, the series is skewed to the right but displays a similar degree of variability.

a pair of scatter graphs, with the changes in the interest rate in panel (a) and the absolute changes in panel (b), both of which are plotted against the level of the federal funds rate the previous day. The dashed lines represent the 10-, 50- and 90-per cent quantiles of the changes in the interest rate. In both panels, the dashed lines of quantiles were drawn with observations grouped into intervals 50 basis points wide.

Much of the variability remains within 25 basis points for interest rates as high as 10 per cent, with the variability gradually increasing in this range. The variability appears much higher in the sparsely-populated region in which the federal funds rate is above 10 per cent. For interest rates arbitrarily close to the lower bound of zero, the variability appears to vanish. The variance increases sharply, however, as soon as the federal funds rate is above the zero lower bound.

In panel (a), the restriction to the zero lower bound is represented by a diagonal line with a slope of negative one. The restriction to the zero lower bound is clearly visible, with all observed changes restricted above the line where the maximum change equals the negative of the current interest rate. In an empirical specification for this series, the variability of the process should increase sharply for low values of the interest rate. Above an interest rate of 2 per cent, the rate of variability should increase gradually. These characteristics will guide the specification of the set of diffusion models described in the next section.

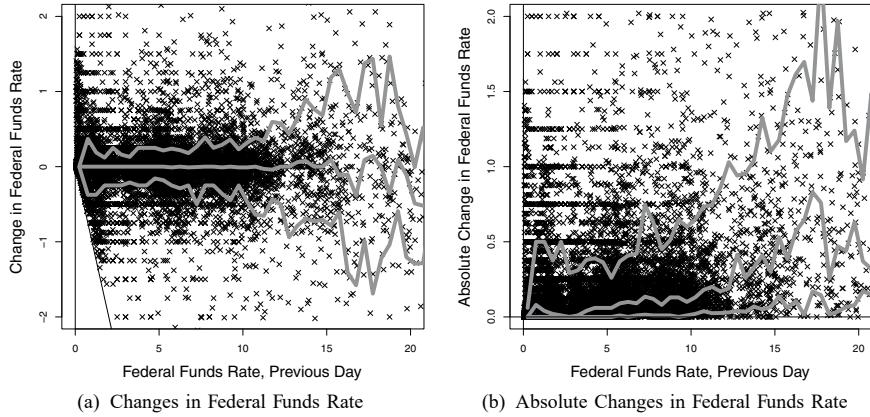


Fig. 4. Quantiles of Daily Changes in the Federal Funds Rate, 1928–2020.

(a) Changes in Federal Funds Rate and (b) Absolute Changes in Federal Funds Rate. *Notes:* Panel (a) depicts a scatter graph of the daily changes in the federal funds rate plotted against the federal funds rate on the previous day. Panel (b) depicts a plot of the absolute value of the daily changes against the federal funds rate on the previous day. The 10-, 50-, and 90-per cent quantiles are superimposed over these plots, indicated by the grey lines. These figures indicate a steep increase in variability of the series for small values of the federal funds rate, followed by a curve that remains fairly flat up to a federal funds rate of 7 per cent. A gradual rise in variability appears as the federal funds rate rises above 10 per cent.

3. DIFFUSION MODELS OF THE INTEREST RATE

The history of diffusion models extends back as far as the history of the Federal Reserve. [Bachelier \(1900\)](#) was the first to apply the stochastic process today referred to as Brownian motion – using the process as a model of prices in financial markets. Soon after, [Einstein \(1905\)](#) made a pioneering effort to characterize Brownian motion formally – using the botanist Robert Brown’s observation of the motion of particles suspended in a liquid as a metaphor, and expanded on the topic in [Einstein \(1926\)](#). [Wiener \(1923\)](#) built a rigorous theoretical foundation, including a statement of the conditions for the existence of Brownian motion, which was named a Wiener process in his honor in some parts of the literature.

A few decades later, [Itô \(1944\)](#) produced what would later be referred to as the Itô Formula, which greatly expanded the set of processes that could be applied to phenomena measured in continuous time. This modeling tool eventually led to an explosion of continuous-time models in finance, especially after [Black and Scholes \(1973\)](#) assumed GBM for the law of motion of securities and portfolios that were valued with their celebrated Black–Scholes option pricing formula. This body of theoretical work was consolidated into many references that are well-known for applications in economics and finance – including [Duffie \(2001\)](#), [Cochrane \(2005\)](#), and [Shreve \(2004\)](#). More generally, [Karlin and Taylor \(1981\)](#) as

well as [Karatzas and Shreve \(1991\)](#) all presented modern introductions to diffusion models.

Although the models have become commonly used to evaluate securities, it is an altogether different matter to estimate the parameters of these models. The technique proposed by [Äit Sahalia \(2002\)](#) is a popular method for an approximate maximum-likelihood estimator (MLE). Phillips and Yu (2009) have also provided a survey of various approaches to likelihood-based estimation of diffusion models. Contributing to this literature, [Äit Sahalia and Park \(2012\)](#) have developed specification tests for diffusion models. More recently, [Choi et al. \(2014\)](#) have developed a framework for model selection among different models of diffusion. [Kim and Park \(2017\)](#) also considered the special case of recurrent diffusions, with an application to high-frequency regression. Using a nonparametric approach, [Äit Sahalia and Park \(2016\)](#) considered an application to nonstationary continuous-time models.

A similar alternative to the estimation of diffusion models involves using a time-series model with time-varying coefficients, such as in [Park and Hahn \(1999\)](#) as well as [Cai et al. \(2009\)](#). Ever since [Engle \(1982\)](#), much attention has been paid to models of time-varying and nonlinear heteroskedasticity. Notable examples include [Han and Park \(2008\)](#) as well as [Chung and Park \(2007\)](#), and Park (2002).

Models with time-varying coefficients have a long history in the macroeconomic literature, especially in term-structure models of the interest rate. This line of research dates back to [Cox et al. \(1981\)](#), and the expectations hypothesis of the term structure of interest rates. This was followed by [Campbell \(1986\)](#) as well as [Campbell and Shiller \(1991\)](#), who provided an early prescription such that default-risk-free zero-coupon bonds could be valued based on expected returns with term premiums that are constant through time. Within this framework, affine term-structure models have been the focus of attention because of their analytical tractability. [Duffie \(2001\)](#) has presented an authoritative treatment of asset pricing and term-structure modeling.

Early examples of interest rate models included the foundational single-factor models pioneered by [Vasicek \(1977\)](#) and [Cox et al. \(1985\)](#). [Vasicek \(1977\)](#) proposed an Ornstein–Uhlenbeck process, also called a mean-reverting process, which is a model having the following form:

$$dr_t = \kappa(\rho - r_t)dt + \sigma dW_t, \quad (1)$$

where ρ is the unconditional mean and κ is known as the speed of mean reversion. [Cox et al. \(1985\)](#) proposed a modification to this model, in the form of a state-dependent rate of diffusion

$$dr_t = \kappa(\rho - r_t)dt + \sigma\sqrt{r_t}dW_t. \quad (2)$$

This formulation is sometimes known as a square-root process. The mean-reverting characteristic of these models make for an interesting comparison, when analyzed in combination with a series constrained by bounds. These models fit within a broader set of models estimated in this chapter.

More generally, a diffusion process X_t is a continuous-time process defined for each $t \in \mathbb{R}_+$. The instantaneous change in X_t satisfies the stochastic differential equation

$$dX_t = \mu(X_t; \boldsymbol{\alpha})dt + \sigma(X_t; \boldsymbol{\beta})dW_t, \quad (3)$$

where W_t is a standard Brownian motion, $\mu(X_t; \boldsymbol{\alpha})$ is the drift function with parameter $\boldsymbol{\alpha}$, and $\sigma(X_t; \boldsymbol{\beta})$ is the diffusion function with parameter $\boldsymbol{\beta}$. Collect the parameters into one vector $\boldsymbol{\theta} = [\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top \in \Theta$ and define \mathcal{D} as the domain of the diffusion process. This definition includes the following examples, each of which has parameter values that lie within a convex parameter space Θ .

Example 1. Some examples of diffusion processes include

- Arithmetic Brownian Motion (ABM):
- $\mu(X_t; \boldsymbol{\alpha}) = \alpha_I$ and $\sigma(X_t; \boldsymbol{\beta}) = \beta_I$ with $\beta_I > 0$ and $\mathcal{D} = \mathbb{R}$.
- Geometric Brownian Motion (GBM):
- $\mu(X_t; \boldsymbol{\alpha}) = \alpha_X X_t$ and $\sigma(X_t; \boldsymbol{\beta}) = \beta_X X_t$ with $\beta_X > 0$ and $\mathcal{D} = \mathbb{R}_+$.
- Ornstein–Uhlenbeck Process (OU):
- $\mu(X_t; \boldsymbol{\alpha}) = \alpha_I + \alpha_X X_t$ and $\sigma(X_t; \boldsymbol{\beta}) = \beta_I$ with $\alpha_X < 0, \beta_I > 0$ and $\mathcal{D} = \mathbb{R}$.
- Square-Root Process (SQR):
- $\mu(X_t; \boldsymbol{\alpha}) = \alpha_I + \alpha_X X_t$ and $\sigma(X_t; \boldsymbol{\beta}) = \beta_X \sqrt{X_t}$ with $\alpha_X < 0, \beta_X > 0$ and $\mathcal{D} = \mathbb{R}_+$.

In the above notation, parameters are assigned subscripts according to the term multiplied by the parameter: α_I is an intercept, β_X is a slope coefficient and β_E appears in the exponent of X_t .

The set of models presented above includes the models proposed by [Vasicek \(1977\)](#) and [Cox et al. \(1985\)](#). In particular, the mean-reverting drift function $\kappa(\rho - r_t)$ is equivalent to $\mu(X_t; \boldsymbol{\alpha}) = \alpha_I + \alpha_X X_t$, in the OU and SQR models shown above. The parameter α_I is the negative of the speed of mean reversion κ . The unconditional mean ρ is captured by the ratio $\alpha_I / \kappa = -\alpha_I / \alpha_X$.

The observed interest rate, denoted by r_t in the literature, takes the place of X_t in the notation above and is the variable of interest in this chapter. In the usual framework, the process X_t is free to move throughout \mathcal{D} , the domain of the diffusion process. In this sense, the process is already bounded by its domain. In this chapter, however, the focus is on the interest rate r_t , which is bounded from below by zero. Several models listed above already impose this constraint: with GBM and the SQR model, the specification of the diffusion function dictates that the rate of diffusion vanishes as the process approaches zero, keeping the process above the lower bound of zero. The OU process and ABM do not restrict the process to positive values, which is a common criticism of [Bachelier \(1900\)](#) for modeling stock prices. I augment the menu of stochastic processes in a way that decouples the specification of the process from the restriction to the region above the lower bound of zero, leaving open the possibility of exploring the full set of processes to model interest rates.

Researchers have developed a budding literature concerned with statistical methodologies for estimating bounded processes, with [Cavaliere \(2005\)](#) as well as

[Cavaliere and Xu \(2014\)](#) having highlighted key examples in econometric theory. This literature built on the foundational treatment of [Harrison \(1985\)](#), setting the stage for the theoretical analysis of stochastic processes within bounds, with the primary example referred to as regulated Brownian motion. Specifically, a realization of regulated Brownian motion X_t is constrained to remain within the interval $[b, \bar{b}]$ at all times t . In the case of interest rates studied in this chapter, only the lower bound applies with $b = 0$ and the upper bound is not binding. The theory presented in [Morin \(2017\)](#) built on this body of knowledge by merging this line of research with the maximum-likelihood methodology developed in [Jeong \(2008\)](#) as well as [Jeong and Park \(2013\)](#) to estimate parametric diffusion models. In this chapter, I take a different approach – implementing a variant of indirect inference, following [Gouriéroux et al. \(1993\)](#), except with bounds imposed.

This combination opens up the set of models available to the applied researcher studying variables that lie within bounds. In the absence of an explicit specification of the bounds, modelers are often constrained to choose a model specification that enforces the bound. As mentioned above, the models GBM and SQR satisfy this criterion, since the rate of diffusion vanishes as the process nears zero. This specification might be in conflict with the reality of a non-zero rate of diffusion near the bound, as was documented above in the analysis of the federal funds rate. Using the flexibility offered by the explicit modeling of the bounds, however, there is no need to be constrained to such a specification. In particular, it is possible to include a positive intercept in the diffusion function, specifying a positive rate of diffusion near the lower bound of zero, matching the conditions observed in the historical series of interest rates. Thus, the processes listed above are augmented in this chapter to allow for a nonzero constant at the lower bound of zero by appending the following set of models.

Example 2. The augmented versions of selected diffusion processes are as follows:

- Augmented Geometric Brownian Motion (GBM-I):
 $\mu(X_t; \alpha) = \alpha_x X_t$ and $\sigma(X_t; \beta) = \beta_x X_t + \beta_I$ with $\beta_I > 0, \beta_x > 0$ and $\mathcal{D} = \mathbb{R}$.
- Augmented Ornstein–Uhlenbeck Process (OU-X):
 $\mu(X_t; \alpha) = \alpha_I + \alpha_x X_t$ and $\sigma(X_t; \beta) = \beta_I + \beta_x X_t$ with $\alpha_2 < 0, \beta_I > 0, \beta_x > 0$ and $\mathcal{D} = \mathbb{R}$.
- Augmented Square-Root Process (SQR-I):
 $\mu(X_t; \alpha) = \alpha_I + \alpha_x X_t$ and $\sigma(X_t; \beta) = \beta_x \sqrt{X_t} + \beta_I$ with $\alpha_x < 0, \beta_I > 0, \beta_x > 0$ and $\mathcal{D} = \mathbb{R}$.

In the naming convention for the augmented models, the hyphenated name has a suffix that indicates the subscript of the parameter appended to the model. Notice that the domain of these processes is no longer restricted to the positive real line, since the rate of diffusion no longer vanishes toward zero. These processes only remain within bounds when regulated to remain above zero, so these processes are now candidates for modeling the interest rate in the bounded specification presented in this chapter.

This list of models represents a first step toward a model of the term structure that incorporates the constraints imposed by the zero lower bound. This approach does sacrifice accuracy by using a single-factor model, since it is well-known that up to three factors will improve prediction, as shown by [Littermann and Scheinkman \(1991\)](#). These factors correspond to the level, the slope, and the curvature of the yield curve. In this chapter, I restrict attention to the level, which is of first-order importance, especially considering the proximity to the zero lower bound. The hypothesis tested here is that a one-factor model that accounts for the zero lower bound will be better-suited than a one-factor model without, for modeling interest rates after the Great Recession and the Great Depression. A question left for further research regards the comparison of performance between a multifactor model with the zero lower bound, versus one that ignores the bounds. The results presented below support this approach as a promising avenue of research. This comparison is made possible using the econometric methodology presented next.

4. INDIRECT INFERENCE

Although the exact likelihood functions are known for the unbounded ABM, GBM, and OU processes, those for the others are not. As a substitute, the likelihood function of the corresponding discrete-time process is used to approximate the likelihood function of each continuous-time model. Under such an approach, however, the MLE has an additional source of bias, often referred to as the convexity effect, which obtains because under the approximate approach the drift and the diffusion functions are assumed constant over the intervals of time between observations. Because the exact discretizations of the processes with bounds are unknown, for all stochastic processes that I considered (except for the ABM process), an alternative estimation strategy was required.

[Gouriéroux et al. \(1993\)](#) estimated the GBM, OU, and, by extension, the ABM processes, as applications of indirect inference. Indirect inference was first introduced by [Smith \(1990\)](#), and in [Smith \(1993\)](#), before it was generalized by [Gouriéroux et al. \(1993\)](#), and later refined by [Gallant and Tauchen \(2006\)](#) as well as [Gouriéroux et al. \(2010\)](#) and, more recently, by [Bruins et al. \(2018\)](#). [Hall and Rust \(2021\)](#) employed the method in a time-series model and introduced the term SMD estimation to describe this procedure. As [Guvenen and Smith \(2014\)](#) have noted, indirect inference is very useful when estimating models for which the likelihood function or the criterion function to be optimized is analytically intractable, or too computationally burdensome to evaluate. This method is admissible here because all the processes that I seek to estimate can be simulated – both with and without the zero lower bound imposed.

The essence of indirect inference is to use an auxiliary model, which is easy to compute, to capture features of the data and then to form a criterion function on which to base the estimation. The mapping from the structural parameters of interest to the auxiliary parameters that optimize this auxiliary model define a *binding function* through which one can indirectly draw inference on the structural

parameters in the model of interest. In this case, following the approach of Gouriéroux et al. (1993), I used the logarithm of the likelihood function for the corresponding discrete-time process as the objective function for the auxiliary model. In the notation from above, this objective function is

$$\begin{aligned} Q_T(\boldsymbol{\alpha}, \boldsymbol{\beta}; \{X_t\}_{t=1}^T) = & \frac{1}{T} \sum_{t=2}^T -\log[\sigma(\boldsymbol{\beta}; X_{t-1})] - \\ & \frac{1}{2T} \sum_{t=2}^T \frac{[X_t - X_{t-1} - \mu(\boldsymbol{\alpha}; X_{t-1})]^2}{\sigma(\boldsymbol{\beta}; X_{t-1})^2}. \end{aligned} \quad (4)$$

This objective function is a suitable choice for an auxiliary model because there exists a clear correspondence between the structural parameters $\boldsymbol{\theta}^\top = [\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]$ that generate the simulated series $\{X_t^*\}_{t=1}^T$ and the auxiliary parameters $\tilde{\boldsymbol{\theta}}^{*\top} = [\tilde{\boldsymbol{\alpha}}^{*\top}, \tilde{\boldsymbol{\beta}}^{*\top}]$ that optimize the objective function $Q_T(\boldsymbol{\theta}, \{X_t^*\}_{t=1}^T)$. This optimization defines a binding function $\mathbf{b}: \Theta \rightarrow \Theta$ defined as $\mathbf{b}(\boldsymbol{\theta}_0) = (\tilde{\boldsymbol{\theta}}_0)$, where $\tilde{\boldsymbol{\theta}}_0$ is the vector of the auxiliary parameters that optimizes the following limiting objective function:

$$Q_\infty(\boldsymbol{\theta}_0; \{X_t\}_{t=1}^T) = \lim_{T \rightarrow \infty} Q_T(\boldsymbol{\theta}_0; \{X_t\}_{t=1}^T).$$

A necessary condition for estimation is that $\mathbf{b}(\cdot)$ is one-to-one, so the matrix of partial derivatives $\partial \mathbf{b}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$, evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, must have full column rank; see Gouriéroux et al. (1993).

The estimated structural parameters in $\boldsymbol{\theta}$ is the vector $\hat{\boldsymbol{\theta}}$ that generates the series $\{X_t^*\}_{t=1}^T$ with auxiliary parameter estimates $\tilde{\boldsymbol{\theta}}^*$ that closely match the auxiliary parameters $\tilde{\boldsymbol{\theta}}$, which are estimated from the observed interest-rate series. The formulation in this chapter is just identified because a one-to-one mapping exists between the structural parameters and the auxiliary parameters. Here, the estimated structural parameters $\hat{\boldsymbol{\theta}}$ produce a series $\{X_t^*\}_{t=1}^T$ with estimates of auxiliary parameters $\tilde{\boldsymbol{\theta}}^*$ that exactly match the auxiliary parameters $\tilde{\boldsymbol{\theta}}$ obtained from the observed interest-rate series. The estimation procedure is summarized in the following steps:

1. Calculate the auxiliary parameters from the data to obtain $\tilde{\boldsymbol{\theta}}$, which is a vector of the parameter estimates from optimizing the objective function for the auxiliary model equation (4) evaluated using the observed interest-rate series $\{X_t\}_{t=1}^T$.
2. Conduct a numerical optimization to search over candidate values of $\hat{\boldsymbol{\theta}}$ by iterating over the following steps:
 - (a) draw a realization of the series $\{X_t^*\}_{t=1}^T$, given the candidate parameter $\hat{\boldsymbol{\theta}}$, following the law of motion specified by the diffusion model, and conforming to the zero lower bound, if imposed;
 - (b) calculate the auxiliary parameters $\tilde{\boldsymbol{\theta}}^*$ from the simulated series $\{X_t^*\}_{t=1}^T$, following the same procedure as in step 1 to optimize the objective function for the auxiliary model equation (4);
 - (c) calculate the weighting matrix $\boldsymbol{\Omega}$, described below, to normalize the auxiliary parameters;

- (d) calculate the weighted distance between the auxiliary parameters from the observed interest-rate series and the simulated data

$$(\tilde{\theta}^* - \tilde{\theta})^\top \Omega^{-1} (\tilde{\theta}^* - \tilde{\theta}). \quad (5)$$

3. Minimize the weighted distance by iterating over steps 2 (a)–(d), to find the parameter $\hat{\theta}$ that minimizes the distance criterion in equation (5).

Given a discrete series $\{X_t\}_{t=1}^T$, the objective function of equation (4) in steps 1 and 2(b) can be optimized using standard numerical methods, such as quasi-Newton methods. To ease the computational burden, I concentrate out the parameter α in $\mu(\alpha; X_t)$ with a weighted least-squares approach, with weights $\sqrt{\sigma(\beta; X_t)}$ to obtain the optimal value of the auxiliary parameter $\tilde{\alpha}$ for a candidate value of $\tilde{\beta}$. Then I conduct the outer numerical optimization over the space of the auxiliary parameter β to obtain the estimate $\tilde{\beta}$.

To generate the simulated series $\{X_t^*\}_{t=1}^T$ in step 2(a), I generated a realization of the discretized process with a small time step: the complete series has two million observations and I dropped 99 observations between each consecutive pair of the remaining 20,000 observations, roughly matching the 24,121 observations in the sample.

The models I estimated are pure time-series models; no exogenous variables were included. [Gouriéroux et al. \(1993\)](#) showed that the optimal weighting matrix Ω^* has the form $\mathbf{J}_0 \mathbf{V}_0^{-1} \mathbf{J}_0$. The matrix \mathbf{J}_0 is consistently estimated by

$$-\frac{\partial^2 Q_T}{\partial \theta \partial \theta^\top}(\tilde{\theta}; \{X_t\}_{t=1}^T), \quad (6)$$

which is calculated by taking numerical derivatives of the objective function with the series $\{X_t\}_{t=1}^T$ generated at the parameter estimates $\hat{\theta}$. Since the objective function Q_T can be written as a sum of contributions to the objective function $\sum_{t=2}^T q_t(\theta; X_t, X_{t-1})$,

$$\mathbf{V}_0 = \lim_{T \rightarrow \infty} \mathbb{V} \left[\frac{1}{\sqrt{T}} \sum_{t=2}^T \frac{\partial q_t}{\partial \theta}(\theta; X_t, X_{t-1}) \right], \quad (7)$$

which I estimated using the approach of [Newey and West \(1987\)](#), as outlined by [Gouriéroux et al. \(1993\)](#). The estimates and standard errors obtained from this procedure are documented in the next section.

5. EMPIRICAL RESULTS

For the first step in the estimation procedure, I estimated the auxiliary parameters from the observed series of the federal funds rate. These estimates are shown in [Table 1](#).¹ For all models, the drift functions are negatively sloped, with a daily reduction on the order of half of a basis point for each percent interest rate. In terms of the diffusion rate, the estimates range between a constant rate of 35 basis points per day, for the OU and ABM models, to a constant slope of

Table 1. Estimated Auxiliary Parameters.

	α_I	α_X	β_I	β_X	β_E
ABM	-0.000159	0.0	0.3500	0.0	0.0
GBM	0.0	-0.000043	0.0	0.6493	1.0
OU	0.017460	-0.004725	0.3495	0.0	0.0
SQR	0.021664	-0.005451	0.0	0.2748	0.5
GBM-C	0.0	-0.001737	0.1998	0.0334	1.0
OU-X	0.016417	-0.004445	0.1990	0.0336	1.0
SQR-C	0.018110	-0.004781	0.1284	0.1174	0.5
FULL	0.018039	-0.004769	0.1331	0.1116	0.5192

Notes: The auxiliary parameters were estimated by optimizing the auxiliary model using the entire series of the federal funds rate from 1928 to 2020. In the models in which some parameter values are implied by the specification of the model, those values are shown in the table alongside the estimates from the other models, with minimal significant digits.

64 basis points for each percentage point of interest, for GBM. For the models with non-zero intercepts, the intercept ranges from thirteen to 20 basis points, with a slope coefficient of 11 down to 3 basis points per day times an exponent of the interest rate.

Table 2 collects the estimates from the models with series that are regulated to remain above zero. Under this approach, I used these series to calculate auxiliary parameters and solved for the values of the structural parameters so that the auxiliary parameters from the simulated series matched those in Table 1.

The ABM process is the simplest model, which was estimated to have a constant downward drift of 6 basis points per day, however, this number is not

Table 2. Estimated Structural Parameters (Zero Lower Bound Imposed).

Model	α_I	α_X	β_I	β_X	β_E	$\Pr(\chi^2_{5-k} > d)$
ABM	-0.0611 (0.5843)	0.0	0.3801 (0.0030)	0.0	0.0	0.9348
GBM	0.0	0.1853 (0.5150)	0.0	0.5328 (0.0905)	1.0	0.0000
OU	-0.0215 (0.0145)	0.0005 (0.0011)	0.3593 (0.0027)	0.0	0.0	0.9373
SQR	0.0225 (0.0008)	-0.0070 (0.2774)	0.0	0.5050 (0.0044)	0.5	0.1009
GBM-C	0.0	-0.0021 (0.0012)	0.2249 (0.0034)	0.0315 (0.0009)	1.0	0.9849
OU-X	-0.0006 (0.0131)	-0.0020 (0.0012)	0.2113 (0.0033)	0.0309 (0.0009)	1.0	0.8845
SQR-C	-0.0109 (0.0844)	-0.0007 (0.0035)	0.2094 (0.0153)	0.0863 (0.0064)	0.5	0.8748

Notes: The estimates of the structural parameters solve for equality between the auxiliary parameters from the observed federal funds rate and those estimates from the series generated from each model, with the realization regulated to remain above zero. The column labeled $\Pr(\chi^2_{5-k} > d)$, lists the *p*-value for a test of the restrictions to the model in each row, each of which is a special case of the full model. The degrees of freedom for this test is $5 - k$, where k is the number of parameters in each model, compared to five parameters in the full model.

statistically significant, since the objective function is very flat in this region, with many drift values mapped to similar auxiliary parameters. Aside from the SQR and GBM processes, the remaining processes also have negative drift functions over the observed range of the interest rate, however, nearly all of those coefficients in those drift functions are statistically insignificant. The SQR process differs in that the drift has the mean-reverting characteristic, with a negative drift above around 3.2 per cent. For the GBM, a positive drift was predicted.

The most important differences were found in the specification of the diffusion function. The ABM and OU processes, each with a constant drift function, have a value near the estimated auxiliary parameter values in the neighborhood of 35 basis points per day. The GBM-C and OU-X models, each of which have a linear diffusion function, have an estimated intercept near 20 basis points per day, and a slope coefficient that increases three basis points per day for each percentage point of the federal funds rate. The GBM and SQR models from [Table 3](#), both of which are missing the intercept in the diffusion function, have much higher values of the slope coefficient β_x of 50 basis points per day, for each percentage point of the interest rate. These steeper slopes are required to fit the noticeable variation in the interest rate when it is close to zero, however, the slope overestimates the variation in the rate of diffusion when the interest rate is higher. The SQR-C model accommodates both the non-zero intercept and the declining rate of increase in the diffusion rate for higher values of the interest rate. In the SQR-C model, the intercept of the diffusion rate is similar to that of the other models augmented with an intercept, with a value of 20 basis points per day at an interest rate of zero. As the interest rate rises, the rate of diffusion rises at a rate of eight basis points times the square root of the federal funds rate, measured in percentage points.

To test for differences between the goodness of fit of the models, I calculated the weighted distance of the auxiliary parameters from those from the full model. All the models are nested in the full model that has all five of the parameters, with drift function $\mu(\alpha; X_t) = \alpha_I + \alpha_X X_t$ and diffusion rate $\sigma(\beta; X_t) = \beta_I + \beta_X X_t^{\beta_E}$, in which the exponent can be estimated. To facilitate a comparison between the models, I calculate the minimized distance between the auxiliary parameters for each of the models and those from the full model. The results appear in the column labeled $\text{Pr}(\chi_{5-k}^2 > d)$, which is the p -value of a test of the restrictions to the model in each row, each of which is a special case of the full model. The degrees of freedom for this test is $5 - k$, where k is the number of parameters in each model, compared to five parameters in the full model.

The results are split into a several-way tie. With the exception of the GBM model, the p -values are greater than 10 per cent, indicating that the models have similar predictions. By definition, the p -values are decreasing in the distance from the auxiliary parameters from the full model. The two models with the farthest distance from the full model – GBM and SQR – are the two models with a vanishing rate of diffusion at the zero lower bound. The remaining models have non-zero intercepts in the diffusion function and the performance is similar whether or not the diffusion rate is increasing in the level of the interest rate. With non-zero rates of diffusion at zero, all of these other models have the potential to cross

below the zero lower bound. This suggests that the most important model specification decision is to choose a regulated process with a strictly positive rate of diffusion. This matches the findings in Figs. 3 and 4, in which the federal funds rate was shown to exhibit a substantial degree of variation near the zero lower bound, as well as relatively flat quantiles of absolute changes in the range from 0 to 7 per cent. Furthermore, the simplest model, the ABM, which is a regulated Brownian motion when combined with bounds, appears to fit the data well enough without the added complexity of the other alternatives.

With the exception of GBM and SQR, the remaining models will produce series that can move below the zero lower bound, if not regulated to remain above the bound. Table 3 collects the estimates ignoring the fact that the interest rate series must remain above zero. Under this approach, the series $\{X_t^*\}_{t=1}^T$ is calculated by strictly following the law of motion with no adjustments. Strictly speaking, all of these models are mis-specified, since they can produce series with negative values. I still investigated these models, however, to compare the coefficients with those from the corresponding bounded processes.

The diffusion functions all exhibit a similar change with this change in specification. The slope coefficients β_x are higher for all three models when bounds are ignored. Furthermore, the intercept terms β_I are lower for all models. This suggests a lower rate of diffusion for lower interest rates, which appears to be a symptom of the censoring from the zero lower bound.

The drift functions changed predictably as well. The drift slope coefficients α_x moved further into negative territory. The intercepts α_I switched to positive values for all but the ABM. The constant drift of the ABM process changed from a large negative value to a negative value near zero. Overall, the pattern suggests a switch between zero-drift or mean-reverting processes, without bounds, and processes with negative drift pushing toward a lower bound.

Table 3. Estimated Structural Parameters (Zero Lower Bound Ignored).

Model	α_I	α_x	β_I	β_x	β_E	$\Pr(\chi_{5-k}^2 > d)$
ABM	-0.0002 (0.0034)	0.0	0.3492 (0.0024)	0.0	0.0	0.7721
OU	0.0209 (0.0052)	-0.0057 (0.0009)	0.3498 (0.0024)	0.0	0.0	0.8223
GBM-C	0.0	-0.0022 (0.0020)	0.2146 (0.0030)	0.0338 (0.0008)	1.0	0.9185
OU-X	0.0176 (0.0079)	-0.0050 (0.0017)	0.2112 (0.0034)	0.0319 (0.0009)	1.0	0.7665
SQR-C	0.0208 (0.0401)	-0.0058 (0.0116)	0.1859 (0.0118)	0.0993 (0.0059)	0.5	0.8620

Notes: The estimates of the structural parameters solve for equality between the auxiliary parameters from the observed federal funds rate and those estimates from the series generated from each model, with no restriction of the series. The column labeled $\Pr(\chi_{5-k}^2 > d)$, lists the *p*-value for a test of the restrictions to the model in each row, each of which is a special case of the full model. The degrees of freedom for this test is $5 - k$, where k is the number of parameters in each model, compared to five parameters in the full model.

In terms of the apparent accuracy of the estimates, the standard errors are smaller, for all but α_x , when the bounds are ignored. Most parameters show a small change in the magnitude of the standard errors, however, the change is particularly large for the intercept of the drift function. The presence of the zero lower bound, when it is ignored in the estimation, gives the illusion that the process is better behaved, with less variability in the estimates of the coefficients in the models.

6. CONCLUSION

In this chapter, I have followed a long history of applications modeling interest rates with diffusion models. This history also includes prolonged periods in which the interest rate remains near the zero lower bound. Thanks to an exhaustive data-collection effort by [Anbil et al. \(2020\)](#), it is now possible to combine these to produce a model that matches this characteristic of the interest rate over long periods of time. I applied recently-developed econometric techniques for bounded processes to estimate simple models of the federal funds rate in the presence of the zero lower bound. I estimated a set of diffusion models and demonstrated the difference in interpretation between the cases with and without the bounds explicitly incorporated into the estimation technique.

When the bounds are taken into account, the estimated drift has a greater negative slope and is higher near the zero lower bound, and the estimated rate of diffusion is also higher near the bound. When ignored, the zero lower bound created the false impression of a mean-reverting process, compared to a process with negative drift moving toward the lower bound, when the bound was taken into account. Ignoring the bound also created the illusion that the coefficients were more precisely estimated. The series appeared more variable when the lower bound was acknowledged, which avoided mistaking a process stuck near bounds for one that is less variable. Although the differences in the estimated drift functions were not statistically significant, the differences in the diffusion functions were more important. The imposition of the lower bound allows greater flexibility for model specification, particularly in achieving a non-zero diffusion rate at the zero lower bound, which matches the behavior observed in the historical series of the federal funds rate.

In this chapter, I have addressed a single issue in a way that has not appeared in the literature. Other specifications exist as well: considering them would be fruitful avenues for future research. In particular, one could extend this analysis to include discontinuities in the series, since changes of this type tend to occur after FOMC meetings. The fit of the model could be considerably improved by introducing a more detailed specification on those dates. A multivariate model that takes the information available at the time of the meeting is an option that also lies outside of the univariate framework presented here. Clearly, the members of the Federal Reserve Board respond to factors that measure the state of the economy, in addition to subjective conditions, such as investor sentiment. The analysis of these rate changes is certainly more complicated than a univariate model

would permit. Still, the approach taken here factors in the zero lower bound, which is shown to have a notable effect on the estimation results. This work is one step toward a multivariate framework that can fruitfully answer such questions. [Morin and Shang \(2020\)](#) have taken another step in this direction.

In hindsight, the current economic crisis during the coronavirus pandemic is not an unprecedented event. The experience during the Great Recession has shown that the economy can return to normal after a visit to the zero lower bound. With the benefit of further hindsight, the experience during the Great Depression has shown that these episodes can be prolonged for many years and may be an expected part of the experience to be considered when evaluating investments with a lifespan of decades. The approach taken in this chapter provides an empirical model of interest rates that is designed to endure for the decades to come.

Another problem that has endured in the literature – for more than a century – is that whenever the pioneering work of [Bachelier \(1900\)](#) is mentioned, the next statement is an indictment of the lack of suitability of arithmetic Brownian motion for modeling non-negative financial time series. Often, GBM is discussed next for its conformity to positive values. The trade-off, however, is that the GBM process specifies a vanishing rate of diffusion as the process nears zero, which was shown to fit the data poorly. In contrast, Bachelier's framework does specify a non-zero rate of diffusion near the bound, demonstrating performance comparable to the other models, and conforms to the characteristics of the historical series of interest rates. At long last, his pioneering model has been given a fair treatment and shown to be respectable with the proper guidance.

NOTE

1. One would normally expect that standard errors be shown with such estimates, however, the estimates from the auxiliary model are known to be biased, which is the primary motivation for using indirect inference, in this case, and for the applications presented in [Gouriéroux et al. \(1993\)](#).

ACKNOWLEDGMENTS

I thank Harry J. Paarsch for helpful comments on the manuscript. I am also grateful for helpful feedback on an earlier version of this manuscript from Morten Nielsen, James MacKinnon, Allan Gregory, and Prosper Dovonan, as well as from seminar participants at Queen's University and annual meetings of the Canadian Economics Association. I also gratefully acknowledge SSHRC for their generous support of this research during my doctoral studies at Queen's University. All errors are my own.

REFERENCES

- Àit Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, 70(1), 223–262.

- Äit Sahalia, Y., & Park, J. Y. (2012). Stationarity-based specification tests for diffusions when the process is nonstationary. *Journal of Econometrics*, 169, 279–292.
- Äit Sahalia, Y., & Park, J. Y. (2016). Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models. *Journal of Econometrics*, 192, 119–138.
- Anbil, S., Carlson, M. A., Hanes, C., & Wheelock, D. C. (2020). *A new daily federal funds rate series and history of the federal funds market, 1928–1954* [Federal Reserve Bank of St. Louis Working Paper Series 2020-016B, Federal Reserve Bank of St. Louis].
- Bachelier, L. (1900). Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 3(17), 21–86.
- Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654.
- Bruins, M., Duffy, J. A., Keane, M. P., & Smith, A. A., Jr. (2018). Generalized indirect inference for discrete choice models. *Journal of Econometrics*, 205(1), 177–203.
- Cai, Z., Li, Q., & Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2), 101–113.
- Campbell, J. Y. (1986). A defense of traditional hypotheses about the term structure of interest rates. *Journal of Finance*, 41(1), 183–193.
- Campbell, J. Y., & Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies*, 58(3), 495–514.
- Cavaliere, G. (2005). Limited time series with a unit root. *Econometric Theory*, 21(5), 907–945.
- Cavaliere, G., & Xu, F. (2014). Testing for unit roots in bounded time series. *Journal of Econometrics*, 178(2), 259–272.
- Choi, H.-s., Jeong, M., & Park, J. Y. (2014). An asymptotic analysis of likelihood-based diffusion model selection using high frequency data. *Journal of Econometrics*, 178, 539–557.
- Chung, H., & Park, J. Y. (2007). Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics*, 137(1), 230–259.
- Cochrane, J. H. (2005). *Asset pricing*. Princeton University Press.
- Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1981). A re-examination of traditional hypotheses about the term structure of interest rates. *Journal of Finance*, 36(4), 769–799.
- Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2), 385–407.
- Duffie, D. (2001). *Dynamic asset pricing theory* (Vol. 4, 3rd ed.). Princeton University Press.
- Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der Physik*, 17, 549–560.
- Einstein, A. (1926). *Investigations on the theory of Brownian movement*. Dover.
- Engle, R. F., III. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.
- Gallant, A. R., & Tauchen, G. (2006). Which moment to match? *Econometric Theory*, 12(4), 657–681.
- Gouriéroux, C., Monfort, A., & Renault, E. (1993). Indirect inference. *Journal of Applied Economics*, 8, S85–S118.
- Gouriéroux, C., Phillips, P. C. B., & Yu, J. (2010). Indirect inference for dynamic panel models. *Journal of Econometrics*, 157(1), 68–77.
- Guvenen, F., & Smith, A. A., Jr. (2014). Inferring labor income risk and partial insurance from economic choices. *Econometrica*, 82(6), 2085–2129.
- Hall, G., & Rust, J. (2021). Estimation of endogenously sampled time series: The case of commodity price speculation in the steel market. *Journal of Econometrics*, 222(1, Part A), 219–243.
- Han, H., & Park, J. Y. (2008). Time series properties of ARCH processes with persistent covariates. *Journal of Econometrics*, 146(2), 275–292.
- Harrison, J. M. (1985). *Brownian motion and stochastic flow systems*. John Wiley and Sons, Inc.
- Itô, K. (1944). Stochastic integral. *Proceedings of the Imperial Academy Tokyo*, 20, 519–524.
- Jeong, M. (2008). *Asymptotics for the Maximum Likelihood Estimation of Diffusion Models*. [PhD thesis]. Texas A & M University, College Station, TX.
- Jeong, M., & Park, J. Y. (2013). *Asymptotic theory of maximum likelihood estimation of diffusion models* [Technical report].

- Karatzas, I., & Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer-Verlag.
- Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press.
- Kim, J., & Park, J. Y. (2017). Asymptotics for recurrent diffusions with application to high frequency regression. *Journal of Econometrics*, 196(1), 37–54.
- Littermann, R., & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1), 54–61.
- Morin, L. (2017). *Keeping diffusion processes within bounds: Using information between observations* [PhD thesis]. Queen's University, Kingston, ON.
- Morin, L., & Shang, Y. (2020). Federal Reserve policy after the zero lower bound: An indirect inference approach. *Empirical Economics*, 60, 2105–2124.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Park, J. Y. (2002). Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics*, 110(2), 383–415.
- Park, J. Y., & Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory*, 15(5), 664–703.
- Phillips, P. C. B., & Yu, J. (2009). Maximum likelihood and Gaussian estimation of continuous time models in finance. In T. G. Andersen (Ed.), *Handbook of financial time series, Mathematics and statistics*. Springer London, Limited. ISBN: 9783540712978.
- Shreve, S. E. (2004). *Stochastic calculus for finance II: Continuous-time models*. Springer Finance. Springer Science + Business Media, LLC.
- Smith, A. A., Jr. (1990). *Three essays on the solution and estimation of dynamic macroeconomic models* [PhD thesis]. Duke University, Durham, NC.
- Smith, A. A., Jr. (1993). Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, 8, S63–S84.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177–188.
- Wiener, N. (1923). Differential space. *Journal of Mathematical Physics*, 2, 131–174.

APPENDIX. DATA PROCESSING

The data were joined from two main sources, both of which are available from the FRED database at the Federal Reserve Bank of St. Louis. For the period from July 1, 1954 to September 15, 2020, the series comprises daily observations of a single series, the *Effective Federal Funds Rate*, which is labeled *DFF* in the FRED database. This series covers the period that begins when the Federal Reserve Bank of New York began to record the daily figure. The data before 1954 were collected by [Anbil et al. \(2020\)](#), who transcribed four series from microfiche records of daily newspapers. The following list describes the series name in the FRED database, the series label, in upper case letters, along with the dates defining the period in which the series were observed. All series were downloaded on July 3, 2020.

- *High Value of the Federal Funds Rate for the Indicated Date Published in The New York Herald-Tribune (FFHTHIGH)* from 1928-04-04 to 1937-08-12.
- *High Value of the Federal Funds Rate for the Indicated Date Published in The Wall Street Journal (FFWSJHIGH)* from 1932-06-01 to 1954-06-28.
- *Low Value of the Federal Funds Rate for the Indicated Date Published in The New York Herald-Tribune (FFHTLOW)* from 1928-04-04 to 1938-03-01.
- *Low Value of the Federal Funds Rate for the Indicated Date Published in The Wall Street Journal (FFWSJLOW)* from 1932-06-01 to 1954-06-30.

These series were then aggregated into a single series using the following procedure, as outlined by [Anbil et al. \(2020\)](#). Each day, the aggregation method depends on the availability of the series because the dates of publication do not completely overlap and all four series were not reported every day.

1. For days when both newspapers reported a high and a low rate, the daily value was calculated using observations from both newspapers, using the simple average of the midpoints of the high and low rates reported in the *Herald Tribune* and *Wall Street Journal*. The high and low were not necessarily bid and offered rates; on dates when the *Wall Street Journal* provided an offer range, the midpoint of the bid-offer range was used to calculate the average value between the two newspapers.
2. For days when one newspaper reported only a single rate, that rate was taken as that newspaper's rate for the market. The average was then calculated with this value combined with the midpoint of the high and low (or bid and ask) provided by the other newspaper.
3. If only one newspaper provided data, the daily value was recorded as the midpoint of rates or the single rate provided by that newspaper.

To extend this series to the current decade, the digital series of the *Effective Federal Funds Rate*, from its first date of availability in the FRED database, was appended to the historical series from April 4, 1928 to June 30, 1954.

After constructing these series, [Anbil et al. \(2020\)](#) verified that connecting these series was reasonable, as the behavior near the end of the newspaper series was comparable to that of the beginning of the official series published by the Federal Reserve. For this paper, weekends and holidays were excluded from the series, to restrict analysis to business days and to avoid dates with partial coverage in the historical series from newspapers.

This page intentionally left blank

CHAPTER 7

A MARKET CRASH OR TAIL RISK? HEAVY TAILS AND ASYMMETRY OF RETURNS IN THE CHINESE STOCK MARKET

Zeyu Xing and Rustam Ibragimov

*Imperial College Business School, United Kingdom, and UBP Investment Management (Shanghai),
China*

ABSTRACT

Rapid stock market growth without real economic back-up has led to the 2015 Chinese Stock Market Crash with thousands of stocks hitting the down limit simultaneously multiple times. The authors provide a detailed analysis of structural breaks in heavy-tailedness and asymmetry properties of returns in Chinese A-share markets due to the crash using recently proposed robust approaches to tail index inference. The empirical analysis points out to heavy-tailedness properties often implying possibly infinite second moments and also focuses on gain/loss asymmetry in the tails of daily returns on individual stocks. The authors further present an analysis of the main determinants of heavy-tailedness in Chinese financial markets. It points out to liquidity and company size as being the most important factors affecting the returns' heavy-tailedness properties. At the same time, the authors do not observe statistically significant differences in tail indices of the returns on A-shares and the coefficients on factors affecting them in the pre-crisis and post-crisis periods.

Keywords: Market crashes; crises; tail risk; heavy tails; gain/loss asymmetry; structural breaks; Chinese stock market

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 181–205

Copyright © 2023 by Zeyu Xing and Rustam Ibragimov

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045B009

1. INTRODUCTION

1.1. Heavy-Tailedness in Financial Markets

Numerous studies in finance and economics indicate that many key variables in these fields, including financial returns and foreign exchange rates, are heavy-tailed and have thick-tailed power law distributions (see, among others, Cont, 2001; Embrechts et al., 1997; Gabaix, 2009, 2016; Ibragimov et al., 2015; Ibragimov & Prokhorov, 2017; Loretan & Phillips, 1994; Ch. 3 in McNeil et al., 2015; Park, 2002, and references therein). For a heavy-tailed power law distributed random variable (r.v.) X (e.g., representing a risk, financial return or exchange rate), one has

$$P(X > x) \sim \frac{c_1}{x^{\xi^+}} \quad (1)$$

$$P(X < -x) \sim \frac{c_2}{x^{\xi^-}} \quad (2)$$

$$P(|X| > x) \sim \frac{c}{x^\xi}, \quad (3)$$

for large x' 's, where $\xi^-, \xi^+, \xi = \min(\xi^-, \xi^+)$, c, c_1, c_2 are positive constants. The parameters ξ^+, ξ^-, ξ in (1)–(3) are referred to as, respectively, the right tail index, left tail index and tail index (or tail exponent) of the variable X . The parameters ξ^+, ξ^- and ξ characterize the degree of heavy-tailedness (the rate of decay of the tails) of power law distributions. Heavy-tailed distributions provide a convenient framework for modeling and quantifying (by their key parameters – the tail index values ξ_+ , ξ_- and ξ) the likelihood and the magnitude of large downfalls, large fluctuations and crises in financial and economic markets. In particular, in (1)–(3), the *smaller* values of the tail index parameters ξ_+ , ξ_- and ξ correspond to a *higher* degree of heavy-tailedness and thus to a larger likelihood of crises, large downfalls and large fluctuations in (e.g., financial returns) time series of observations on the variable X , and vice versa (see the discussion in Ch. 1 in Ibragimov et al., 2015; Ibragimov & Prokhorov, 2017). The tail index parameters are further important as they govern existence of moments of X with, for instance, the variance of X being defined and finite: $Var(X) < \infty$ if and only if $\xi > 2$, and, more generally, the p th moment $E|X|^p$, $p > 0$, being finite: $E|X|^p < \infty$ if and only if $\xi > p$.¹ The most of the empirical literature on heavy-tailed distributions agrees that, in the case of developed financial markets, the returns and foreign exchange rates' tail indices ξ belong to the interval $(2, 4)$, thus implying finite variances and infinite fourth moments (Ibragimov & Ibragimov & Walden, 2015; Ibragimov & Prokhorov, 2017).² Importantly, the stylized facts of heavy-tailedness and volatility clustering are exhibited by widely used GARCH-type models for financial returns and foreign exchange rates (see the results and discussion in Section 8.4 in Davis & Mikosch, 1998; Embrechts et al., 1997; Han & Park, 2008; Ibragimov et al., 2020; Ch. 4 in McNeil et al., 2015; Mikosch & Starica, 2000b; Park, 2002). One should note that finiteness of variances and higher moments for economic

and financial indicators like financial returns and exchange rates is crucial in the analysis of many models in economics and finance and also for applicability of classical statistical and econometric approaches, including regression and least squares methods. In a similar fashion, the problem of heavy-tailedness with potentially infinite fourth moments, nonlinear dependence and potential nonstationarity of economic and financial time series needs to be taken into account in applications of regression and autocorrelation-based methods, and related inference procedures in their analysis (see, among others, the discussion in [Granger & Orr, 1972](#), and in a number of more recent studies, e.g., [Anatolyev, 2019](#); [Chung & Park, 2007](#); Section 7 in [Cont, 2001](#); [Davis & Mikosch, 1998](#); [Embrechts et al., 1997](#); Ch. 1 in [Ibragimov et al., 2015](#); [Ibragimov et al., 2020](#); [Loretan & Phillips, 1994](#); [Mikosch & Starica, 2000b](#); [Miller & Park, 2010](#), and the references therein).

[Gabaix et al. \(2003, 2006\)](#) review the empirical results that imply that, in developed financial markets, the tail index of stock returns is close to 3 and develop a theoretical model explaining this empirical regularity that the authors call the “Cubic Law of the Stock Returns.” In the model, heavy-tailedness of financial returns is implied by trading actions of largest market participants (mutual funds and other institutional investors) that have a size distribution with the tail index $\xi = 1$ (Zipf’s law). The empirical results in the literature imply that the “Cubic Law of the Stock Returns” does not hold in the case of emerging and developing country financial returns and foreign exchange rates as many of them, including those in Chinese markets, have tail indices smaller than 3, and even tail indices smaller than 2 and infinite variances are not uncommon (see [Chen & Ibragimov, 2019](#); [Gu & Ibragimov, 2018](#); [Ibragimov et al., 2013](#); Ch. 3 in [Ibragimov et al., 2015](#) and references therein). This implies that the model in [Gabaix et al. \(2003, 2006\)](#) may need to be modified in the case of emerging and developing markets, e.g., with possible deviations of the distribution of sizes of market participants from Zipf’s law due to the governments’ regulatory interventions. The analysis in [Quintos et al. \(2001\)](#) and [Candelon and Straetmans \(2006\)](#) point to important structural breaks in heavy-tailedness properties of developed and emerging country stock index returns and foreign exchange rates, in particular, those due to the Asian financial crisis.³ [Ankudinov et al. \(2017\)](#) provides the analysis of tail index regressions for financial returns in Russia that indicates importance of stock liquidity and company size as determinants of the returns’ heavy-tailedness.

1.2. Chinese Equity Market and the 2015 Crash

In terms of scale, China’s economy is currently the second largest one in the world and is on the track of exceeding that of the United States. However, institutions and financial management in Chinese stock markets work both similar and different from the western world ([Jiang et al., 2017](#)).

The Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) date back to the 1990s. Both the exchanges are order-driven without specialists or market makers with both upward cap and downward floor of 10% daily price fluctuation limit from the closing price of the prior day. In March 2010, China

introduced a pilot scheme and eased the ban on short-selling and margin-trading for a certain list of stocks. [Chang et al. \(2014\)](#) find that this scheme has improved Chinese stock market efficiency and decreased its volatility.

A-shares are regular domestic stocks settled in renminbi (RMB). There are also stocks settled in foreign currencies other than RMB. B-shares are denominated in US or Hong Kong dollars with the same cash flow rights as A-shares. Cross-listed shares are stocks listed both in Chinese mainland and Hong Kong or foreign markets, such as H-shares (listed in Hong Kong), N-shares (listed in New York), S-shares (listed in Singapore) and L-shares (listed in London). The Shanghai-Hong Kong Stock Connect and Shenzhen-Hong Kong Stock Connect were introduced in 2014 and 2016 respectively, since when qualified investors are able to directly invest in Hong Kong stock markets (see [Jiang et al., 2017](#) and also the review in [Chen & Ibragimov, 2019](#)).

Before having peaked on June 12, 2015, stock market in China had increased around 150% in one single year, while the nominal GDP growth rate had been on the downwards track from 14.23% in 2007 to 6.9% in 2015. However, it took far less time for the market to drop, when it went down 40% from around 5,166 to 3,052 points in less than four months.

[Zeng et al. \(2016\)](#) summarize the causes of China's 2015 stock market crash as the inconsistency of economic fundamentals and market performance. Besides the slowing GDP growth rate, stock prices skyrocketed in companies with meager earnings or even losses. On the demand side, a huge amount of inexperienced investors flooded into the market with their own savings and the number of investors added up to 7% of the population then ([Sornette et al., 2015](#)). On the supply side, short selling and trading on the margin have been approved during that period by authorities, contributing to the market leverage.

More than 40 measures had been enacted to rescue the market by the government and authorities. However, by the end of August, as the SSE Composite Index fell down below 4,500 points, it became clear that the government has failed to boost the stock market as was initially planned. Among so many actions deployed, only a few have worked except one, which was the establishment of the so-called "National Team" that directly purchased shares from the market (see [Zeng et al., 2016](#)).

1.3. Research Problems and Contributions

Despite the existence of a few studies on heavy-tailedness properties of developed markets reviewed in Section 1.1, the research on heavy-tailedness of emerging markets, in particular, those in China, and especially their determinants remains limited.

This chapter attempts to partially fill this gap in the literature by providing a detailed econometric analysis of several important research questions on heavy-tailedness characteristics of financial markets in China.

- i. Using recently proposed robust approaches to tail index inference, we provide a rigorous analysis of heavy-tailedness and asymmetry properties of returns in Chinese A-share markets.

- ii. We further provide a detailed analysis of structural changes in these properties due to the 2015 crash.
- iii. The chapter also presents a study of the main determinants of heavy-tailedness in Chinese financial markets, including stock market factors, stock liquidity and firm-specific variables and attributes such as company size, ownership structure and sector affiliation.
- iv. In addition, we provide the results on the analysis of potential structural breaks in the relationship between the degree of heavy-tailedness in A-share returns and the explanatory factors considered.

Among other results, the chapter provides the estimates of the degree of heavy-tailedness of daily returns on individual stocks that point to their fat-tailedness properties with possibly infinite variances and second moments. The estimates further indicate gain/loss asymmetry in the tails of the returns' distributions.

The results the chapter point out to liquidity and company size as being the most important factors affecting A-share returns' heavy-tailedness properties. At the same time, we do not observe statistically significant differences in tail indices of the returns on A-shares and the coefficients on factors affecting them in the pre-crisis and post-crisis periods.

1.4. Organization of the Chapter

The chapter is organized as follows. Section 2 reviews tail index inference approaches employed in the analysis. Section 3 describes the data and variables used in the analysis. Section 4 provides the results of the empirical analysis and their discussion. In Section 6, we make some concluding remarks.

2. INFERENCE ON HEAVY-TAILEDNESS

Several approaches to inference about the tail index of heavy-tailed distributions (1)–(3) are available in the literature (see, among others, the reviews in [Beirlant et al., 2004](#); [Embrechts et al., 1997](#); Ch. 3 in [Ibragimov et al., 2015](#)). The two most commonly used are Hill's tail index estimates and the OLS log–log rank-size regression approach to tail index estimation.

Below, we describe Hill's and log–log rank-size regression estimates $\hat{\xi}_{Hill}$ and $\hat{\xi}_{RS}$ of the tail index ξ in power law distributions (3). Hill's and log–log rank-size regression estimates $\hat{\xi}_{Hill}^+$, $\hat{\xi}_{RS}^+$ and $\hat{\xi}_{Hill}^-$, $\hat{\xi}_{RS}^-$ of the right and left tail indices ξ^+ and ξ^- (1)–(2) are defined in a similar way, with the use of the largest positive returns and largest negative returns instead of the returns' largest absolute values.

Let r_1, r_2, \dots, r_N be a sample of returns that have power law distribution (3). Furthermore, let, for $n < N$;

$$|r|_{(1)} \geq |r|_{(2)} \geq \dots \geq |r|_{(n)} \quad (4)$$

be decreasingly ordered largest absolute values of observations in the sample (in practical applications, one usually takes the number n of observations in the tails of power law distribution used for estimation of the tail index ξ to be equal to some small fraction, e.g., 5% or 10%, of the total sample size N : $n \approx 0.05N, 0.1N$). Hill's estimator $\hat{\xi}_{\text{Hill}}$ of the tail index ξ in (3) is given by

$$\hat{\xi}_{\text{Hill}} = \frac{n}{\sum_{t=1}^n (\log |r|_{(t)} - \log |r|_{(n+1)})} \quad (5)$$

with the standard error $S.E. = \frac{1}{\sqrt{n}} \hat{\xi}_{\text{Hill}}$. The corresponding 95% confidence interval for the true value of the tail index ξ is given by

$$(\hat{\xi}_{\text{Hill}} - 1.96 \times \frac{1}{\sqrt{n}} \hat{\xi}_{\text{Hill}}, \hat{\xi}_{\text{Hill}} + 1.96 \times \frac{1}{\sqrt{n}} \hat{\xi}_{\text{Hill}}). \quad (6)$$

It was reported in a number of studies that inference on the tail index using Hill's estimator suffers from several problems, including sensitivity to dependence, deviation from power laws in the tails and small sample sizes (see, among others, the discussion in Ch. 6 in [Embrechts et al., 1997](#); [Gabaix & Ibragimov, 2011](#); Ch. 3 in [Ibragimov et al., 2015](#)). Motivated by these problems, several studies have focused on alternative robust approaches to the tail index estimation, including small-sample weighted analogues of Hill's estimator (see [Huisman et al., 2001](#)) and nonlinear tail index estimation approaches (see [Embrechts et al., 1997](#)). A popular simple approach to tail index estimation that appears to be more robust than inference procedures based on Hill's estimates is that based on log–log rank-size regressions.⁵ The log–log rank-size regression tail index estimation approach is motivated by the linear relationships like $\log(P(|X| > x)) \sim c - \xi \log(x)$ for large x 's implied by (3). [Gabaix and Ibragimov \(2011\)](#) show that the empirical analogues of such relationships that correct the bias in tail index estimation to the first-order can be based on regressions of shifted ranks of power law distributed observations on their sizes, that is, the regressions $\log(\text{Rank} - 1/2) = a - b \log(\text{Size})$ with the optimal shift of 1/2 in ranks. More precisely, following the approach, one runs the following OLS regression:

$$\log(t - 1/2) = a - b \log |r|_{(t)}, \quad t = 1, 2, \dots, n, \quad (7)$$

and takes b as the log–log rank-size regression estimate $\hat{\xi}_{\text{RS}}$ of the tail index ξ in (3). [Gabaix and Ibragimov \(2011\)](#) show that the standard error of the log–log rank-size regression tail index estimator $\hat{\xi}_{\text{RS}}$ is different from the OLS standard error and is given by $S.E. = \sqrt{\frac{2}{n}} \hat{\xi}_{\text{RS}}$. The corresponding 95% confidence interval for the true value of the tail index ξ is given by

$$\left(\hat{\xi}_{\text{RS}} - 1.96 \times \sqrt{\frac{2}{n}} \hat{\xi}_{\text{RS}}, \hat{\xi}_{\text{RS}} + 1.96 \times \sqrt{\frac{2}{n}} \hat{\xi}_{\text{RS}} \right). \quad (8)$$

The analysis of (a)symmetry in heavy-tailedness properties of upward and downward fluctuations, that is large upward moves and large downfalls, in financial returns r_t can be based on estimates of the right and left tail indices ξ^+ , ξ^- in power laws (1)–(2) and their standard errors and confidence intervals constructed as described in this section (see also Ch. 3 in Ibragimov et al., 2015; Section 3.2 in Ibragimov et al., 2015, for examples of the analysis of asymmetry in right and left tails of the distribution of emerging country exchange rates using their right and left tail indices).

In a similar way, the analysis of structural breaks in heavy-tailedness properties of financial returns due to the 2015 crash can be conducted using the estimates $\hat{\xi}_{\text{pre}}$ and $\hat{\xi}_{\text{post}}$ of the tail indices in the pre-crash and post-crash periods, the standard errors of the estimates described in this section and the corresponding confidence intervals.⁶

3. DATA

The analysis is based on the data for the period from 13/04/2012 to 04/07/2019 collected from Bloomberg. The data include the time series of daily closing prices, daily volume, daily bid-ask spread, quarterly market capital, daily P/B ratio, yearly return on equity (ROE) and total investment to total assets ratio.

The analysis of structural breaks due to the 2015 crash in the degree of heavy-tailedness of financial returns and in its dependence on explanatory factors considered is conducted using the data for the pre-crash period from 13/04/2012 to 12/06/2015 and the post-crash break period from 13/06/2015 to 04/07/2019.

Many companies in the sample have a record of hundreds of days on suspension. To deal with such missing data problems, the analysis focuses on shares that have at least 1,600 of non-zero returns in the 1,885 trading days considered. The sample considered consists of 1,038 public companies.

4. EMPIRICAL ANALYSIS

Due to the large volume of stock indices, we present and describe in detail the results of tail index estimation for four companies in the sample. The results of the analysis for the full sample of companies are presented in the online Appendix.

4.1. Tail Index Estimates

Table 1 provides Hill's and the log–log rank-size regression estimates $\hat{\xi}_{\text{Hill}}$ and $\hat{\xi}_{\text{RS}}$ of the tail index ξ of the returns on the shares of companies considered, together with their standard errors and the corresponding 95% confidence intervals (see Section 2 for details). Due to the moderate size of financial returns time series used in the analysis, tail index estimation uses the truncation levels n equal to 7.5% and 15% of the total sample size N of time series observations: $n \approx 0.075N, 0.15N$. The analysis below using analogues of Hill's plots for log-log rank-size regressions shows that the tail index estimates are generally robust to the truncation level choice.

Table 1. Tail Index Estimates for Absolute Returns.

Stock	Truncation (%)	N	Log–Log Rank-Size		Hill’s Estimate	
			$\hat{\xi}_{RS}$	$CI_{95\%,RS}$	$\hat{\xi}_{Hill}$	$CI_{95\%,Hill}$
After the 2015 China’s Stock Market Crash						
000963 CH Equity	7.5	951	3.54	(2.38, 4.7)	3.17	(2.44, 3.9)
	15.0	951	3.28	(2.52, 4.04)	2.92	(2.44, 3.4)
600021 CH Equity	7.5	887	2.84	(1.89, 3.79)	2.19	(1.67, 2.72)
	15.0	887	2.08	(1.58, 2.58)	1.67	(1.39, 1.95)
600770 CH Equity	7.5	884	4.98	(3.29, 6.67)	3.21	(2.44, 3.98)
	15.0	884	2.55	(1.94, 3.16)	2.02	(1.68, 2.36)
601010 CH Equity	7.5	910	2.97	(1.98, 3.96)	2.20	(1.68, 2.73)
	15.0	910	2.08	(1.59, 2.57)	1.67	(1.39, 1.95)
Before the 2015 China’s Stock Market Crash						
000963 CH Equity	7.5	743	4.36	(2.76, 5.96)	3.90	(2.89, 4.91)
	15.0	743	3.67	(2.71, 4.63)	2.90	(2.36, 3.44)
600021 CH Equity	7.5	717	2.61	(1.63, 3.59)	1.98	(1.46, 2.51)
	15.0	717	2.09	(1.54, 2.64)	1.88	(1.53, 2.23)
600770 CH Equity	7.5	746	3.87	(2.45, 5.29)	3.11	(2.3, 3.91)
	15.0	746	3.24	(2.4, 4.08)	2.92	(2.38, 3.46)
601010 CH Equity	7.5	745	2.74	(1.73, 3.75)	2.25	(1.67, 2.84)
	15.0	745	2.11	(1.56, 2.66)	1.87	(1.53, 2.22)
Through the Whole Period of 2012 to 2019						
000963 CH Equity	7.5	1694	3.89	(2.94, 4.84)	3.45	(2.85, 4.05)
	15.0	1694	3.47	(2.87, 4.07)	2.92	(2.56, 3.28)
600021 CH Equity	7.5	1604	2.81	(2.1, 3.52)	2.11	(1.73, 2.48)
	15.0	1604	2.09	(1.72, 2.46)	1.75	(1.53, 1.97)
600770 CH Equity	7.5	1630	3.91	(2.93, 4.89)	2.80	(2.31, 3.3)
	15.0	1630	2.74	(2.25, 3.23)	2.27	(1.99, 2.55)
601010 CH Equity	7.5	1655	2.90	(2.18, 3.62)	2.22	(1.83, 2.61)
	15.0	1655	2.10	(1.73, 2.47)	1.76	(1.54, 1.98)

The results in [Table 1](#) suggest that the obtained Hill’s estimates of tail indices of absolute returns are in general smaller than their log–log rank-size regression estimates. This is further confirmed by [Fig. 1](#) that provides the histograms of Hill’s and log–log rank-size regression tail index estimates of returns with 15% truncation across the companies considered. In both the pre-crisis and post-crisis periods and over the whole period considered, Hill’s estimates are distributed leftward on the x axis as compared to the log–log rank-size regression estimates. Both of the histograms have a long tail on the right hand side indicating less pronounced heavy-tailedness for many companies. The estimates are distributed nearly symmetrically around the center. The average differences of Hill’s estimate and the log–log rank-size estimate are 0.4787, 0.4617 and 0.4431 post-crisis, pre-crisis and in the whole period, respectively.

Importantly, with the 15% truncation level, there are only seven companies for which the 95% confidence intervals for the tail indices of their shares constructed using the log–log rank-size regression and Hill’s estimates do not intersect. For these companies, this is the case for confidence intervals for the tail index of absolute returns over the whole period. However, for the 7.5% truncation level, there are 148 out of 9,342 cases where the above confidence intervals do not intersect

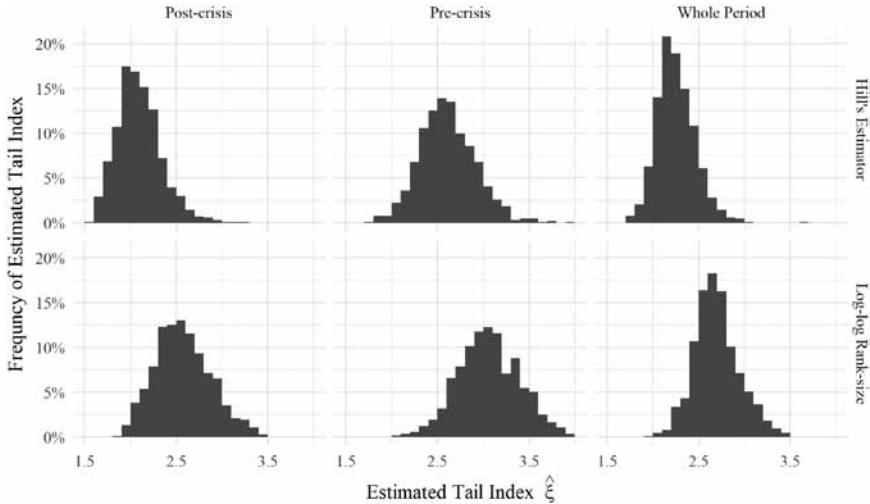


Fig. 1. Histogram of Tail Index Estimates with 15% Truncation.

with each other. The difference may be due to strong sensitivity of Hill's estimates to truncation levels used (Embrechts et al., 1997). Overall, the above results point out to similarity of Hill's and log–log rank-size estimates of the tail indices of financial returns in the Chinese stock market in the period from April 2012 to July 2019, even though Hill's point estimates are in general smaller.

Due to the disadvantages of Hill's estimators discussed in Section 2 and robustness of log–log rank-size regression approach, the analysis and the discussion below will be mostly based on log–log rank-size regression approach to tail index estimation with the optimal shift of 1/2 in ranks (see the discussion in Gabaix & Ibragimov, 2011 and Section 2).

Similar to Ibragimov et al. (2013), in order to illustrate the appropriateness of the tail truncation levels (7.5% and 15%) used in this section, we follow the analysis and suggestions in Embrechts et al. (1997) and Mikosch and Starica (2000b) and present, in Figure 2, the analogues of Hill's plots for the log–log rank-size regression tail index estimates for four companies in the sample. These are graphs of the log–log rank-size regression point estimates $\hat{\xi}_{RS}$ of the tail indices of the returns on the companies' shares, together with the corresponding 95%-confidence intervals in (8) for the true tail index values computed using log–log rank-size regressions,

Similar to the four companies dealt with in the figure, the log–log rank-size estimates for returns on shares of companies in the sample, the log–log rank-size tail index estimates start to stabilize above the truncation level roughly 6% while the length of the 95% confidence intervals shrinks down evidently as well. The conclusions indicate that the use of the truncation levels of 7.5% and 15% in tail index estimation is reasonable. In general, the log–log rank-size regression tail index estimates are robust to the choice the truncation level choice greater than 6%, as most of the corresponding confidence intervals for the tail index

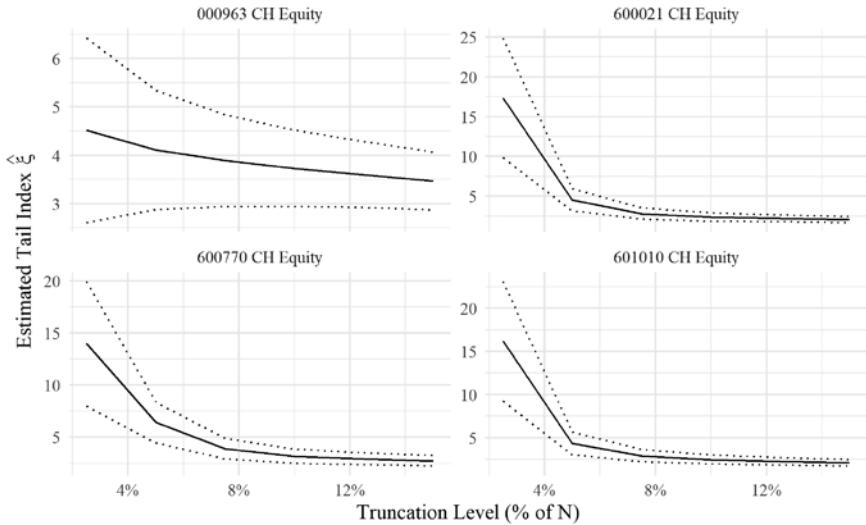


Fig. 2. The Dynamics of Daily Returns.

intersect. At the same time, in contrast to the analysis for emerging country foreign exchange rates in Ibragimov et al. (2013), due to smaller sample sizes in this paper, the confidence intervals constructed using small truncation levels are rather wide and do not intersect with those constructed using the above larger truncation levels.

The plots in Fig. 3 illustrate the dynamics of daily returns from four sampled public companies with different tail indices and the degrees of heavy-tailedness. The blank parts of the diagrams for the returns on 600021 CH and 600770 CH equities correspond to their long periods of trading suspensions. It was more than normal for companies in China to suspend their trading due to high market volatility or other major issues during and after the 2015 stock market crash.

The histograms in Fig. 4 further illustrates the differences in heavy-tailedness of the distribution of the returns plotted in Fig. 3. As is seen from Figs. 3–4, heavy-tailedness of the returns' distribution and presence of outliers becomes more pronounced as the tail indices decrease. In addition, as is seen from the figures the decrease in tail indices implying more pronounced heavy-tailedness is accompanied by the increase in the likelihood and the magnitude of large downfalls and large fluctuations in the returns' time series.

In the plots in Fig. 3, one observes clusters of returns in the intervals $[-10\%, -9\%], (9\%, 10\%]$. This is due to the cap (floor) of 10% on the daily price variation in the Chinese market. The clusters at the boundary of 10% indicate a so-called magnet effect of daily price limits. Liquidity risk and behavioral investors catalyze the process to touch and trigger the limit (Cho et al., 2003). Due to the presence of clusters at the boundary of 10% (Hill's and log-log rank-size regression) tail index estimates with small truncation levels used in estimation are expected to be large (Chen & Ibragimov, 2019), which is why, in particular, the tail

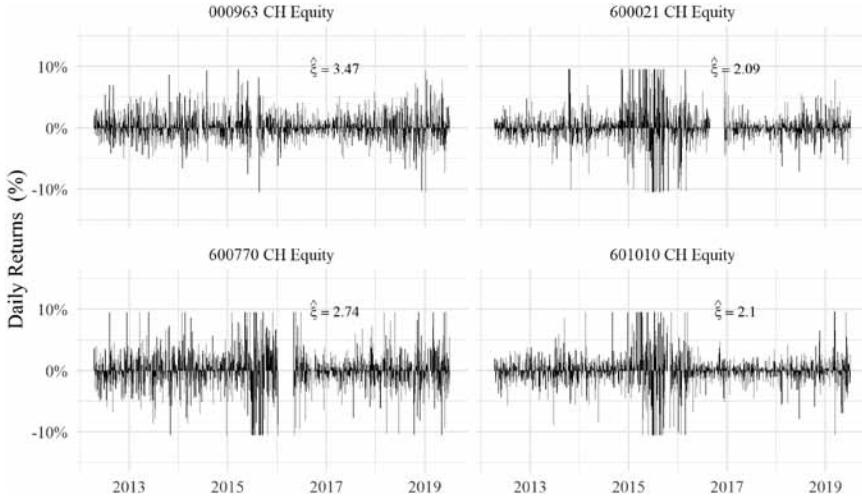
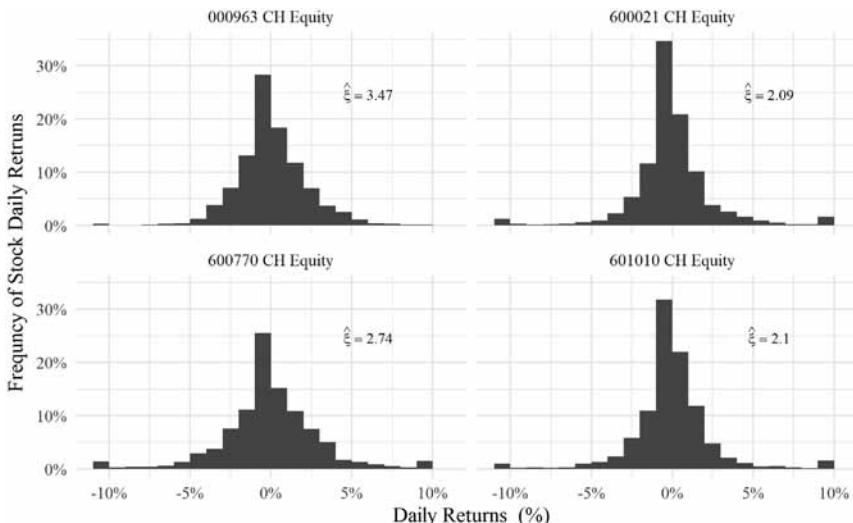


Fig. 3. Histograms of Daily Returns.

Fig. 4. Hill's Plots for Log–Log Rank-Size Regression Tail Index Estimates $\hat{\xi}_{RS}$.

index estimates at the starting points with small truncation levels in the graphs in the Fig. 2 are so high as compared to the estimates with larger truncation levels.

4.2. (A)symmetry in the Left and Right Heavy-Tailedness

This section provides the analysis of asymmetry in heavy-tailedness in the right and left tails of the distribution of the returns considered. It also analyses structural changes in the asymmetry due to the 2015 market crash.

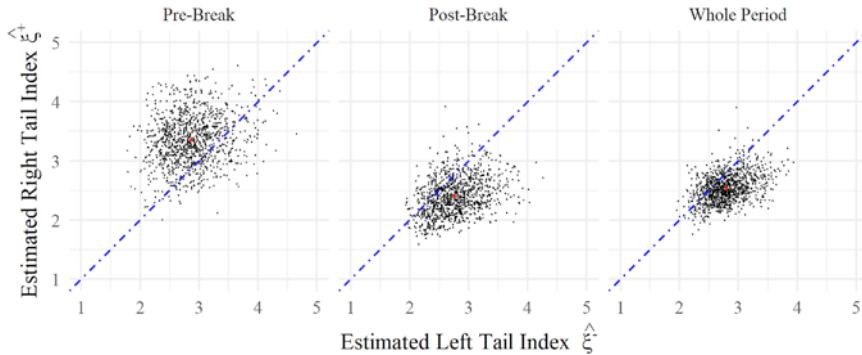


Fig. 5. Scatterplots of the Left and Right Tail Index Estimates ($\hat{\xi}^-$, $\hat{\xi}^+$).

[Fig. 5](#) provides scatterplots of the log–log rank-size regression estimates $\hat{\xi}_{\text{RS}}^+$ and $\hat{\xi}_{\text{RS}}^-$ of the right and left tail indices ξ^+ and ξ^- in power laws (1) and (2) for financial returns considered. One can observe, due to smaller sample sizes used in estimation, a larger variance in estimates of right tail and left tail indices ξ^+ and ξ^- in (1) and (2) in the pre- and post-break periods as compared to the estimates over the whole period.

In [Fig. 5](#), if a point is located downwards to the diagonal line, then this suggests that the right tail index ξ^+ is smaller than the left tail index ξ^- : $\xi^+ < \xi^-$. That is, in this case, heavy-tailedness is more pronounced in the right tail of the distribution of the return on the corresponding stock. The conclusions are reversed for a point located above the diagonal line.

According to tail index estimation results, before the market crash on June 12th, 2015, 832 of 1,038 companies had a larger estimate of the right tail indices ξ^+ as compared to the estimates of the left tail indices ξ^- . This number declines to 191 post-crisis. This effect appears to be more pronounced for large downfalls in the whole period, with the points $(\hat{\xi}_{\text{RS}}, \hat{\xi}_{\text{RS}}^+)$ for 815 of 1,038 stocks lying below the diagonal line (although very close to it).

One should note that the 95% confidence intervals for the right and left tail indices ξ and ξ for all the stocks considered that are constructed using their estimates in [Fig. 5](#) intersect. This implies that the null hypothesis $H_0: \xi^- = \xi^+$ of symmetry in the degree of heavy-tailedness properties of the right and left tails of the returns' distributions cannot be rejected in favor of the two-sided alternative $H_a: \xi^- \neq \xi^+$ at 5% level. That is, the right and left tail indices ξ^+ and ξ^- of the returns' distributions are statistically indistinguishable (see also [Ibragimov et al., 2013](#), for the analysis and conclusions on symmetry in the degree of heavy-tailedness in right and left tails of distributions of emerging country foreign exchange rates).

4.3. Structural Breaks in Heavy-Tailedness

Similar to the analysis of heavy-tailedness (a)symmetry in the previous section, this section of the paper provides the analysis of structural breaks in heavy-tailedness

properties of returns' distribution due to the 2015 crash using the estimates and confidence intervals for tail indices in the pre-crash and post-crash periods (see Ibragimov et al., 2013, for the related analysis of structural breaks in heavy-tailedness properties of distributions of emerging country foreign exchange rates due to the beginning of the 2008 financial crisis). Fig. 6 provides the scatterplots of the pre-crash and post-crash estimates $(\hat{\xi}_{\text{pre}}^-, \hat{\xi}_{\text{post}}^-)$, $(\hat{\xi}_{\text{pre}}^+, \hat{\xi}_{\text{post}}^+)$ and $(\hat{\xi}_{\text{pre}}, \hat{\xi}_{\text{post}})$ of tail indices of the returns in (1)–(3).

In the scatterplots in Fig. 6, if a point is located below the diagonal line, then this suggests a larger value of the corresponding tail index in the pre-crash period as compared to its post-crash value. In this case, therefore the corresponding return has a thinner tail before June 12th, 2015 as compared to the period after that date. In contrast, if a point in the scatterplots is located above the diagonal line, then this suggest that the corresponding pre-crash tail index value is smaller than its post-crash value, and thus the corresponding public company's return has a heavier tail in the period before June 12th, 2015 as compared to the period after that date.

As is seen from the scatterplots for the pre-crash and post-crash left tail indices of large stock price downfalls in Fig. 6 there appears to be a significant change in the tail indices at the time of the crash. The pre-crash left tail index value is smaller than its post-crash value only in the case of 25 of 1,038 return time series. This indicates that, for nearly all of the returns in the sample, the left tails become heavier after the stock market crash. This conclusion may be due to the bear market with large stock price downfalls replacing the bull market before the crash.

In the corresponding scatterplot for the right tail indices of large stock price upward movements in Fig. 6, one observes less deviations of the points $(\hat{\xi}_{\text{pre}}^+ > \hat{\xi}_{\text{post}}^+)$ from the diagonal line. Specifically, the points lie above the diagonal for 603 out of 1,038 companies: $(\hat{\xi}_{\text{pre}}^+ < \hat{\xi}_{\text{post}}^+)$, while the points for the other 425 companies lie below the diagonal with $(\hat{\xi}_{\text{pre}}^+ > \hat{\xi}_{\text{post}}^+)$. Although more positive large

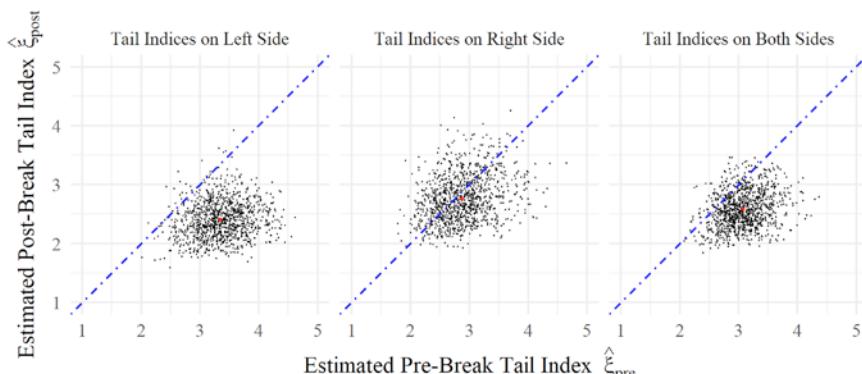


Fig. 6. Scatterplots of the Pre-Crash & Post-Crash Tail Index Estimates $(\hat{\xi}_{\text{pre}}, \hat{\xi}_{\text{post}})$.

upward movements are expected during the bull market suggesting that pre-crisis right tail indices are expected to be smaller than their post-crisis values (with more pronounced heavy-tailedness in the right tails of returns' distributions in the pre-crisis period), the relative symmetry of the scatterplot of the pre- and post-crash estimates of the right tail indices may be explained by a positive and active market environment in the post-crisis period. In particular, the post-crash period includes the first half of 2019. During the first half of 2019, the SSE Composite opened on 2nd January at 2497.88 points and closed on 4th July at 3005.25 points, while SSEC hit 3288.45 as its highest price on 8th April with the maximum price growth of 31.73% during the period. Such a positive and active market environment brings positive returns, which could neutralize the effect of the bear market. As this period is a part of the post-crisis period, the estimates $\hat{\xi}_{\text{pre}}^+$ and $\hat{\xi}_{\text{post}}^+$ are closer to each other.

As shown in Fig. 6, due to larger sample sizes used in estimation, the tail index estimates for absolute returns (3) have smaller variance than the estimates of right and left tail indices in (1)–(2). The points $(\hat{\xi}_{\text{pre}}, \hat{\xi}_{\text{post}})$ with pre- and post-crash estimates of the tail index $\hat{\xi}$ are located above the diagonal for returns of only 105 of 1,038 companies. The conclusion that, in general, the returns' tail indices become smaller post-crash thus implying more pronounced heavy-tailedness is in accordance with the property that more large stock price downfalls is observed during the bear market.

Similar to the heavy-tailedness (a)symmetry analysis in Section 4.2, for most of the returns in the sample, the 95% confidence intervals with 15% truncation levels for the pre-break and post-break tail indices intersect. There are only 3 and 13 cases for negative and absolute returns, respectively, where the intersections of the confidence intervals are empty. Even with the lower truncation level of 7.5%, the total number of stocks with empty intersections of the confidence intervals adds up to 96 out of 3,114 cases (1,038 companies \times 3 types of tail indices).

Fig. 7 illustrates the above three cases where the confidence intervals for the left tail indices in the pre-crash and post-crash periods do not intersect and thus a structural break in the left tail indices at the 2015 crash date is observed. The change in the distribution of negative returns of these stocks is visible from the returns' histograms. Taking 002138 CH Equity as an example, the pre-break left tail index estimate is $\hat{\xi}_{\text{pre,RS}}^- = 4.55$ with the 95% confidence interval $CI_{95\%,\text{RS}}^- = (2.83, 6.27)$. In the post-crisis period, the left tail index estimate becomes $\hat{\xi}_{\text{post,RS}}^- = 2.11$ with the 95% confidence intervals $CI_{95\%,\text{RS}}^- = (1.43, 2.79)$, implying that the tails of the negative returns' distribution become heavier post-crisis. This, in turn, explains the fact that more frequent and larger in magnitude downfalls are observed in the returns' time series considered in the post-crisis period.

Similar conclusions are also implied by the corresponding diagrams for the 13 returns' time series with statistically significant changes in the tail indices of their absolute values. Table 2 provides the estimates and the corresponding 95%

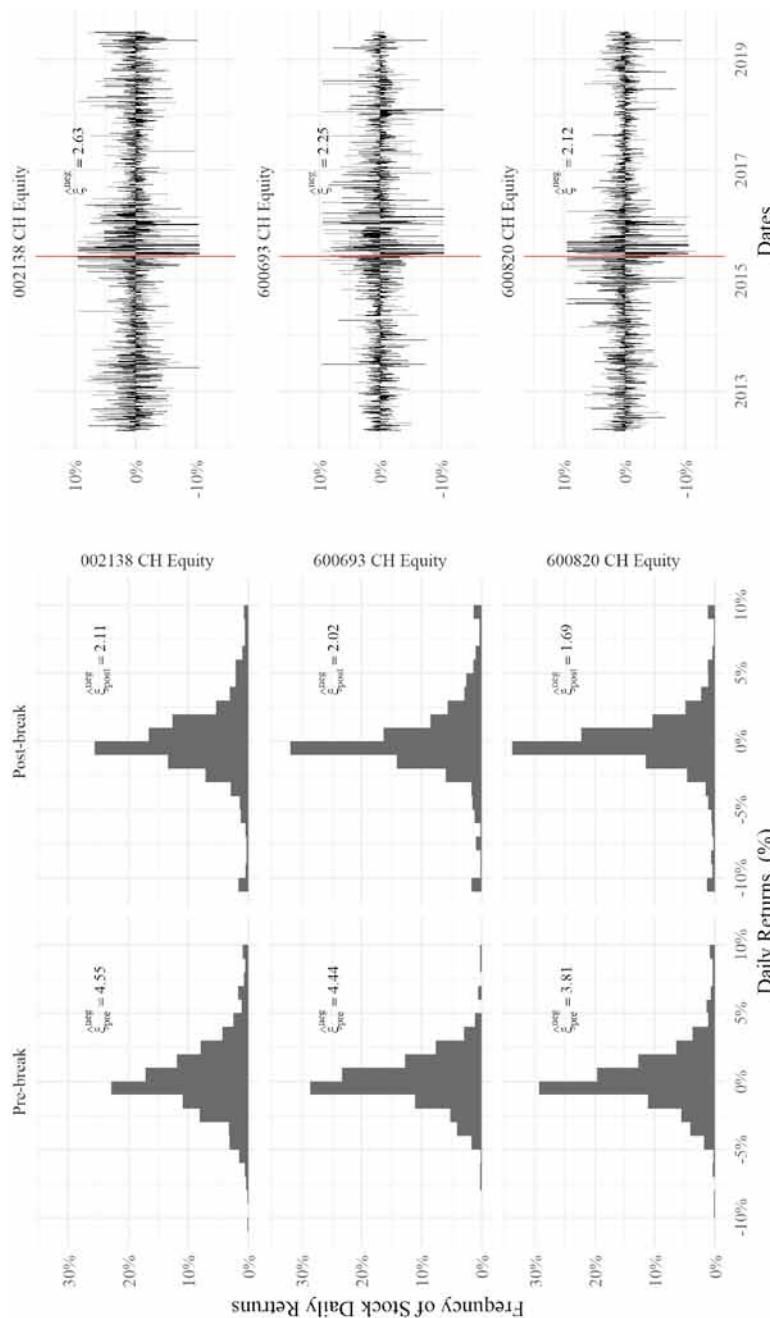


Fig. 7. Daily Returns Variation Between Pre-/Post-Crisis with $\hat{\xi}_{rs,15\%}$.

confidence intervals for the tail indices of these returns in the pre- and post-crisis periods. The fact that the confidence intervals do not intersect implies that the null hypothesis that the tail indices are the same in the pre- and post-crisis periods: $H_0 : \xi_{\text{pre}} = \xi_{\text{post}}$ is rejected in favor of the alternative that the post-break tail index is smaller: $H_a : \xi_{\text{pre}} > \xi_{\text{post}}$ (at 2.5% level of significance) and thus heavy-tailedness of these returns becomes more pronounced in the post-crash period.

As indicated above, except for the above individual cases, the 95% confidence intervals for the tail indices in (1)–(3) do intersect. Thus, the hypothesis on equality of the tail indices in the pre-crisis and post-crisis periods: $H_0 : \xi_{\text{pre}} = \xi_{\text{post}}$ is not rejected in favor of the alternative $H_a : \xi_{\text{pre}} \neq \xi_{\text{post}}$ at 5% significance level. Thus, the returns' tail indices in the pre-break and post-break periods appear to be statistically indistinguishable, and there is no evidence of statistically significant change in the tail indices due to the 2015 crash for most of the returns.

The results in Fig. 7 and Table 2 and the corresponding results for other companies in the sample emphasize the importance of the structural break in the tail indices of the returns considered. In particular, the 95% confidence intervals for the tail indices in the pre-crash period $CI_{95\%,\text{Pre}}$ in Table 2 are located on the right of 2, thus implying the tail indices greater than 2 and finite variances for the returns' time series. At the same time, the confidence intervals for the post-crash period contain the value of 2 and thus imply possibly infinite variances for the returns. In the full sample, there are 571 more stocks for which the confidence intervals for the returns' tail indices contain the value of two thus implying possibly infinite second moments and variances for the returns. As discussed in the introduction, finiteness of variances and higher moments for financial returns

Table 2. Stocks with Significant Variation of Tail Index Estimator on Absolute Returns.

Stock	Truncation (%)	Pre-Crash			Post-Crash		
		N_{Pre}	$\hat{\xi}_{\text{Pre}}$	$CI_{95\%,\text{Pre}}$	N_{Post}	$\hat{\xi}_{\text{Post}}$	$CI_{95\%,\text{Post}}$
000726 CH Equity	0.15	721	3.65	(2.68, 4.62)	951	2.04	(1.57, 2.51)
002073 CH Equity	0.15	729	4.13	(3.04, 5.22)	941	2.42	(1.86, 2.98)
002433 CH Equity	0.15	746	3.89	(2.88, 4.90)	943	2.22	(1.70, 2.74)
002459 CH Equity	0.15	733	3.99	(2.94, 5.04)	892	2.31	(1.76, 2.86)
002521 CH Equity	0.15	748	3.76	(2.78, 4.74)	944	2.07	(1.59, 2.55)
300206 CH Equity	0.15	755	4.23	(3.13, 5.33)	947	2.47	(1.90, 3.04)
300258 CH Equity	0.15	762	4.15	(3.08, 5.22)	965	2.48	(1.91, 3.05)
600252 CH Equity	0.15	749	3.35	(2.48, 4.22)	876	1.96	(1.49, 2.43)
600509 CH Equity	0.15	750	3.63	(2.68, 4.58)	923	2.07	(1.58, 2.56)
600674 CH Equity	0.15	747	3.54	(2.62, 4.46)	938	2.06	(1.58, 2.54)
600757 CH Equity	0.15	658	3.53	(2.55, 4.51)	952	2.02	(1.55, 2.49)
600969 CH Equity	0.15	751	3.67	(2.72, 4.62)	962	2.14	(1.65, 2.63)
601058 CH Equity	0.15	728	3.87	(2.85, 4.89)	888	2.22	(1.69, 2.75)

and exchange rates is crucial for many models in finance and economics also for applicability of widely used standard statistical and econometric approaches, including the regression and least squares methods.

5. EMPIRICAL ANALYSIS: HEAVY-TAILEDNESS DETERMINANTS

5.1. Tail Index Regressions: Factors Affecting Heavy-Tailedness

The analysis of the determinants of heavy-tailedness of financial returns in this section is based on tail index regressions that relate the (estimates of) tail indices of A-share returns obtained in the previous section to several firm-specific factors. The estimated tail index regressions evaluate the effects of specific characteristics of market participants – Chinese companies – on the degree of heavy-tailedness of the companies' stock returns.⁷

More precisely, we provide estimates of dependence of tail indices of A-share returns on the factors in the following model:

$$\xi_i = f(\text{SOE}_i, \text{FinSect}_i, \text{SSECLList}_i, \text{ForList}_i, \text{Liq}_i, \text{Size}_i, \text{Value}_i, \text{Profit}_i, \text{Invest}_i) \quad (9)$$

where

ξ_i is the tail index of daily returns on shares of company i ,

SOE_i is a dummy that equals one if the government is one of stakeholders of company i , and zero otherwise,

FinSect_i is a dummy variable that equals one if company i belongs to the financial sector,

SSECLList_i is a dummy variable that equals one if company i is listed on the SSE Composite Index,

ForList_i is a dummy variable that equals one if company i is listed in foreign exchanges,

Liq_i is a dummy variable that equals one if the volume weighted average bid-ask spread (in percentage) of company i is in the first decile $D_1(\text{Spread}^{\text{VWA}})$,⁸

Size_i is the natural logarithm of the quarterly average market capitalization of company i ,

Value_i is a dummy variable that equals one if the price-to-book (P/B) ratio characterizing the value of company i is in the ninth decile $D_9(P/\text{Bratio})$,

Profit_i is a dummy variable that equals one if the ROE of company i is in the ninth decile $D_9(\text{ROE})$,⁹

and Invest_i is a dummy variable that equals one if the total investment to total assets (I/A) ratio of company i is in the ninth decile $D_9(I/A \text{ Ratio})$.¹⁰

The analysis is based on OLS regressions of tail index estimates $\hat{\xi}$ on company-specific factors on the right-hand side of (9).¹¹

5.2. Tail Index Regressions: Estimation

5.2.1. Absolute returns

[Table 3](#) presents the estimation results for tail index regressions (9) for tail index estimates $\hat{\xi}_{RS}$ for absolute returns obtained using log–log rank-size regression approach with 15% truncation level, for the whole sample, the pre-crash period and the post-crash period, respectively. The overall F -statistics on the nine firm-specific factors indicate their joint significance at 1% level.

We note that, in the tail regression, a positive value of the estimate of the coefficient on a particular factor implies that an increase in the factor leads, on average, to an *increase* in the *tail index* and thus *decreases* the degree of heavy-tailedness (see the discussion in Section 1.1).

The intercept appears to be highly statistically significant and rather large as compared to the coefficients on the factors in the regressions for all three periods

Table 3. Regression Analysis of Heavy-Tailedness of Absolute Stock Returns on Factors.

	Dependent Variable:		
	Tail Index on Absolute Returns		
	Whole Period	Pre-break	Post-break
Constant	2.500 *** (0.083)	3.065 *** (0.117)	2.615 *** (0.107)
SOE	-0.112 *** (0.015)	-0.119 *** (0.022)	-0.104 *** (0.019)
FinSect	-0.099 *** (0.050)	-0.235 *** (0.075)	-0.021 (0.065)
SSECList	-0.088 *** (0.015)	-0.076 *** (0.023)	-0.077 *** (0.019)
ForList	-0.059 *** (0.027)	-0.153 *** (0.041)	-0.023 (0.034)
Liquid	0.111 *** (0.027)	0.110 *** (0.040)	0.175 *** (0.034)
Size	0.030 *** (0.009)	0.011 (0.014)	0.002 (0.012)
Value	0.125 *** (0.024)	0.107 *** (0.036)	0.162 *** (0.030)
Profit	0.083 *** (0.025)	0.065 * (0.038)	0.065 ** (0.032)
Invest	-0.014 (0.027)	0.002 (0.040)	-0.049 (0.034)
Observations	1,038	1,038	1,038
R^2	0.221	0.121	0.152
Adjusted R^2	0.214	0.114	0.144
F Statistic _(9,1028)	32.429 ***	15.763 ***	20.425 ***

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

considered.¹² This suggests that heavy-tailedness of Chinese stock returns is more influenced by general market trends rather than the individual company characteristics. This conclusion is similar to heavy-tailedness properties of the Russian stock market (see [Ankudinov et al., 2017](#)) and to the fact that, according to the analysis in [Morck et al. \(2000\)](#) and [Jin and Myers \(2006\)](#), in emerging economies, the country-level volatility of returns turns out to be significantly higher compared to the firm-specific volatility.

The regression estimates in [Table 3](#) point to the higher level of heavy-tailedness for financial sector companies as compared to those compared to those representing the real sector of economy, with statistically significant negative values of the coefficient at the variable FinSect in regressions for the whole sample and the pre-crash period. The effect is expected as higher degree of financial leverage results in higher returns volatility of financial intermediaries. The difference between the degree of heavy-tailedness of the returns of financial sector companies as compared to those in the real sector appears to be particularly large and significant in the pre-crisis period, while it is not statistically significant in the post-crash period. According to the estimates, during the pre-crash period, the financial sector companies had on average tail indices smaller by 0.235 as compared to the real sector companies. These conclusions are apparently due to the fact that the bull market before the crash was led by financial companies and banks. We also note that statistical significance of the coefficient is in contrast to the conclusions for tail index regressions in the case of Russian stock market [Ankudinov et al. \(2017\)](#) where it is in general absent.

Stock liquidity appears to reduce the degree of returns' heavy-tailedness. The effect is expected since, as discussed, for example, in [Ankudinov et al. \(2017\)](#), all other factors held equal, actively traded stocks are expected to have lower volatility due to the regular and prompt revaluation by the market as any significant difference between a stock's fundamental and market values emerges. In contrast, for illiquid shares, rare and limited in value transactions may significantly affect their market prices. The effect of (il)liquidity on the degree of returns' heavy-tailedness is statistically significant in all the periods considered. The tail indices of the returns on stocks of the most liquid companies (with the volume-weighted bid-ask spread percentage in the 10th percentile of the sample) are, on average, greater than those for less liquid companies by about 0.11 in the pre-crash period and over the whole time period considered. Illiquidity amplifies heavy-tailedness in the post-crash bear market with a larger average difference (of 0.175) between the returns on stocks of the most liquid and less companies.

The returns on stocks of larger companies have, on average, lower degree of heavy-tailedness as compared to those for smaller one. The effect of company size on the degree of heavy-tailedness is statistically significant in regressions over the whole time period considered. The above negative relation between the degree of heavy-tailedness and companies size is expected since, as discussed in [Ankudinov et al. \(2017\)](#), large firms are generally more stable financially due to government support for very large “too big to fail” companies, better access to capital markets and diversified sources of funding. These effects are stronger in emerging markets under financial constraints resulting from, among other factors, the lesser

development of financial intermediation. Higher stability of large firms tends to reduce volatility of their stock prices, including that in the tails.

According to the tail index regressions, state owned companies tend to have, on average, higher degree of heavy-tailedness in their stock returns as compared to other companies. On average, higher degree of heavy-tailedness is also observed for companies listed on the SSE Composite Index SSE and on foreign exchanges. These conclusions are surprising since state-owned enterprises are expected to be more stable, especially in emerging markets like China and Russia (see the discussion in [Ankudinov et al., 2017](#)) due to government's preferential policies, low-cost funding, etc. and thus to have less volatile stock prices. In addition, listings on the SSE or foreign exchanges would be expected to reduce the degree of heavy-tailedness due to their regulatory requirements on corporate governance, financing, transparency, and other factors that favor more stable companies. The analysis of the above conclusions from the empirical results is an important problem that is left for further research.

Higher value and profitability indicated by a higher price-to-book (P/B) ratio and higher ROE tend to be associated with lower degrees of heavy-tailedness, with statistically significant coefficients on these regressors over all the periods considered. The analysis of the effects of these variables on the degree of heavy-tailedness and explanations for them merits a further investigation.

The investment factor does not appear to be statistically significant in any of the regression models considered.

5.2.2. One-Sided Returns

[Table 4](#) presents the estimation results for analogues of tail regressions dealt with in the previous section for the right and left tail indices ξ^+ and ξ^- in power law models (1) and (2). Again, the analysis is based on log-log rank-size regression estimates $\hat{\xi}_{RS}^+$ and $\hat{\xi}_{RS}^-$ with 15% truncation regressed on firm-specific factors in (9), As in the previous section, the tail index regression estimates are provided for the whole time period considered as well as for the pre-crash and post-crash periods.

As in the case of the analysis of tail index regressions for absolute returns, the overall F -statistics appear to be significant in all the periods considered. In addition, similar to the previous section, high significance is observed for the intercepts in all the regressions, pointing out to the apparently greater importance of general market trends rather than the individual company characteristics for heavy-tailedness properties (and thus the likelihood and the magnitude) of both the large downfalls and large upward movements in the returns and price time series in the three periods. Importantly, a statistically significant change is observed in the intercept of the regression for left tail indices, with its point estimates declining from a rather high value of 4.2 indicating rather low degree of heavy-tailedness pre-crash to the value of 3.1.

As expected, the magnitude of essentially all the coefficients in tail index regressions estimated separately for the returns' right and left tail indices in [Table 4](#)

Table 4. Tail Index Estimates for Positive and Negative Returns.

	Dependent Variable:					
	Tail Index on One-side Returns					
	Left Tail			Right Tail		
	Whole Period	Pre-break	Post-break	Whole Period	Pre-break	Post-break
Constant	2.663 *** (0.087)	4.175 *** (0.141)	3.089 *** (0.116)	2.189 *** (0.114)	2.139 *** (0.147)	2.116 *** (0.139)
SOE	-0.085 *** (0.015)	-0.081 *** (0.027)	-0.053 ** (0.020)	-0.133 *** (0.020)	-0.141 *** (0.028)	-0.145 *** (0.024)
FinSect	0.015 (0.052)	-0.296 *** (0.090)	0.114 (0.070)	-0.211 *** (0.069)	-0.174 * (0.094)	-0.241 *** (0.084)
SSECList	-0.073 *** (0.015)	-0.055 ** (0.027)	-0.061 *** (0.020)	-0.095 *** (0.020)	-0.097 ** (0.029)	-0.093 *** (0.025)
ForList	-0.012 (0.028)	-0.098 ** (0.049)	0.014 (0.037)	-0.106 *** (0.037)	-0.197 *** (0.051)	-0.079 * (0.044)
Liquid	0.108 *** (0.028)	0.077 (0.048)	0.217 *** (0.037)	0.118 *** (0.037)	0.150 *** (0.050)	0.131 *** (0.044)
Size	-0.008 (0.010)	-0.085 *** (0.017)	-0.072 *** (0.013)	0.076 *** (0.013)	0.094 *** (0.018)	0.080 *** (0.015)
Value	0.119 *** (0.025)	0.074 * (0.043)	0.115 *** (0.033)	0.079 ** (0.033)	0.118 *** (0.045)	0.126 *** (0.039)
Profit	0.044 * (0.027)	-0.022 (0.045)	0.062 * (0.035)	0.168 *** (0.035)	0.099 ** (0.047)	0.104 ** (0.042)
Invest	-0.036 (0.028)	-0.058 (0.048)	-0.046 (0.037)	-0.002 (0.037)	0.038 (0.050)	-0.060 (0.044)
Observations	1,038	1,038	1,038	1,038	1,038	1,038
R ²	0.133	0.129	0.097	0.207	0.132	0.161
Adjusted R ²	0.125	0.122	0.089	0.200	0.124	0.153
F Statistic (df = 9; 1028)	17.512***	16.952***	12.250***	29.821***	17.340***	21.852***

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

are smaller than those for the regressions for tail indices of absolute returns in the previous section.

The analysis of confidence intervals for the coefficients in regressions in [Table 4](#) does not indicate their statistically significant difference for the right and left tails.

In addition, importantly, with the exception of the intercept, the confidence intervals for all the coefficients in the regressions in the pre-crash and post-crash periods intersect. This points out to absence of statistically significant changes in the regressions' coefficients due to the 2015 crash, and in the effects of firm-specific factors on heavy-tailedness properties of large downfalls and large upward movements in the return and price time series.

6. CONCLUSION

Many studies in the literature have focused on the analysis of stylized facts of developed financial markets, including heavy-tailedness properties of financial returns and foreign exchange rates (see, among others, the review in [Cont \(2001\)](#)). The empirical research mostly agrees, in particular, that, in the case of developed markets, tail indices of financial returns and foreign exchange rates lie in the interval (2,4), thus implying finite variances and infinite fourth moments. At the same time, the research on heavy-tailedness properties of emerging financial markets still remains limited.

This chapter partially fills this gap in the literature by providing a detailed analysis of heavy-tailedness properties of financial returns on more than a thousand A-shares in China using the recently developed robust tail index inference approaches. Among other results, the analysis points to the tail index estimate smaller than two for several stocks considered, thus implying possibly infinite second moments and variances. These results are particularly important as infinite second moments and variances lead to the inapplicability of many classical econometric and statistic models including least-squares analysis and (auto-) correlation based approaches, while standard auto-correlation based methods require infinite fourth moments (see, among others, the discussion in [Davis & Mikosch, 1998](#); [Cont, 2001](#); [Ibragimov et al., 2015](#); [Mikosch & Starica, 2000a](#)).

We further focus on the analysis of (a)symmetry and structural breaks in heavy-tailedness properties of Chinese financial markets due to the 2015 crash. The returns' distributions display some (gain/loss) asymmetry in their right and left heavy-tailedness properties that varies before and after the crash, with a relatively fatter right tail observed before the 2015 market crash (see [Fig. 5](#)). However, none of the stocks shows a statistically significant difference between the degree of heavy-tailedness in the left and right tails.

In terms of the structural breaks in the tail indices, differences in the degree of heavy-tailedness in the pre-crash and post-crash periods generally exist, and we find 3 cases where left tails become heavier statistically significantly and 13 cases where the tails of the returns' large fluctuations of either sign get fatter after the break (see [Fig. 7](#) and [Table 2](#)). Importantly, the break causes fluctuations in the tail indices for most of the stocks that imply infinite second moments.

Motivated, in part, by the analysis in [Fama and French \(2015\)](#) and [Ankudinov et al. \(2017\)](#), the chapter also provides a detailed regression analysis of firm-specific characteristics and attributes affecting heavy-tailedness properties of returns on their A-shares before and after the break. Important effects on the returns' heavy-tailedness properties are observed, in particular, for the momentum, liquidity, and company size. The analysis also points out to the importance of government ownership, sector affiliation, and dual listing as determinants of returns' heavy-tailedness properties.

Further research may focus on applications of t -statistic approaches in [Ibragimov and Müller \(2010, 2016\)](#) to robust inference on coefficients in tail index regressions (see Section 5.2). It may also focus on applications of the above approaches and QLR-type procedures in robust tests for structural breaks in

tail indices and the coefficients at their determinants, possibly with unknown date. The research in these directions is currently under way by the authors and co-authors, and will be presented elsewhere.

NOTES

1. Naturally, therefore, the standard OLS regression methods and autocorrelation-based time series analysis methods are in principle inapplicable directly and need to be modified in the case of heavy-tailed time series with tail indices ξ smaller than two and infinite (or undefined) variances.

2. The property that the financial returns' tail indices are smaller than 4 and their fourth moments are infinite implies that the use of the common measure of heavy-tailedness, the kurtosis, is inappropriate: For example, under $\xi \in (2, 4)$, its estimate given by the sample kurtosis diverges to infinity in probability as the sample size grows. Thus, the sample kurtosis of financial returns is expected to take on increasingly larger values as the sample size increases (see the results and the discussion in [Han & Park, 2008](#); [Park, 2002](#)).

3. See also [Koedijk et al. \(1992\)](#) for the analysis of changes in the tail behavior of Latin American foreign exchange rates in response to changes in foreign exchange rate regimes.

4. The standard errors and confidence intervals for the tail index estimators dealt with are asymptotically valid under the assumption that r_t are i.i.d. observations from power laws (1)–(3). As discussed below, the log–log rank-size regression approach to tail estimation appears to be more robust to dependence in observations r_t , including GARCH-type dependence typical for financial returns, as compared to that based on Hill's tail index estimates (see also the discussion in [Gabaix & Ibragimov, 2011](#) and the online appendix to that paper).

5. In particular, the log–log rank-size regression tail index estimation approaches appear to be more robust than those based on Hill's estimates to nonlinear dependence in the form of GARCH-type volatility dynamics as in the case of real-world economic and financial and to deviations from power laws (1)–(3) in the form of slowly varying functions (see the discussion in [Gabaix & Ibragimov, 2011](#) and the numerical results in the online appendix to that paper, https://scholar.harvard.edu/files/xgabaix/files/rank_1_2_appendix.pdf)

6. See [Ibragimov et al. \(2013\)](#) and [Chen and Ibragimov \(2019\)](#) for applications of the approach in the analysis of structural breaks due to the beginning of the 2008 global financial crisis in the degree of heavy-tailedness of developed and emerging country foreign exchange rates and the returns on several A- and H-shares in China.

7. See [Ankudinov et al. \(2017\)](#), for estimates of related tail index regressions for financial markets in Russia.

8. The volume weighted average daily bid-ask spread percentage used in the analysis characterizes how liquid an underlying asset is. The definition of the liquidity dummy Liq_i using the 10th percentile of the volume weighted bid-ask spread percentage is motivated by [Fama and French \(2015\)](#). [Ankudinov et al. \(2017\)](#) regards a stock as liquid if its weighted average bid-ask spread percentage is less than 0.04 with an average number of transactions of minimum 1,000 per month.

9. The literature has also used other characteristics of a company's profitability, including the earnings yield, expected profitability, and others.

10. The value one for $Invest_i$ thus suggests that the underlying company invests in an aggressive style.

11. As indicated above, due to it is robustness as compared to Hill's estimates, tail index regressions (9) in the next section are based on log–log rank-size regression tail index estimates.

12. One should that the reported standard errors of the coefficients of the tail index regressions and the comparisons and the analysis of their significance should be considered as only indicative as the standard errors do not account for uncertainty in (the first-stage) estimation of tail indices that are used as dependent variables in the regressions. A more

detailed analysis of the effects of different factors on the degree of heavy-tailedness and tail indices of financial returns would require extensions of recently developed maximum likelihood-type methods for power laws with factor-dependent tail indices (see Ma et al., 2019; Wang & Tsai, 2009, and references therein) to the case of time series. The analysis may also use robust t -statistic inference approaches developed in Ibragimov and Müller (2010, 2016) that do not require consistent estimation of limiting variances of estimators dealt with (e.g., those of tail index regression coefficients). Applications of these approaches are left for further research.

ACKNOWLEDGMENTS

We thank the Editors, an anonymous referee and the participants at the seminar series at the Centre for Econometrics and Business Analytics (CEBA), St. Petersburg University, and iCEBA-2021, 2022 conferences for helpful comments and suggestions. Rustam Ibragimov's research for this chapter was supported in part by a grant from the Russian Science Foundation (Project No. 22-18-00588).

REFERENCES

- Anatolyev, S. (2019). Volatility filtering in estimation of kurtosis (and variance). *Dependence Modeling*, 7, 1–23.
- Ankudinov, A., Ibragimov, R., & Lebedev, O. (2017). Heavy tails and asymmetry of returns in the Russian stock market. *Emerging Markets Review*, 32, 200–219.
- Beirlant, J., Goegebeur, Y., Teugels, J., & Segers, J. (2004). *Statistics of extremes: Theory and applications*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd.
- Candelon, B., & Straetmans, S. (2006). Testing for multiple regimes in the tail behavior of emerging currency returns. *Journal of International Money and Finance*, 25, 1187–1205.
- Chang, E. C., Luo, Y., & Ren, J. (2014). Short-selling, margin-trading, and price efficiency: Evidence from the Chinese market. *Journal of Banking & Finance*, 48, 411–424.
- Chen, Z., & Ibragimov, R. (2019). One country, two systems? The heavy-tailedness of Chinese A-and H-share markets. *Emerging Markets Review*, 38, 115–141.
- Cho, D. D., Russell, J., Tiao, G. C., & Tsay, R. (2003). The magnet effect of price limits: evidence from high-frequency data on Taiwan stock exchange. *Journal of Empirical Finance*, 10(1–2), 133–168.
- Chung, H., & Park, J. Y. (2007). Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics*, 137, 230–259.
- Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1, 223–236.
- Davis, R. A., & Mikosch, T. (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *The Annals of Statistics*, 26, 2049–2080.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*, Vol. 33 of *Applications of Mathematics*. Springer.
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116, 1–22.
- Gabaix, X. (2009). Power laws in economics and finance. *Annual Review of Economics*, 1, 255–293.
- Gabaix, X. (2016). Power laws in economics: An introduction. *Journal of Economic Perspectives*, 30, 185–206.
- Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2003). A theory of power-law distributions in financial market fluctuations. *Nature*, 423, 267–270.
- Gabaix, X., Gopikrishnan, P., Plerou, V., & Stanley, H. E. (2006). Institutional investors and stock market volatility. *Quarterly Journal of Economics*, 121, 461–504.

- Gabaix, X., & Ibragimov, R. (2011). Rank-1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29, 24–39.
- Granger, C. W. J., & Orr, D. (1972). Infinite variance and research strategy in time series analysis. *Journal of the American Statistical Association*, 67, 275–28.
- Gu, Z., & Ibragimov, R. (2018). The “Cubic law of the stock returns” in emerging markets. *Journal of Empirical Finance*, 46, 182–190.
- Han, H., & Park, J. Y. (2008). Times series properties of ARCH processes with persistent covariates. *Journal of Econometrics*, 146, 275–292.
- Huisman, R., Koedijk, K. G., Kool, C. J. M., & Palm, F. (2001). Tail-index estimates in small samples. *Journal of Business & Economic Statistics*, 19, 208–216.
- Ibragimov, M., Ibragimov, R., & Kattuman, P. (2013). Emerging markets and heavy tails. *Journal of Banking & Finance*, 37, 2546–2559.
- Ibragimov, M., Ibragimov, R., & Walden, J. (2015). *Heavy-tailed distributions and robustness in economics and finance, Vol. 214 of lecture notes in statistics*. Springer.
- Ibragimov, R., & Müller, U. K. (2010). *t*-statistic based correlation and heterogeneity robust inference. *Journal of Business and Economic Statistics*, 28, 453–468.
- Ibragimov, R., & Müller, U. K. (2016). Inference with few heterogeneous clusters. *Review of Economics and Statistics*, 98(1), 83–96.
- Ibragimov, R., Pedersen, R. S., & Skrobotov, A. (2020). *New approaches to robust inference on market (non-)efficiency, volatility clustering and nonlinear dependence* [Working paper]. Imperial College Business School and Department of Economics, University of Copenhagen. <https://arxiv.org/abs/2006.01212>
- Ibragimov, R., & Prokhorov, A. (2017). *Heavy tails and copulas: Topics in dependence modelling in economics and finance*. World Scientific.
- Jiang, F., Jiang, Z., & Kim, K. A. (2017). Capital markets, financial institutions, and corporate finance in China. *Journal of Corporate Finance*, 63, Article 101309.
- Jin, L., & Myers, S. (2006). R^2 around the world: New theory and new tests. *Journal of Financial Economics*, 79, 257–292.
- Koedijk, K., Stork, P., & de Vries, C. (1992). Differences between foreign exchange rate regimes: The view from the tails. *Journal of International Money and Finance*, 11, 462–473.
- Loretan, M., & Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, 1, 211–248.
- Ma, Y., Jiang, Y., & Huang, W. (2019). Tail index varying coefficient model. *Communications in Statistics – Theory and Methods*, 48, 235–256.
- McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools* (revised ed.). Princeton University Press.
- Mikosch, T., & Starica, C. (2000a). Is it really long memory we see in financial returns. *Extremes and Integrated Risk Management*, 12, 149–168.
- Mikosch, T., & Starica, C. (2000b). Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process. *Annals of Statistics*, 28, 1427–1451.
- Miller, J. I., & Park, J. (2010). Nonlinearity, nonstationarity, and thick tails: How they interact to generate persistence in memory. *Journal of Econometrics*, 155, 83–89.
- Morck, R., Yeung, B., & Yu, W. (2000). The information content of stock markets: Why do emerging markets have synchronous stock price movements? *Journal of Financial Economics*, 58, 215–260.
- Park, J. Y. (2002). Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics*, 110, 383–415.
- Quintos, C., Fan, Z., & Phillips, P. C. (2001). Structural change tests in tail behaviour and the Asian crisis. *Review of Economic Studies*, 68, 633–663.
- Sornette, D., Demos, G., Zhang, Q., Cauwels, P., Filimonov, V., & Zhang, Q. (2015). *Real-time prediction and post-mortem analysis of the Shanghai 2015 stock market bubble and crash* [Swiss Finance Institute Research Paper No. 15-31]. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2693634
- Wang, H., & Tsai, C.-L. (2009). Tail index regression. *Journal of the American Statistical Association*, 104, 104–487.
- Zeng, F., Huang, W.-C., & Hueng, J. (2016). On Chinese government’s stock market rescue efforts in 2015. *Modern Economy*, 7, 411–418.

This page intentionally left blank

PART III

PANDEMIC, CLIMATE, AND DISASTER

This page intentionally left blank

CHAPTER 8

PREDICTING CRASHES IN OIL PRICES DURING THE COVID-19 PANDEMIC WITH MIXED CAUSAL-NONCAUSAL MODELS

Alain Hecq and Elisa Voisin

Maastricht University, Netherlands

ABSTRACT

This chapter aims at shedding light upon how transforming or detrending a series can substantially impact predictions of mixed causal-noncausal (MAR) models, namely dynamic processes that depend not only on their lags but also on their leads. MAR models have been successfully implemented on commodity prices as they allow to generate nonlinear features such as locally explosive episodes (denoted here as bubbles) in a strictly stationary setting. The authors consider multiple detrending methods and investigate, using Monte Carlo simulations, to what extent they preserve the bubble patterns observed in the raw data. MAR models relies on the dynamics observed in the series alone and does not require economical background to construct a structural model, which can sometimes be intricate to specify or which may lack parsimony. The authors investigate oil prices and estimate probabilities of crashes before and during the first 2020 wave of the COVID-19 pandemic. The authors consider three different mechanical detrending methods and compare them to a detrending performed using the level of strategic petroleum reserves.

Keywords: Noncausal models; detrending; forecasting; predictive densities; bubbles; crashes; simulations-based forecasts; Hodrick-Prescott filter; COVID-19 pandemic

JEL Codes: C22; C53

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 209–233

Copyright © 2023 by Alain Hecq and Elisa Voisin

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:[10.1108/S0731-90532023000045B010](https://doi.org/10.1108/S0731-90532023000045B010)

1. INTRODUCTION

This chapter aims at forecasting Brent and WTI oil price series during the first wave of the COVID-19 pandemic outbreak in 2020 using the recent literature on mixed causal-noncausal autoregressive models (hereafter *MAR*). Namely, time series processes with lags but also leads components and non-Gaussian errors. This new specification can, in a parsimonious way, model locally explosive episodes in a strictly stationary setting. It can therefore capture nonlinear features such as bubbles (which is defined here as a persistent increase followed by a sudden crash), often observed in commodities prices, while standard linear autoregressive models (e.g., *ARMA* models) cannot do so. *MAR* models have successfully been implemented on several commodity price series (see *inter alia* [Cubadda et al., 2019](#); [Fries & Zakoian, 2019](#); [Gouriéroux & Zakoian, 2017](#); [Hecq & Voisin, 2021](#); [Hecq et al., 2020](#); [Karapanagiotidis, 2014](#); [Lof & Nyberg, 2017](#)).¹ Similarly to [Gouriéroux and Zakoian \(2013\)](#), our goal when introducing a lead component in oil prices is not to provide an economic justification for the existence of a rational bubble. However, the link with a present value model between prices and dividends [Campbell and Shiller \(1987\)](#) can enrich the discussion and it also explains the difficulties to find economic fundamentals for oil prices. This motivates our choice to use proxies such as technical methods to extract the bubble component. Let us indeed consider a general model (see [Diba & Grossman, 1988](#)) in which the real current stock price P_t is linked to the present value of next period's expected stock price P_{t+1} , dividend payments D_{t+1} and an unobserved variable u_{t+1} ,

$$P_t = \frac{1}{1+r} \mathbb{E}_t [P_{t+1} + \alpha D_{t+1} + u_{t+1}], \quad (1)$$

with \mathbb{E}_t the conditional expectation given the information set known at time t .

The discount factor is $\frac{1}{1+r}$ with r being a time-invariant interest rate. The general solution of (1) is (e.g., [Diba & Grossman, 1988](#))

$$P_t = \sum_{i=1}^{\infty} \left(\frac{1}{1+r} \right)^i \mathbb{E}_t [\alpha D_{t+i} + u_{t+i}] + B_t = P_t^F + B_t, \quad (2)$$

where the actual price deviates from its fundamental value P_t^F by the amount of the rational bubble B_t . As shown by [Gourieroux et al. \(2020\)](#), *MAR* processes provide stationary solutions for the modeling of the bubbles component in (2) (see also [Fries, 2021](#)).

However, oil prices are challenging time series to forecast and model (see [Baumeister & Kilian, 2016](#) and for a survey on oil prices forecasting see [Alquist et al., 2013](#)). Unlike for equity prices, measuring commodities fundamentals might not be as straightforward [Brooks et al. \(2015\)](#), [Pindyck \(1993\)](#) and [Alquist and Kilian \(2010\)](#) consider the convenience yield, that is, a premium associated with holding an underlying product instead of derivative securities or contracts. It typically increases when costs associated with physical storage are low. Yet, not only is the convenience yield not easy to measure but there also are other factors driving each of the demand and supply side of crude oil: the level of stocks, economic activity, geopolitical considerations, shifts in expectations regarding the

oil market, etc. While there is a large literature on modeling and forecasting the price of oil using structural models that incorporate economic fundamentals (see Kilian & Zhou, 2020b), our model is parsimonious and exploits the statistical properties of oil prices only.

As shown in Fig. 3 in Section 4, oil prices series do not appear to be stationary over time. Consequently, before estimating *MAR* models we intend to extract a smooth time-varying trend to render the series stationary without affecting the dynamics. By extracting a trend from the series we do not claim to identify the fundamental values of oil prices but instead detrend the series while preserving the dynamics of the prices in the remaining cycle and more specifically the noncausal component. As such, we obtain stationary series that retain their forward-looking aspect and which can be modeled as *MAR* processes. Obviously, a wrong detrending can give misleading results if it alters the dynamics of the cycle. Consequently, investigating the impact of different technical detrending filters on the identification of *MAR* models is the first contribution of this chapter. Similarly to what Canova (1998) does for business cycles, we investigate the extent to which the identification of causal and noncausal dynamics are sensitive to different filters. We then study the consequences on the predictive densities of oil prices after applying different detrending methods. Inspired by the work of Kilian and Murphy (2014), who constructed a structural VAR model of the global market for crude oil, we make use of US crude oil Strategic Petroleum Reserve (SPR), a sub-part of total petroleum stocks, for a potential trend in oil prices in Section 4. Hence, the second contribution of this chapter is to compare the *MAR* estimations and predictions of oil price series after using technical detrending with the results obtained after detrending with the SPR levels.

The rest of this chapter is as follows. Section 2 describes mixed causal-noncausal models and explains the different technical detrending methods employed in this analysis, leaving the locally explosive components in the cycle. In Section 3, the impact of the different detrending filters on model identifications is investigated using a Monte Carlo study, based on trends estimated in oil prices series. We investigate the identification of the models but also the magnitude of the coefficients estimated as they are the main drivers of the predictions. Section 4 analyzes the impact of these filters on the WTI and the Brent crude oil price series for ex-post and real-time analyses. We compare the results with those obtained after detrending with US SPR levels. We show how each detrending approach affects probabilities that oil price crashes in the period capturing the first 2020 wave of the COVID-19 pandemic. Section 5 concludes.

2. MIXED CAUSAL-NONCAUSAL MODELS AND FILTERING

2.1. The Model

MAR(r, s) denotes dynamic processes that depend on their r lags as for usual autoregressive processes but also on their s leads in the following multiplicative form

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t, \quad (3)$$

with L the backward operator, i.e., $Ly_t = y_{t-1}$ gives lags and $L^{-1}y_t = y_{t+1}$ produces leads. When $\Psi(L^{-1}) = (1 - \psi_1 L^{-1} - \dots - \psi_s L^{-s}) = 1$, namely when $\psi_1 = \dots = \psi_s = 0$, the univariate process y_t is a purely causal autoregressive process, denoted $MAR(r,0)$ or simply $AR(r)$ model, $\Phi(L)y_t = \varepsilon_t$. Reciprocally, the process is a purely noncausal $MAR(0,s)$ model $\Psi(L^{-1})y_t = \varepsilon_t$, when $\phi_1 = \dots = \phi_r = 0$ in $\Phi(L) = (1 - \phi_1 L - \dots - \phi_r L^r)$. The roots of both the causal and noncausal polynomials are assumed to lie outside the unit circle, that is $\Phi(z) = 0$ and $\Psi(z) = 0$ for $|z| > 1$ respectively. These conditions imply that the series y_t admits a two-sided moving average representation $y_t = \sum_{j=-\infty}^{\infty} \gamma_j \varepsilon_{t-j}$, such that $\gamma_j = 0$ for all $j < 0$ implies a purely causal process y_t (with respect to ε_t) and a purely noncausal model when $\gamma_j = 0$ for all $j > 0$ [Lanne and Saikkonen \(2011\)](#). Error terms ε_t are assumed *iid* (and not only weak white noise) non-Gaussian (with potentially infinite variance) to ensure the identifiability of the causal and the noncausal parts ([Breidt et al., 1991](#); [Gouriéroux & Zakoïan, 2015](#)). While noncausal models are strictly stationary, their conditional moments are time-varying. A purely stationary noncausal $MAR(0,1)$ Cauchy-distributed process, has a unit root in its conditional mean and exhibit ARCH-type effects (see [Cavaliere et al., 2018](#); [Gouriéroux & Zakoïan, 2017](#)).

Fig. 1 shows a purely causal (a) and a purely noncausal (b) trajectories induced by the same Student's $t(2)$ -distributed errors, both with coefficient 0.8 and 200 observations. For the purely causal process, a shock is unforeseeable and affects the series only once it happened, inducing a large jump in the series. On the other hand, for purely noncausal processes, a shock impacts the process ahead of time, mirroring the purely causal trajectory. Indeed, we see that the series already reacts to a positive shock by increasing until a sudden crash, creating bubble patterns. This anticipative aspect is widely observed in financial and economics time series. The detrended Brent crude oil prices as shown in **Fig. 5** noticeably exhibit such features, the most apparent episode being the 2008 financial crisis. A combination

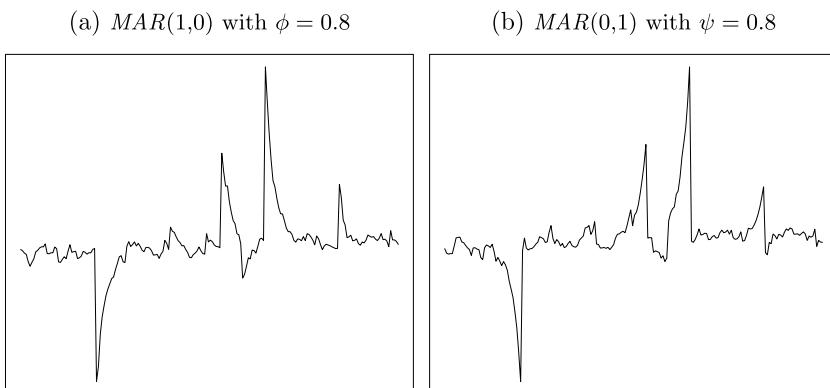


Fig. 1. Purely Causal (a) and Noncausal (b) Trajectories.

of causal and noncausal dynamics consequently creates some asymmetry around a shock, varying with the magnitude of the respective coefficients.

The advantage with oil prices is that they already underwent bubbles in the past, and those previous locally explosive episodes will help identifying *MAR* models. In the case where series are for the first time following a long and abnormal increase, an explosive process is difficult to distinguish from a stationary locally explosive one.

The focus of this chapter is on the probabilities of crashes. Predictions are performed using the approximation methods of [Gouriéroux and Jasiak \(2016\)](#) and [Lanne et al. \(2012\)](#) since no closed-form of the predictive density exists when the errors of the process follow a Student's *t* distribution. For a detailed analysis of the two approximation methods see [Hecq and Voisin \(2021\)](#).²

2.2. Filtering the Data

The requirement of y_t being stationarity for both lag and lead polynomials gave rise to different strategies to transform nonstationary series to stationary ones. [Hecq et al. \(2020\)](#) and [Cubadda et al. \(2019\)](#) assume³ that their commodity price series are $I(1)$ and work with the returns Δy_t . However, this operation eliminates most of the locally explosive behaviors and the transformed series consist of many spikes instead.

In this chapter, we capture the trending behavior of the observed series denoted \tilde{y}_t in different ways using the general form

$$\tilde{y}_t = f_t + y_t,$$

where

$$\Phi(L)\Psi(L^{-1})y_t = \varepsilon_t.$$

In this framework, \tilde{y}_t is the (potentially nonstationary) observed series and f_t a generic trend function. The deviation of \tilde{y}_t from its trend is an $MAR(r, s)$ process. Several authors, although sometimes not explicitly, use this decomposition. [Cavaliere et al. \(2018\)](#) opt for the choice of a particular time period with no trend and hence use only an intercept $f_t = \mu$. [Hencic and Gouriéroux \(2015\)](#) detrend \tilde{y}_t using a polynomial trend function of order three. In summary, we could consider several choices among the following deterministic trends,

$$\begin{aligned} f_t^{(1)} &= \mu, \\ f_t^{(2)} &= \mu + \beta D_t, \quad \text{with } D_t = 1 \text{ when } t \geq t_{break} \text{ and } 0 \text{ otherwise,} \\ f_t^{(3)} &= \alpha_0 + \alpha_1 t + \dots + \alpha_k t^k, \quad \text{with } k \text{ some positive integer and } t = 1, 2, \dots, T. \end{aligned}$$

Note (see Section 4) that since a larger order of polynomial allows for more flexibility, we consider polynomial trends of order four and six for the trending pattern of the monthly oil prices series considered in this analysis. More complex trends, constructed as a combination of the aforementioned examples could also be considered, such as (multiple) breaks in trends for instance.

Hecq and Voisin (2021) use the Hodrick-Prescott (HP) filter before detecting bubbles in Nickel monthly prices. The HP filter, as opposed to the aforementioned deterministic trends, extracts the trend process $f_t^{(4)}$ via a minimization that relies on a penalizing parameter denoted λ .

$$\min_{\{f_t^{(4)}\}_{t=1}^T} \left\{ \sum_{t=1}^T (\tilde{y}_t - f_t^{(4)})^2 + \lambda \sum_{t=3}^T [(f_t^{(4)} - f_{t-1}^{(4)}) - (f_{t-1}^{(4)} - f_{t-2}^{(4)})]^2 \right\}.$$

The larger this parameter, the smoother the trend component is (i.e., with λ approaching infinity, the extracting trend becomes linear). For details about the HP filter see Hodrick and Prescott (1997). It is now commonly accepted to use $\lambda = 1,600$ for quarterly data. For other frequencies, the rule of thumb consists in adjusting the parameter to the frequency relative to quarterly data,

$$\lambda = \left(\frac{\text{number of observations per year}}{4} \right)^i \times 1,600,$$

with either $i = 2$ (Backus & Kehoe, 1992) or $i = 4$ (Ravn & Uhlig, 2002), yielding respectively a penalizing parameter of 14,400 and 129,600 for monthly series. Most criticisms of the HP filter concern its application on series with complicated stochastic and deterministic trends. Phillips and Shi (2019) propose an adaptation of the filter improving its accuracy for such series.⁴ We investigate in Section 3 the potential dynamic distortions that can be induced by HP filtering (see among others Hamilton, 2018) but find no significant distortions of the mixed causal-noncausal dynamics.

Note that we are not interested in the exact value of a forecast but rather in its direction and potential magnitude. This is why we extract smooth trends to preserve the dynamics in the series. This allows to estimate predictive densities of oil prices based on the statistical properties of the data alone in a parsimonious way, and not from the construction of complicated structural models. However, wrongly detrending the series could have a significant impact on the estimation of the noncausal dynamics of the process, which could in turn strongly under- or over-estimate the longevity of explosive episodes and therefore of the probabilities of crashes and of turning points.

3. MONTE CARLO ANALYSIS – EFFECTS OF DETRENDING

The aim of this section is to analyze the effect of wrongly detrending a series, both on the identification of the *MAR* model and on the subsequent predictions performed with the resulting model. We base this analysis on stylized facts observed in oil prices series.

3.1. Accuracy of Detrending

We simulate 5,000 trajectories for 12 distinct data generating processes (hereafter *dgp*), composed of a trend and a stationary dynamic process denoted as cycle. All *dgps* are generated by Student's *t*-distributed errors with 2 degrees of freedom,

a value frequently observed in financial time series, and with 400 observations. For the cycles, we consider purely noncausal processes with a lead coefficient of 0.8, purely causal processes with a lag coefficient of 0.6 and mixed causal-noncausal processes with a lag coefficient of 0.6 and a lead coefficient of 0.8. The heavy-tailed distribution generates extreme values, inducing bubble-like phenomena in processes with noncausal components. We are interested in mostly forward looking processes characterized by long lasting bubbles hence the choice of coefficients. We consider three different deterministic trends: a linear trend with breaks (denoted *breaks*) and two trend polynomials up to orders 4 and 6 (denoted respectively τ^4 and τ^6 for simplicity). The coefficients of the trends were estimated on the monthly WTI crude oil prices series between 1986 and 2019. Fig. 2 depicts the three mentioned trends to which purely causal, noncausal and mixed causal-noncausal trajectories are added. Additionally, we consider processes with an intercept only. This results overall in 12 sets of 5,000 trajectories of the form $\tilde{y}_t = f_t + y_t$.

Four detrending methods are employed for each trajectories, with the general form $\tilde{y}_t = \hat{f}_t + \hat{y}_t$. Estimated polynomial trends of orders 4 and 6 and HP filters with $\lambda = 14,000$ and $\lambda = 129,600$ are applied (respectively denoted t^4 , t^6 , HP_1 and HP_2).⁵ To gauge and compare the accuracy of the detrending methods, Table 1 shows the average mean square errors (MSE) between the true cycle of \tilde{y}_t (y_t) and the one obtained after detrending (\hat{y}_t). The average MSEs are computed over the 5,000 replications of each *dgp* and for the four detrending approaches,

$$MSE_{k,d} = \frac{1}{5,000} \sum_{i=1}^{5000} \frac{1}{400} \sum_{t=1}^{400} (y_t^{(k,i)} - \hat{y}_t^{(k,i,d)})^2,$$

where k indicates the *dgp*, d the detrending method used, and i the i -th replication with $1 \leq i \leq 5,000$.

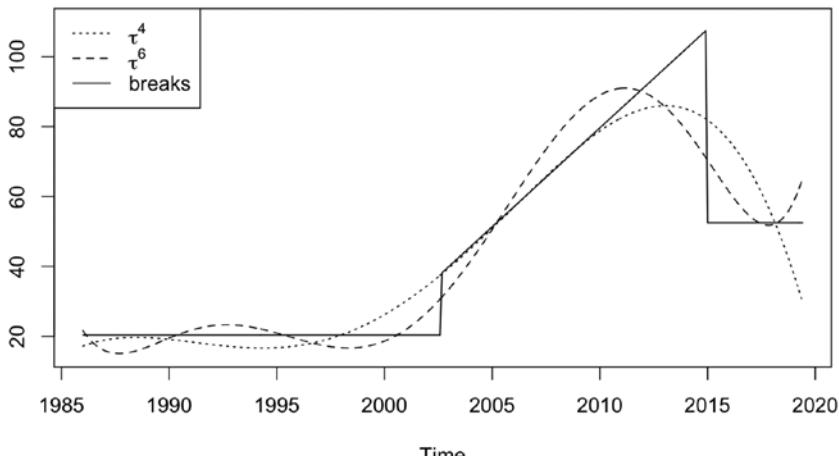


Fig. 2. Trends Estimated on WTI Oil Prices Series.

Table 1. Average Mean Squared Errors Between True Cycles and Detrended Series.

DGP	Detrended with			
	t^4	t^6	HP_1	HP_2
$MAR(0,1) + \text{no trend}$	5.23	7.61	11.44	7.15
$MAR(0,1) + \tau^4$	4.55	6.03	9.62	7.50
$MAR(0,1) + \tau^6$	62.42	6.38	11.35	11.26
$MAR(0,1) + \text{breaks}$	79.02	55.78	31.84	47.65
$MAR(1,1) + \text{no trend}$	22.69	31.05	48.58	30.81
$MAR(1,1) + \tau^4$	42.74	65.18	91.42	57.60
$MAR(1,1) + \tau^6$	85.91	39.57	61.21	43.02
$MAR(1,1) + \text{breaks}$	101.48	86.93	78.36	77.18
$MAR(1,0) + \text{no trend}$	1.20	1.64	2.55	1.58
$MAR(1,0) + \tau^4$	0.96	1.34	2.14	2.70
$MAR(1,0) + \tau^6$	59.24	2.45	4.21	6.73
$MAR(1,0) + \text{breaks}$	76.42	52.19	26.30	44.10

Notes: Are reported the average MSEs over 5,000 trajectories with sample size $T = 400$. HP_1 corresponds to the HP filter with $\lambda = 14,400$ and HP_2 to the HP filter with $\lambda = 129,600$.

The MSEs are minimized when the correct polynomial trend is employed or when the lower order is employed (4 in this case) in the absence of trend in the *dgp*. However, underestimating the order of the polynomial trend leads to significantly larger discrepancies. Distortions between the true cycle and the detrended series are larger for mixed causal-noncausal processes than for purely causal or noncausal processes. Furthermore, in the presence of noncausal dynamics the HP filter with $\lambda = 14,400$ (HP_1) distorts more the series than HP_2 . Hence, we can expect that a low penalizing parameter in the HP filter mostly captures some of the noncausal dynamics. However, HP_1 distorts the least the cycles to which the linear trend with breaks was added. It is the method that best manages to mimic this non-smooth trend due to this flexibility induced by its low penalizing parameter.

3.2. Effects of Detrending on Model Identification

To investigate the impact of detrending on dynamic processes, we perform *MAR* estimations on the raw and detrended series from each *dgp*. The estimation of *MAR* models first consists in estimating the pseudo causal lag order. Since the autocorrelation structure of mixed or purely causal and noncausal processes are identical, we can estimate the order of autocorrelation (p) with information criteria by OLS. Once this order p is estimated, the identification of the lag and lead orders (r and s respectively) is performed by maximum likelihood among all $MAR(r,s)$ models such that $r + s = p$ Lanne and Saikkonen (2011). We do so using the MARX package in R (Hecq et al., 2017).

Table 2 presents the frequencies of identifying wrong models in each of the 12 *dgp*, based on the detrending methods, with a maximum pseudo causal lag order of 4.⁶ Proportions of a wrongly identified the pseudo lag order in the first step of the estimation using BIC are reported ($p \neq 1$ and $p \neq 2$), as well as the proportions of wrongly identified *MAR* models, namely when at least one of the lag or lead order mis-identified. We also report the frequency with which no noncausal dynamics is identified ($s = 0$). For the purely causal processes, we only report in the last column ($s > 0$), i.e., the frequency with which spurious noncausal dynamics is detected.

Let us first focus on the models with noncausal dynamics (the *MAR(0,1)* and *MAR(1,1)* *dgps*) for which we report the frequencies with which we over- or underestimate the pseudo causal lag order in the first step of the estimation. We can see that HP_1 under-performs relative to the other approaches. Indeed, around twice as many lag orders are wrongly estimated in the first step on average, with a maximum of 22.84% for the *MAR(1,1)* processes with breaks in the linear trend. However, this non-smooth trend seems to be difficult to capture by the filters

Table 2. Percentages of Mis-Identified MAR Models.

Detrending method	wrong			wrong			$s > 0$
	$p \neq 1$	MAR	$s = 0$	$p \neq 2$	MAR	$s = 0$	
	MAR(0,1) + no trend			MAR(1,1) + no trend			MAR(1,0) + no trend
raw	5.50	5.50	0.00	4.74	4.74	0.00	0.52
t^4	5.46	5.58	0.14	4.64	4.68	0.04	0.72
t^6	5.70	5.86	0.22	4.94	5.04	0.06	0.88
HP_1	10.78	11.24	0.52	8.18	8.44	0.22	1.86
HP_2	6.84	7.10	0.28	5.56	5.66	0.06	1.02
	MAR(0,1) + τ^4			MAR(1,1) + τ^4			MAR(1,0) + τ^4
raw	12.10	43.84	35.04	9.70	16.38	7.22	32.76
t^4	6.44	6.72	0.28	4.70	4.74	0.04	0.78
t^6	6.76	6.96	0.20	4.86	4.94	0.08	0.76
HP_1	10.28	10.88	0.60	7.64	7.90	0.22	2.52
HP_2	6.24	6.56	0.32	5.36	5.56	0.14	1.92
	MAR(0,1) + τ^6			MAR(1,1) + τ^6			MAR(1,0) + τ^6
raw	13.18	36.14	26.04	9.04	15.56	7.12	35.44
t^4	7.30	7.36	0.08	4.00	4.14	0.04	60.02
t^6	6.54	6.68	0.14	4.48	4.68	0.04	0.92
HP_1	9.40	9.84	0.44	7.90	8.24	0.22	2.86
HP_2	5.94	6.12	0.18	4.86	5.12	0.06	3.46
	MAR(0,1) + breaks			MAR(1,1) + breaks			MAR(1,0) + breaks
raw	4.54	4.92	0.68	6.34	7.68	1.38	94.60
t^4	3.44	4.00	0.60	8.68	8.74	0.22	38.24
t^6	3.40	3.86	0.58	10.70	10.86	0.24	28.24
HP_1	4.00	4.58	0.62	22.84	23.24	0.26	7.54
HP_2	3.38	3.70	0.40	12.70	12.82	0.18	28.92

Notes: During the first stage of the model identification, the maximum number of lags in the pseudo lag model is set to 4. Results are in percentages of the 5,000 trajectories. $T = 400$. HP_1 corresponds to the HP filter with $\lambda = 14,400$ and HP_2 to the HP filter with $\lambda = 129,600$.

considered in this analysis. We can see from the last five rows of [Table 2](#) that detrending this type of processes with breaks – with the four methods employed here – does not improve the correct identification of the orders of the model, and can even make it worse for $MAR(1,1)$. This can be explained by the construction of the trend, mimicking somehow a bubble pattern, with a long and persistent expansion when the linear trend is present and followed by a sudden crash when the series returns to a stationary process. This might be mistaken for noncausal dynamics, ensuring a non-zero lead order identification when the series is not detrended. This claim is supported by the results in the last column, indicating large proportions of wrongly detected noncausal dynamics for each detrending approaches, with 7.54% for HP_1 and more than 28% for the others. For the $dgps$ with other trends (or only intercept) HP_1 wrongly estimates the pseudo causal lag order at most 10.78% of the time. For the three other detrending methods the pseudo lag order is wrongly identified in less than 7.3% of the cases. Note that when the lag order is wrongly identified, it is almost always due to over-identification. The discrepancy between the two HP filters is explained by the low penalizing parameter in HP_1 , allowing the trend to mimic the series too much. By that, some of the dynamics of the MAR process are absorbed by the trend.

It is notably more harmful not to detrend when necessary than the contrary. As can be seen on the upper rows of [Table 2](#), applying polynomial trends or HP_2 do not increase the proportions of wrongly identified models by more than 1.6% compared to estimations on the raw series. However, when the existing trend is ignored, the pseudo lag order is wrongly estimated twice as much on the raw series than for the detrended series, and the MAR models are wrongly identified up to 6 times more than the best performing detrending method. Furthermore, the incorrect identification of the pseudo lag order p accounts for most of the proportion of wrongly identified MAR models. If p is correctly estimated, the model is also correctly identified in more than 99% of the cases. Note that the pseudo causal lag order identified is never zero, meaning that no detrending completely absorbs all dynamics. Besides, in no more than 0.62% the detrending methods killed the noncausal dynamics, as is indicated by the columns $s = 0$.

Let us now consider the last column, displaying the results for purely causal processes. We here investigate whether detrending can create spurious noncausal dynamics ($s > 0$). We find that (ignoring the dgp composed of the trend with breaks) as long as the polynomial trend order is not underestimated, in less than 3.46% of the cases noncausal dynamics was wrongly detected. For the processes with a polynomial trend of order 6, detrending with a polynomial trend of order 4 creates spurious noncausal dynamics in 60.02% of the cases.

Overall, for a dgp with noncausal dynamics, the impact of ignoring a trend is quite significant while detrending when not necessary has negligible effects on model identification. Both the polynomial trends and the HP filter with $\lambda = 129,600$ (HP_2) perform equally well with respect to identifying the correct orders of the model. Choosing a penalizing parameter λ too low alters the dynamics of the process as shown by the results from HP_1 . All of the approaches almost always retain the noncausal dynamics, but rarely create spurious noncausal

dynamics when nonexistent in the *dgp* (except when the polynomial trend order is underestimated). The lead order is not always the correct one but in less than 0.62% for all cases no noncausal dynamics is found. The presented results only report identification of the model lag and lead orders. To have a better understanding of the impact of the detrending methods on the dynamics, focus needs to be put on the impact on the estimated coefficients and parameters of the models identified.

Detailed results on the impact on estimated coefficients are available in Appendix B in the online material. Overall, we find that due to low penalization, HP_1 absorbs too much of the dynamics (mostly the noncausal ones) in the resulting trend. Hence, for monthly data, we advise to use the HP filter with penalization parameter 129,600. It is also rather harmful to underestimate the order of the polynomial trend, which results in a significantly larger lead coefficient. When the fundamental trend consists of breaks (mimicking bubbles), the smooth detrending methods do not succeed in capturing the trend and this translates in much more persistent noncausal dynamics. We also investigate the effect of detrending white noise series; while for the raw series, 6.82% of the models were identified with dynamics, 7.34% were identified with dynamics for the HP filtered series with penalizing parameters 129,600. Hence, we find no significant creation of dynamics when applying the HP filter to a white noise.

4. PREDICTING CRASHES IN OIL PRICES

This section investigates the impact of detrending both for in-sample and real-time analyses. WTI and Brent crude oil monthly prices series are employed, ranging from June 1987 to December 2020. The series consist of end-of-period prices, which enables us to adequately time our analysis based on the outbreak of the COVID-19 pandemic and the appearances of worldwide regulations and lockdowns to counter its spread. Fig. 3 shows that both series are characterized by bubble episodes, which we define in this chapter as rapidly increasing episodes

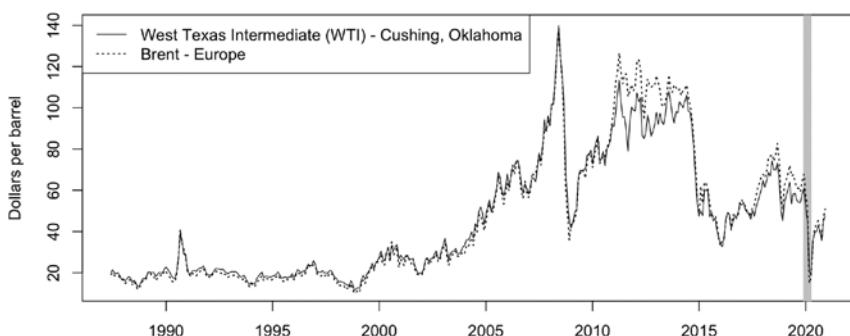


Fig. 3. Monthly Crude Oil Prices.

followed by a sharp decline, the main one being during the financial crises in 2008. The series are also characterized by various sudden crashes. The highlighted gray bar represents the period of interest in this analysis. The earliest point of the period is December 2019; at this point almost no information was available on the coronavirus and no worldwide outbreak had already taken place. Then, we can see that as the outbreak started and regulations were increasingly being imposed worldwide, the price of crude oil significantly dropped. Brent crude oil prices fell from around \$68 at the end of December 2019 to around \$15 by the end of March 2020, point at which most European countries imposed national lock-downs. The restrictions of movement within and between countries thus induced a sharp and sudden decrease in the demand for crude oil.

As shown in Fig. 3, the series are probably nonstationary but considering their growth rate would eliminate the locally explosive episodes that are interesting to exploit. The two series appear almost identical until the 2008 financial crisis, period from which we can observe more apparent discrepancies. The last part of the samples is rather noisy and volatile, and estimating a trend on such a part is not straightforward.⁷ We seek to extract a smooth trend without affecting the dynamics of the series. Based on the findings of Section 3, we consider the deterministic polynomial trends of orders 4 and 6 as well as the HP filter with $\lambda = 129,600$ (denoted t^4 , t^6 and HP respectively). We furthermore employ an economic variable – described in the following section – as another trend to compare economically motivated detrending with mechanical detrendings. The analysis focuses on the probabilities for oil prices to drop and investigates the potential magnitude of such decrease. We first consider an in-sample analysis, that is, the trends and the *MAR* models are estimated over the whole sample, from June 1987 to December 2020. Then, we fix the estimated parameters and use this information to perform one-month ahead density forecasts for the months of January, February, March and April 2020. The in-sample analysis includes as much information as possible and therefore reduces estimation uncertainty. We then compare the in-sample analysis to a real-time forecast exercise. In the real time analysis, we re-estimate the trends and the *MAR* models at each point of the period of interest. That is, we consider an expanding sample and perform one-month ahead density forecasts for points that are out-of-sample.

4.1. Economic Variables to Detrend Series

There is an extensive literature on modeling oil prices using economic variables. As an example, Kilian and Murphy (2014) construct a structural VAR model for the real price of oil, making use of stationary transformations of economic variables, namely the real economic activity index constructed in Kilian (2009) as well as inventories and production of crude oil. In this analysis, we however do not construct a structural model for the price of oil, but instead we investigate ways of detrending prices without altering the inherent dynamics of the process. As such, we suggest employing the US crude oil Strategic Petroleum Reserve (hereafter SPR) levels. These reserves were established primarily to reduce the impact of disruptions in supplies of petroleum stocks Kilian and Zhou (2020a). This variable

therefore incorporates not only expectations regarding the economic activity but also regarding the production of crude oil. US SPR stock is depicted against WTI crude oil prices in Fig. 4.

SPR is significantly less volatile than total crude oil stocks as it is a last resort reserve and is not often made use of as it requires approval of the US President.⁸ This characteristic of the series makes it a good candidate for the smooth trend we intend to extract from oil price series. We hence detrend prices (both nominal and real) by taking the residuals from a standard OLS regression of prices on crude oil SPR levels.⁹

4.2. In-sample Analysis

To save space, Fig. 5 only depicts the detrended Brent series,¹⁰ after the polynomial trends, the HP filter and SPR levels were used to detrend the whole sample. The *SPR*-detrended series consists of the residuals obtained from a standard OLS regression of the prices on the SPR levels. We can see that the *HP*-detrended

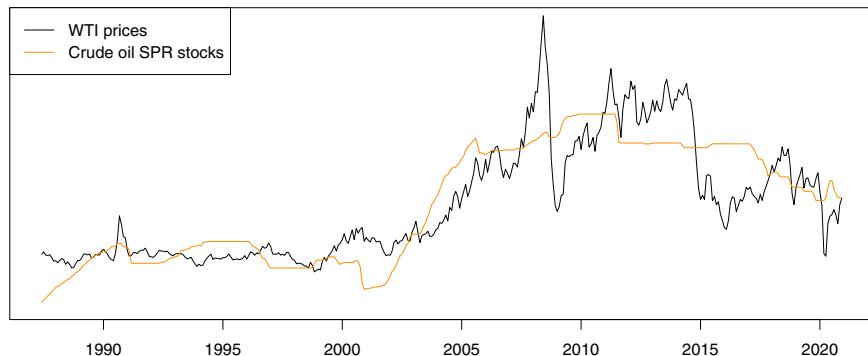


Fig. 4. Raw WTI Prices and US Crude Oil SPR Stocks.

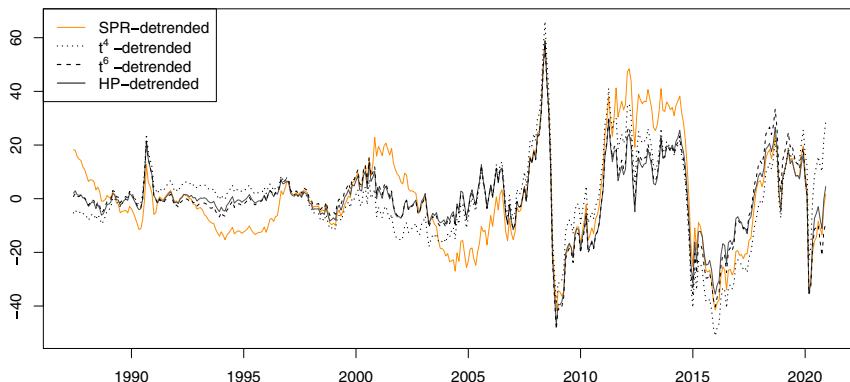


Fig. 5. Detrended Brent Prices.

series (black solid line) and the t^6 -detrended series (dashed line) are very much alike over the majority of the sample. The polynomial trend of order 4 (dotted line) however seems to induce some more variations than the other two mechanical detrending. This is especially visible at the beginning and at the end of the sample, stemming from the lack of flexibility of such trend due to its lower order. This latter detrending method suggests that the end of 2020 is as extreme as the period between 2010 and 2014, during which prices were in fact twice as large. We can see that overall *SPR*-detrended series follows a similar pattern than the others but shows slightly more persistent dynamics. This could stem from the fact that the *SPR* series, while being rather smooth, still displays more dynamics than the 3 other trends considered here. Hence, it could slightly alter the dynamics in the remaining cycle. *SPR*-detrended series has a correlation of 0.84 with both *HP*- and t^6 -detrended series. Note that until the end of the 1980s, there was a persistent increase in *SPR* due to the creation and initial filling of the reserves which started in 1977, explaining the induced downward trend at the beginning of the detrended sample.

We estimate *MAR* models with Student's t -distributed errors and set the maximum pseudo lag length in the first stage on the estimation to 4. All resulting models are *MAR*(1,1) and are reported in Table 3. We report the lag and lead coefficients as well as the degrees of freedom of the distribution and their respective standard errors in parentheses. Models estimated on series that were detrended with a polynomial trend of order 6 and with the HP filter are the most similar, as suggested by Fig. 5. Models estimated after detrending with the polynomial trend of order 4 slightly deviate from the two others and always have a larger lead coefficient, hence indicating more persistence in the explosive episodes. Recall that Section 3 suggests that underestimating the trend order in mixed causal-noncausal models

Table 3. Estimated MAR Models.

Series	<i>MAR</i> (1,1) estimations per detrending method											
	t^4			t^6			<i>HP</i>			<i>SPR</i>		
	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$
WTI	0.25 (0.03)	0.88 (0.01)	2.25 (0.36)	0.29 (0.03)	0.82 (0.02)	1.93 (0.29)	0.29 (0.03)	0.80 (0.02)	1.85 (0.28)	0.24 (0.03)	0.90 (0.01)	2.60 (0.37)
WTI _{real}	0.22 (0.04)	0.88 (0.02)	3.05 (0.49)	0.26 (0.04)	0.83 (0.02)	2.75 (0.50)	0.26 (0.04)	0.81 (0.02)	2.63 (0.49)	0.22 (0.04)	0.91 (0.02)	3.44 (0.69)
Brent	0.31 (0.03)	0.89 (0.01)	1.93 (0.27)	0.31 (0.03)	0.86 (0.02)	1.82 (0.33)	0.31 (0.03)	0.83 (0.02)	1.83 (0.31)	0.31 (0.03)	0.92 (0.01)	2.18 (0.34)
Brent _{real}	0.26 (0.04)	0.90 (0.02)	2.59 (0.55)	0.27 (0.04)	0.86 (0.02)	2.48 (0.56)	0.27 (0.04)	0.84 (0.02)	2.45 (0.57)	0.25 (0.04)	0.92 (0.02)	2.90 (0.53)

Notes: The models are obtained with a maximum pseudo lag order of 4 and for each series the model identified was an *MAR*(1,1). ϕ is the lag coefficient, ψ is the lead coefficient and γ the degrees of freedom of the Student's t distribution. The polynomial trend are trends up to the order indicated and the *HP* filtering is performed with a penalization parameter $\lambda = 129,600$. In parentheses are reported the standard error of the coefficients estimated obtained with the MARX package [Hecq et al. \(2017\)](#).

induces on average an overestimation of the noncausal coefficient. All series are mostly forward looking, as the lead coefficients are at least 0.8 while the lag coefficients are at most 0.31.¹¹ We can see that, as expected, *SPR*-detrended series are slightly more persistent in their noncausal dynamics with a lead coefficient up to 0.1 larger than other detrending methods and also slightly larger degrees of freedom induced by more persistent extreme events. The identification of the dynamics is overall consistent across series and their transformation. Note that adjusting the series for inflation leads to larger estimated degrees of freedom for the Student's *t* distribution but overall to similar dynamics.

Lacking closed-form expressions for the predictive densities, we use the two data-driven approaches mentioned in Section 2. We employ the simulations-based approach of Lanne et al. (2012), which only depends on the model estimated and the last observed point and compare, it approximate the density by use of simulations. We compare this method with the sample-based approach of Gouriéroux and Jasiak (2016), which uses past values in the forecasting step to approximate the conditional density. Table 4 shows the one-month ahead probabilities that the series will decrease (hence be lower than its last observed value) and the probabilities that the series will drop by more than 1 standard deviation (the standard deviations are calculated empirically over the whole sample). Forecasts are performed for January, February, March and April 2020 and results

Table 4. One-Step Ahead Probabilities.

Series	Detrended with	Jan.		Feb.		Mar.		Apr.	
		samp.	sims.	samp.	sims.	samp.	sims.	samp.	sims.
Probability of a decrease									
<i>WTI</i>	<i>t</i> ⁴	0.444	0.423	0.784	0.762	0.722	0.681	0.828	0.825
	<i>t</i> ⁶	0.414	0.437	0.873	0.851	0.726	0.748	0.583	0.705
	<i>HP</i>	0.411	0.422	0.869	0.836	0.701	0.730	0.544	0.675
	<i>SPR</i>	0.432	0.440	0.808	0.768	0.691	0.687	0.738	0.781
	Probability of a decrease > 1 s.d.								
	<i>t</i> ⁴	0.052	0.044	0.016	0.017	0.006	0.008	0.018	0.015
	<i>t</i> ⁶	0.041	0.042	0.007	0.011	0.004	0.005	0.177	0.322
	<i>HP</i>	0.047	0.045	0.007	0.012	0.005	0.006	0.227	0.399
<i>Brent</i>	<i>SPR</i>	0.012	0.010	0.005	0.005	0.002	0.003	0.005	0.014
	Probability of a decrease								
	<i>t</i> ⁴	0.379	0.346	0.806	0.800	0.718	0.696	0.879	0.852
	<i>t</i> ⁶	0.398	0.400	0.886	0.864	0.792	0.768	0.569	0.770
	<i>HP</i>	0.397	0.396	0.880	0.853	0.757	0.745	0.500	0.720
	<i>SPR</i>	0.386	0.390	0.861	0.824	0.786	0.731	0.678	0.860
	Probability of a decrease > 1 s.d.								
	<i>t</i> ⁴	0.044	0.035	0.016	0.016	0.007	0.008	0.071	0.106

Notes: For the simulations-based approach (sims.) the truncation parameter $M = 100$ and 1,000,000 simulations were used. Standard deviations (s.d.) are calculated over the detrended samples and are around 15 for all nominal series.

from the two prediction methods are reported for each of the detrended nominal series. We focus on the nominal series as they are the prices people observe and because the estimated models for real series are noticeably similar.¹² While we advocate the use of predictive densities to get the best picture of potential future prices, we choose 2 arbitrary probabilities to present for a matter of comparison and to save space. Nonetheless, the probabilities for any event can be computed from the methods used here, and they could for instance be employed in the construction of risk measures.

At the end of December 2019 oil prices were around \$60 per barrel, they had been fluctuating around this price over the last three years. All detrending methods yield values for December that are above the 90th percentile of the samples, suggesting high but not extreme levels. At that point in time, no international alerts regarding the risk of a pandemic had been made yet. Probabilities that prices will drop in January are roughly 0.4 for all series and for both forecasting methods. However, probabilities that prices will drop by more than 1 standard deviation are at most 0.052. This confirms that crude oil prices are in a period of volatile and rather high prices, but it does not suggest a bubble behavior with a potential large drop. This can also be seen by the difference between the sample-based and simulations-based predictions. [Hecq and Voisin \(2021\)](#) show that discrepancies between the two approaches mostly arise during extreme episodes. Here, they do not differ by more than 3.3% for the probabilities of a decrease, and by no more than 0.9% for the probabilities of a sharper decrease.

At the end of January 2020, international alerts regarding the spread of the novel coronavirus had been made, which induced an unforeseeable drop in prices. Yet, the t^4 -detrended series only fell by half a standard deviation and the other two by 75% (resp. 80%) of a standard deviation for the Brent (resp. WTI) series. Values remained however above median values. Forecasts based on both methods suggest a continuity in the decrease for February with probabilities ranging from 0.76 to 0.88, yet, they indicate almost zero probability that the drop will be substantial (more than a standard deviation). They hence suggest a return to median values, meaning a return to fundamental prices. Both prediction methods again provide results diverging by no more than 3.3%. By the end of February 2020, mass gatherings started to be forbidden and the first advice for the quarantine of individuals to contain the spread of the virus had been made. The increasing worldwide pressure hence kept pushing prices down. Yet, no decrease in the detrended series was larger than 60% of a standard deviation, which was once again in line with the predictions. The series reached their median levels, forecasts for March suggested that series would remain stable around those values, yet favoring a further slight decrease as prices had been declining for the last three consecutive periods. Probabilities of a sharp drop decreased even more toward zero and both prediction methods yielded again similar probabilities.

In March 2020, the worldwide situation worsened significantly and the World Health Organization declared COVID-19 a global pandemic. Many countries imposed strict movement restrictions within and across borders, and curfews and lock-downs were implemented. This sudden drop in crude oil demand led to a considerable fall in prices, WTI prices fell by 55% and Brent prices by 71%. Values

of the detrended series fell by more than 2 standard deviations and reached the 2nd and 3rd percentile for *HP*-, *SPR*- and t^6 -detrending. This indicates a negatively explosive episode, and therefore a negative bubble below fundamental prices. The t^4 -detrending values correspond to at least the 10th percentile, suggesting a less extreme episode, compared to the previous behavior of the series. Until this point both predicting methods yielded similar probabilities. However, the discrepancy between the probabilities now attains 0.24 difference, where the simulations-based probabilities of a decrease are always larger than the sample-based probabilities. [Hecq and Voisin \(2021\)](#) show that the discrepancies between the sample- and simulations-based approaches widen during explosive episodes. This is why probabilities for t^4 -detrending series are still very similar across the forecasting methods as opposed to the other detrending methods. They also show that the larger the lead coefficient, the more the sample-results tend to yield larger probabilities of a turning point than the ones computed with simulations. This stems from the fact that the series had attained a few times this point before (in 2008 and in 2015) and turned back toward median value. It is therefore, based on the learning mechanism of the sample-based approach, less likely that the series will keep on decreasing. It is important to notice that even though prices dropped significantly, probabilities that they will keep on decreasing are lower than before for *HP*- and t^6 -detrended series as well as for *SPR*-detrended Brent. However, compared to previous forecasts, probabilities now suggest that if the series actually kept on decreasing, it could likely be by more than 1 standard deviation as it has now entered an explosive episode. *SPR*-detrended WTI series has larger probabilities of decrease than for the previous month, however, as can be noticed, the probabilities of the sharper decrease for both *SPR*-detrended series are much closer to 0 than with other detrending. This stems from the larger degrees of freedom as well as larger lead coefficient and slightly lower lag coefficient.

[Fig. 6](#) illustrates the evolution of the predictive densities of the *HP*-detrended Brent series over the time span. On the *x-axis* are the predictions and on the *y-axis* their corresponding probability density. The vertical dashed line corresponds to the last value, that is, in graph (a), the vertical line is the detrended value of Brent prices for December 2019. We can clearly observe the bi-modality of the distribution when the series deviates from median values, as shown for the forecasts of January and April, which exacerbate during the explosive negative episode. The range and shape of the density also explains the discrepancies between probabilities of a decrease and probabilities of a decrease of more than 1 standard deviation.¹³

To illustrate the valuable information provided by the predictive densities of *MAR* models, graph (a) of [Fig. 7](#) depicts the predictive density for April 2020 using a Gaussian *AR(2)* model instead of an *MAR(1,1)* on *HP*-detrended Brent prices. The predictive density is obtained using the closed-form of the conditional normal distribution. We can see that the mode of the density corresponds to a further decrease, but it now lacks the bi-modality and therefore does not suggest a return to central values as does the *MAR* predictive density shown on graph (d) of [Fig. 6](#). As such, once the series enters a locally – here negative – explosive episode, the *AR(2)* only predicts a continuing decrease of the prices. Graph (b) of

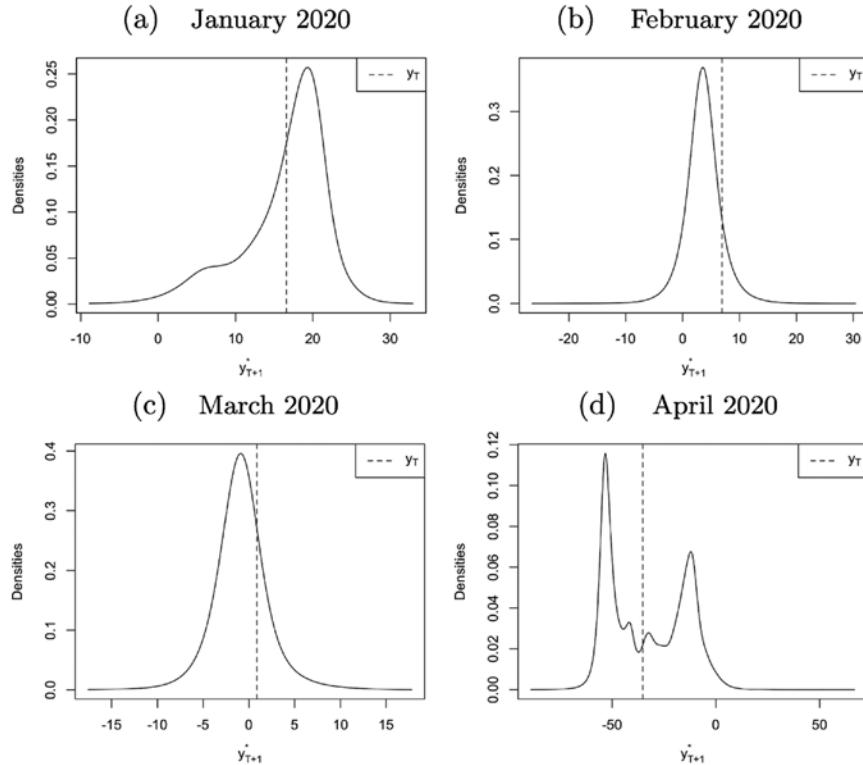


Fig. 6. One-step Ahead Predictive Densities of *HP*-Detrended Brent Prices
Obtained With the Sample-Based Prediction Method.

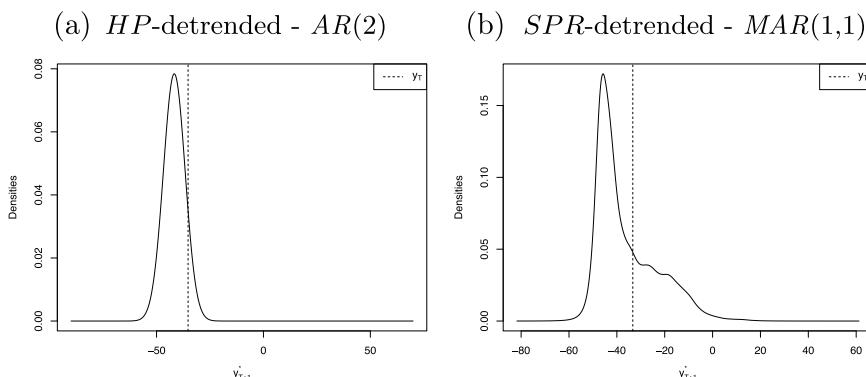


Fig. 7. One-Step Ahead Predictive Densities of *HP*-Detrended Brent Prices
Obtained With the Sample-Based Prediction Method.

[Fig. 7](#) displays the sample-based predictive density of the *SPR*-detrended Brent series. We can see that the larger lead coefficient implies a lower rate of decrease (this can be seen as the distance between the two modes), but it indicates larger probabilities of a further decrease as large lead coefficients imply longer lasting explosive episodes. This is why the right mode, which corresponds to a return to central values has a much lower weight on the density.

The *MAR* models employed here are univariate, hence no exogenous information is incorporated, as opposed to *MARX* models (see [Hecq et al., 2020, 2022](#)). Disregarding exogenous variables facilitates forecasting but can sometimes lead to consequential lack of information. For instance, it is expected that crude oil prices should be lower-bounded as they cannot decrease indefinitely and become increasingly negative. Simulations-based probabilities cannot take that into account as they are only based on the model estimated. Sample-based probabilities however, since prices have never become negative (or at least not long enough to be visible on monthly series), will tend to limit the probabilities that it will happen in the future, even without incorporating additional information within the model, based on its learning mechanism.

Overall, *HP*- and t^6 -detrending provide similar results both for estimation and predictions. *SPR*-detrending, as mentioned earlier yields slightly different dynamics which might stem from the dynamics that are inherent to the stock variable itself. Detrending with t^4 yields slightly different results for the estimation but which in turn yields quite different results for predictions. We saw in [Fig. 5](#) that detrending with a trend polynomial of order 4 induced different dynamics in the remaining cycle than the other mechanical detrending. This also corroborates the results found in Section 3 about the risks of underestimating the order of a polynomial trend on the dynamics of the series. We can also see in [Fig. 5](#) that the main differences between all detrending methods appear at the end of the sample. *HP* and *SPR* detrending are almost identical while t^6 provides slightly lower values. On the other hand, t^4 -detrending yields significantly larger value than the others for the end of the sample.

4.3. Real-time Analysis

To illustrate the difficulties and the limitations of detrending and forecasting in real time, we compare the results obtained in real time to the ones obtained in-sample for Brent prices with t^4 , t^6 and *HP* detrending. We did not include *SPR* detrending in this Section as we are interested in detrending methods that are affected by sample expansion and while with *SPR* detrending we still need to re-estimate the model at each point, the trend itself does not change. [Table 5](#) shows the estimated *MAR* models for the expanding samples after each detrending. We can see that the expansion of the sample, even with the inclusion of the large drop of March 2020 did not affect the identification of the model nor the dynamics. Lead and lag coefficients vary by no more than 0.03. The estimated degrees of freedom of the Student's *t* distribution are rather stable until the data point of March is included, which induced decrease between 0.07 and 0.1 for all series, getting therefore closer to the parameter estimated ex-post. This stability in the

Table 5. Estimated MAR Models on Different Brent Prices Samples.

MAR(1,1) estimations per detrending method									
Sample	t^4			t^6			HP		
	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$	ϕ	ψ	$t(\gamma)$
In-sample	0.31 (0.03)	0.89 (0.01)	1.93 (0.27)	0.31 (0.03)	0.86 (0.02)	1.82 (0.33)	0.31 (0.03)	0.83 (0.02)	1.83 (0.31)
→ Dec	0.30 (0.03)	0.89 (0.01)	2.06 (0.27)	0.29 (0.03)	0.86 (0.02)	1.98 (0.33)	0.29 (0.03)	0.84 (0.02)	1.97 (0.31)
→ Jan	0.30 (0.03)	0.89 (0.01)	2.05 (0.31)	0.29 (0.03)	0.86 (0.02)	1.97 (0.33)	0.28 (0.03)	0.84 (0.02)	1.97 (0.31)
→ Feb	0.30 (0.03)	0.89 (0.01)	2.07 (0.31)	0.31 (0.03)	0.85 (0.02)	1.97 (0.32)	0.30 (0.03)	0.83 (0.02)	1.97 (0.31)
→ Mar	0.30 (0.03)	0.89 (0.02)	1.97 (0.36)	0.31 (0.03)	0.85 (0.02)	1.89 (0.36)	0.30 (0.03)	0.83 (0.02)	1.90 (0.33)

Note: See Table 3.

estimation of the models suggest that probabilities should not significantly differ either.

To investigate the sensitivity of the detrending methods to the addition of new data points, Fig. 8 shows how the detrended series vary based on the stopping point of the sample. The dashed line corresponds to the ex-post detrended series, hence when all data points until December 2020 are included. Then, the expanding samples are depicted from the light blue curve (sample stopping in December 2019) to the black curve (sample until March 2020). While detrending with t^4 induced the most spurious dynamics over the sample, it seems, as well as the HP filter, to be less affected by the addition of the new points than the t^6 -detrending. In graph (a), we can see that the 4 detrended series are almost identical, even once the point for March is added. In graph (c), corresponding to the HP-detrended series, we can see that the 3 first detrended series are almost identical but that the inclusion of March creates a slight shift in the detrended series. In this case also, the inclusion of even later points will induce further shifts of the estimated trend. However, for the polynomial trend of order 6, as depicted in graph (b), we can see that the inclusion of each point creates a noticeable shift in the estimated trend. From this, we expect the t^6 -detrended series to be the ones for which the probabilities differ the most from the in-sample probabilities. Indeed, even if the estimated model is almost identical, the substantial discrepancies between the real-time and ex-post detrended series may impact probabilities, especially during (mildly) explosive episodes.

Fig. 9 depicts the evolution of the one-month ahead probabilities with expanding window. In black are the in-sample probabilities and in gray the real-time probabilities. For the real-time analysis, the trend and the model is re-estimated at each point. The full lines are the probabilities of a decrease and the dashed lines are the probabilities of a decrease of more than 1 standard deviation. Graph (a) (resp. (b)) represents the sample-based (resp. simulations-based) probabilities.

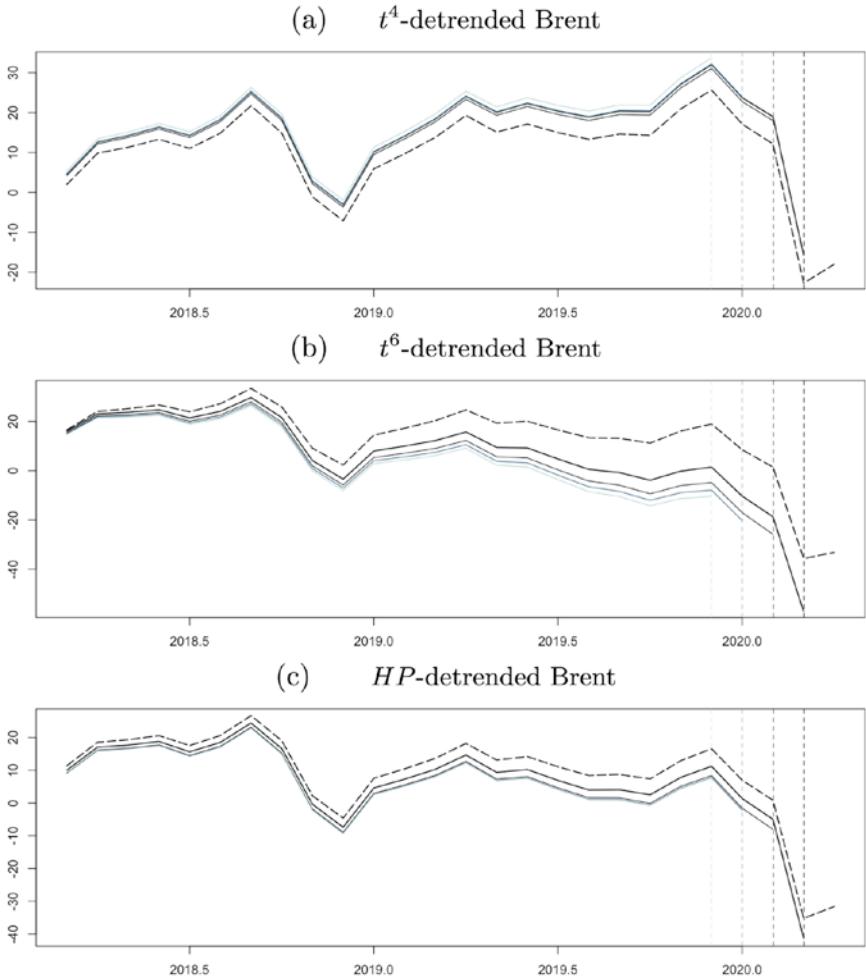


Fig. 8. In-sample (Dashed Curve) Versus Real-Time Detrending of Brent Prices.

As expected, the simulations-based probabilities are the least affected by the re-estimation of the model at each point, since we did not observe significant alteration in the estimations. However, as shown in Fig. 8, it is indeed the t^6 -detrending that is the most sensitive to the expansion of the sample. Furthermore, we can see that mostly the probabilities of a decrease are affected, as the probabilities of a drop of more than 1 standard deviation are not significantly deviating from the in-sample probabilities. Overall, this indicates that real-time forecasting would have indicated on average lower probabilities of a decrease, at each point and for both approaches. Yet, it would have indicated equal, if not slightly higher, probabilities for the larger drop. Hence, probabilities of more extreme events, namely

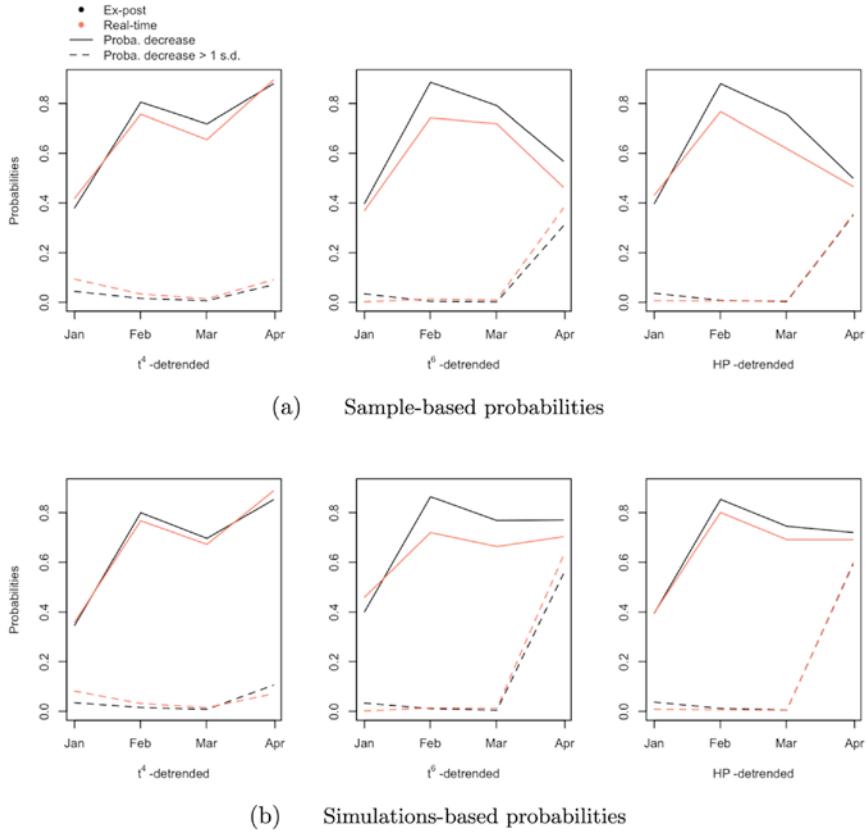


Fig. 9. Evolution of In-Sample (Black Solid and Dashed Lines) and Real-Time (Grey Solid and Dashed Lines) Probabilities Over Time.

the tails of the predictive densities, seem to be the least affected by alteration of the trend.

Overall, it seems that the *HP*-filter is the least sensitive to the change of sample size within this analysis. Results with t^4 -detrending also emphasize the risks of underestimating the order of the trend. Moreover, while the *HP*-filter and the polynomial trend of order 6 perform similarly in this analysis, assuming the order of a trend polynomial requires additional understanding regarding the deviations of the series from its fundamental trend. Deterministic trends appear also to be more sensitive to the addition of points in a real-time exercise than the *HP* filter. Furthermore, while simulations-based probabilities are not characterized by the learning mechanism of the sample-based approach, they are less affected by expanding samples, as long as the model estimated remains consistent. However, as mentioned earlier, with a model that lacks exogenous information, the sample-based approach relying more on past behavior can potentially offset the shortcomings.

5. CONCLUSION

This chapter aims at shedding light upon how transforming or detrending a series can substantially impact predictions of mixed causal-noncausal models. Assuming a polynomial trend of order 4 for WTI and Brent series probably alters the dynamics in the remaining cycle. The HP filter (with penalizing parameter $\lambda = 129,600$) does not require any further assumptions with respect to the trend and can therefore be an adequate filter in cases where the trend is unknown. Knowing the actual trend or using exogenous variables for it is also not straightforward. We use US crude oil strategic petroleum reserves (SPR) to detrend oil price series to illustrate this option. We show that by detrending with SPR we obtain similar results to the *HP* and polynomial trend of order 6 detrending. However, detrending with a variable that has seasonality or dynamics will alter the dynamics left in the cycle. Overall, caution is needed when detrending a series, and some filtering such as polynomial trends may require additional understanding regarding the deviations of the series from its fundamental trend. Nonetheless, once the series is detrended, resulting in a stationary series, using *MAR* models is a straightforward approach to model nonlinear time series. They capture the locally explosive episodes observed in oil prices in a strictly stationary setting. While the bimodality of the predictive density would not be detected with standard Gaussian *ARMA* models, it could be detected with complex nonlinear models, but such model lacks the parsimonious characteristic of *MAR* models. The data-driven prediction methods may lack theoretical grounds but provide valuable information based on the estimated model and on past behaviors of the series in a parsimonious way. This chapter focuses on one-step ahead predictions of decrease in crude oil prices during the first wave of the COVID-19 pandemic.

NOTES

1. An alternative strategy to ours is to consider autoregressive processes with breaks in coefficients. Indeed, autoregressive processes with successively unit roots, explosive and stable stationary episodes are also able to capture locally explosive episodes. See among many others [Phillips et al. \(2011\)](#) and the survey papers by [Homm and Breitung \(2012\)](#) or [Bertelsen \(2019\)](#). Yet, for the purpose of forecasting, we argue for the choice of a model with constant coefficients as more adequate.
2. A description of how the methods are used in this analysis can be found in Appendix A in the online material.
3. The locally explosive features of the data make unit root tests doubtful.
4. In our case, the proposed boosting algorithm absorbs too much dynamics and captures the bubble in the trend component.
5. The estimated trend polynomials are denoted t^4 and t^6 to distinguish them from the trend polynomials part of the *dgps* τ^4 and τ^6 .
6. Results when the pseudo lag order is fixed to the correct one ($p = 1$ or $p = 2$ for mixed models) are available upon request.
7. A Figure of the prices deflated with the consumer price index can be found in Appendix C in the online material.
8. Limited release can be allowed by the Secretary of Energy for crude oil loans to non-governmental entities, as is described by the Energy department of the United States.
9. As shown in Fig. 3, WTI and Brent price series seem to follow a similar trend; we therefore also employ US SPR stocks to detrend Brent prices. We find statistical support

for cointegration between prices (both nominal and real prices of WTI and Brent) and US SPR levels. Hence the remaining cycles are, as intended, stationary.

10. Data and results for the prices-adjusted series that are not presented here are available upon request.

11. The bi-modality of the coefficient distribution in the estimation can lead, in the optimization of the likelihood function, to a local maximum (Bec et al., 2020). This phenomenon is subject to initial values and can induce a switch between the lag and lead coefficients. This was however thoroughly checked in the analysis.

12. Results for price-adjusted series can be found in Appendix C in the online material, probabilities slightly vary however the patterns described in the results for nominal series are identical.

13. Results for all other series, available upon request, follow a similar pattern.

ACKNOWLEDGMENTS

The authors would like to thank Francesco Giancaterini, an anonymous referee and the editors for valuable comments and suggestions. All remaining errors are ours.

REFERENCES

- Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics*, 25(4), 539–573.
- Alquist, R., Kilian, L., & Vigfusson, R. J. (2013). Forecasting the price of oil. In G. Elliott & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 2, pp. 427–507). Elsevier.
- Backus, D. K., & Kehoe, P. J. (1992). International evidence on the historical properties of business cycles. *The American Economic Review*, 82(4), 864–888.
- Baumeister, C., & Kilian, L. (2016). Forty years of oil price fluctuations: Why the price of oil may still surprise us. *Journal of Economic Perspectives*, 30(1), 139–160.
- Bec, F., Nielsen, H. B., & Saïdi, S. (2020). Mixed causal–noncausal autoregressions: Bimodality issues in estimation and unit root testing 1. *Oxford Bulletin of Economics and Statistics*, 82(6), 1413–1428.
- Bertelsen, K. P. (2019). *Comparing tests for identification of bubbles*. Department of Economics and Business Economics, Aarhus University.
- Breidt, F. J., Davis, R. A., Li, K.-S., & Rosenblatt, M. (1991). Maximum likelihood estimation for non-causal autoregressive processes. *Journal of Multivariate Analysis*, 36(2), 175–198.
- Brooks, C., Prokopczuk, M., & Wu, Y. (2015). Booms and busts in commodity markets: Bubbles or fundamentals? *Journal of Futures Markets*, 35(10), 916–938.
- Campbell, J. Y., & Shiller, R. J. (1987). Cointegration and tests of present value models. *Journal of Political Economy*, 95(5), 1062–1088.
- Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, 41(3), 475–512.
- Cavaliere, G., Nielsen, H. B., & Rahbek, A. (2018). Bootstrapping noncausal autoregressions: with applications to explosive bubble modeling. *Journal of Business & Economic Statistics*, 38(1), 55–67.
- Cubadda, G., Hecq, A., & Telg, S. (2019). Detecting co-movements in non-causal time series. *Oxford Bulletin of Economics and Statistics*, 81(3), 697–715.
- Diba, B. T., & Grossman, H. I. (1988). Explosive rational bubbles in stock prices? *The American Economic Review*, 78(3), 520–530.
- Fries, S. (2021). Conditional moments of noncausal alpha-stable processes and the prediction of bubble crash odds. *Journal of Business & Economic Statistics*, 40(4), 1596–1616.
- Fries, S., & Zakoïan, J.-M. (2019). Mixed causal–noncausal ar processes and the modelling of explosive bubbles. *Econometric Theory*, 35(6), 1234–1270.

- Gouriéroux, C., & Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, 37(3), 405–430.
- Gouriéroux, C., & Zakoïan, J.-M. (2013). Explosive bubble modelling by noncausal process. *CREST*. Centre de Recherche en Economie et Statistique.
- Gouriéroux, C., & Zakoïan, J.-M. (2015). On uniqueness of moving average representations of heavy-tailed stationary processes. *Journal of Time Series Analysis*, 36(6), 876–887.
- Gouriéroux, C., & Zakoïan, J.-M. (2017). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3), 737–756.
- Gourieroux, C., Jasiak, J., & Monfort, A. (2020). Stationary bubble equilibria in rational expectation models. *Journal of Econometrics*, 218(2), 714–735.
- Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, 100(5), 831–843.
- Hecq, A., & Voisin, E. (2021). Forecasting bubbles with mixed causal-noncausal autoregressive models. *Econometrics and Statistics*, 20, 29–45.
- Hecq, A., Lieb, L., & Telg, S. (2017). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The MARX package. Available at SSRN 3015797.
- Hecq, A., Issler, J. V., & Telg, S. (2020). Mixed causal–noncausal autoregressions with exogenous regressors. *Journal of Applied Econometrics*, 35(3), 328–343.
- Hecq, A., Issler, J. V., & Voisin, E. (2022). A short term credibility index for central banks under inflation targeting: an application to Brazil. arXiv. <https://doi.org/10.48550/ARXIV.2205.00924>
- Hencic, A., & Gouriéroux, C. (2015). Noncausal autoregressive model in application to bitcoin/USD exchange rates. In V.-N. Huynh, V. Kreinovich, S. Sriboonchitta, & K. Suriya (Eds.), *Econometrics of risk* (pp. 17–40). Springer.
- Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: An empirical investigation. *Journal of Money, Credit, and Banking*, 29(1), 1–16.
- Homm, U., & Breitung, J. (2012). Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics*, 10(1), 198–231.
- Karapanagiotidis, P. (2014). *Dynamic modeling of commodity futures prices* [MPRA Paper 56805]. University Library of Munich, Germany.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3), 1053–69.
- Kilian, L., & Murphy, D. P. (2014). The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics*, 29(3), 454–478.
- Kilian, L., & Zhou, X. (2020a). Does drawing down the US Strategic Petroleum Reserve help stabilize oil prices? *Journal of Applied Econometrics*, 35(6), 673–691.
- Kilian, L., & Zhou, X. (2020b). *The econometrics of oil market VAR models* [CESifo Working Paper Series, (8153)].
- Lanne, M., & Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, 3(3). <https://doi.org/10.2202/1941-1928.1080>
- Lanne, M., Luoto, J., & Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, 28(3), 623–631.
- Lof, M., & Nyberg, H. (2017). Noncausality and the commodity currency hypothesis. *Energy Economics*, 65, 424–433.
- Phillips, P. C., & Shi, Z. (2019). Boosting the Hodrick-Prescott filter [Cowles Foundation Discussion Paper No. 2192]. <http://dx.doi.org/10.2139/ssrn.3447546>
- Phillips, P. C., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1), 201–226.
- Pindyck, R. S. (1993). The present value model of rational commodity pricing. *The Economic Journal*, 103, 511–530.
- Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of economics and statistics*, 84(2), 371–376.

This page intentionally left blank

CHAPTER 9

DEPTH-WEIGHTED FORECAST COMBINATION: APPLICATION TO COVID-19 CASES

Yoonseok Lee^a and Donggyu Sul^b

^a*Syracuse University, Syracuse, New York, United States*

^b*University of Texas at Dallas, Richardson, Texas, United States*

ABSTRACT

The authors develop a novel forecast combination approach based on the order statistics of individual predictability from panel data forecasts. To this end, the authors define the notion of forecast depth, which provides a ranking among different forecasts based on their normalized forecast errors during the training period. The forecast combination is in the form of a depth-weighted trimmed mean. The authors derive the limiting distribution of the depth-weighted forecast combination, based on which the authors can readily construct prediction intervals. Using this novel forecast combination, the authors predict the national level of new COVID-19 cases in the United States and compare it with other approaches including the ensemble forecast from the Centers for Disease Control and Prevention (CDC). The authors find that the depth-weighted forecast combination yields more accurate and robust predictions compared with other popular forecast combinations and reports much narrower prediction intervals.

Keywords: Forecast depth; forecast combination; panel forecast; prediction interval; robust forecast; COVID-19

JEL Classifications: C32; C33; C53

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 235–260

Copyright © 2023 by Yoonseok Lee and Donggyu Sul

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045B011

1. INTRODUCTION

Since the seminal work by [Bates and Granger \(1969\)](#), forecast combinations have been successfully used in many empirical studies when multiple forecasts of the same variable are available. It is also known that combined forecasts often produce better forecasts than the *ex ante* best single forecasting model. If we know the individual forecasting models and their information sets (i.e., the predictors) that produce multiple forecast values, we take the model averaging approach and obtain the optimal weights by minimizing the forecast mean squared error loss. When the number of candidate models and the information sets are large, however, it can be very costly to pool the information sets particularly in real-time forecast (e.g., [Diebold & Pauly, 1990](#)). In practice, it is often the case that the individual forecasting models are not fully known and only the forecast reports are available from the forecasting agencies. In such cases, we can apply ensemble methods such as bagging and boosting to combine forecasts. Such approaches, however, typically require a long training period but a small number of forecasting models, so that the weights can be properly estimated. For this reason, it is common to use the equally-weighted average of forecasts or the weighted average based on the inverted forecast mean squared errors (e.g., Stock & Watson, 2001). The equal weight approach is the most popular because it is simple but typically outperforms the estimated optimal weights, which is often called the “forecast combination puzzle.” See [Clemen \(1989\)](#), Stock and Watson (2001, 2006), and Timmermann (2006), for instance, for survey of this literature.

In this chapter, we propose a forecast combination based on the order statistics of individual predictability when many forecasts are available. We assume (cross-sectional) data rich environment of the forecasts but we do not know of each forecasting model nor its information set. In other words, we only have a set of forecast values reported from many different forecasting agencies, where we do not know their forecasting models, data sets, or predictors. The weights can be obtained using the cross-sectional information and hence we do not need a long training period to estimate the weights. For this reason, we can apply this novel method to a very short panel forecast data set, which is not typically the case for the inverted forecast mean squared error combination. This feature is very useful in practice especially when the time series of interest highly fluctuates and hence the forecasting accuracy from each forecasting agency is not consistent over time.

More precisely, we develop the *forecast depth*, by modifying the notion of data depth (e.g., [Lee & Sul, 2022a](#); [Zuo & Serfling, 2000](#)) in the context of forecast combination, which measures the nearness of each vector of forecasts toward the vector of observed values over the training period. The weights for forecast combination are proportional to the forecast depth. Rather than deriving the optimal forecasting weights, we seek for a robust forecast combination against erroneous forecasts (i.e., outliers). Note that the forecast depth naturally provides a ranking among the forecasting agencies. To design a more robust forecast combination toward extremely poor forecasts during the training period, we trim out the forecasts by the agencies who belong to the lowest ranking group. In this sense, this novel weighting scheme shares the idea of the rank-based approach

(e.g., [Aiolfi & Timmermann, 2006](#)) and the idea of trimming (e.g., [Granger & Jeon, 2004](#)) in forecast combinations. The depth-weighted forecast combination is in the form of the L-statistic, and thus it is more robust toward very bad forecasts than the popular combination methods, including the equal weight combination and the inversed forecast mean squared error combination.

The main contribution of this chapter can be summarized in three folds. First, we develop the forecast depth, based on which we readily rank the forecasting performance over multiple periods and construct a robust forecast combination in the form of the depth-weighted trimmed mean. This approach only requires the forecast values and does not need to know each forecasting model. The number of forecasts can be large and the training period can be very short. Second, we derive the limiting distribution of the depth-weighted forecast combination with trimming, which can be used to construct a prediction interval of forecast combination without relying on subsampling or bootstrap. Both the weight and the trimming scheme depend on the forecast depths in the sample, and hence they are treated stochastic in deriving the limiting distribution. Since the proposed form of forecast combination encompasses popular ones as its special cases, such as the equally-weighted forecast combination, the trimmed forecast combination, and the median forecast, this result also provides prediction intervals of those forecast combinations as well. Third, we apply the depth-weighted forecast combination to predict the national level of new COVID-19 cases in the United States. The new forecast combination yields lower forecast mean squared errors than the ensemble forecast reported by the CDC as well as other popular forecast combination approaches including the equally-weighted combination and inversed forecast mean squared error combination. It provides forecast that is very robust to erroneous or extremely bad predictions in the pool. In addition, it reports much narrower prediction intervals.

It is worth noting that the depth-weighted forecast combination uses cross-sectional distribution information in prediction, which can be time-varying. In the recent works, Joon Y. Park has developed novel approaches to estimate distributional dynamics and unknown trends in time series distribution. For example, see [Chang et al. \(2016\)](#), [Hu et al. \(2017\)](#), and [Chang et al. \(2020\)](#). Once the unknown stochastic trend in the distribution of many forecasts is estimated, this information can be used to model the dynamics of forecast depth and to reinforce the forecast combination especially in long horizon forecasting when a long panel forecast data set is available. We do not consider this method here, but it will be a very promising and interesting topic for future research.

The rest of the chapter is organized as follows. Section 2 defines the forecast depth and develops the depth-weighted forecast combination. Section 3 compares the forecast depth with the projection depth and the inversed forecast mean squared errors. Section 4 derives the limiting distribution of the depth-weighted forecast combination and provides the prediction interval. Using the panel of forecasts on new COVID-19 cases in the United States, Section 5 examines the performance of the proposed forecast combination and compares it with other popular forecast combination approaches. Section 6 concludes with some remarks. The proof of the main theorem is in the Appendix.

2. FORECAST DEPTH AND FORECAST COMBINATION

We denote y_t^0 as the observed true value of interest at time t and $\{y_{i,t}\}$ be the multiple competing forecasts for y_t^0 from different forecasting agencies $i = 1, \dots, n$ using different forecasting models or predictors. For the simplicity, we let y_t^0 and $y_{i,t}$ be scalar values, though the idea below can be extended to the vector case when we jointly predict multiple variables.

We suppose there are n different h -step ahead forecasts $y_{i,t+h}$ for y_{t+h}^0 . We do not know how each agency produces her own forecast value $y_{i,t+h}$; in other words, we do not know her forecasting model nor predictors. We consider the h -step ahead forecast combination in the form of a linear combination of the forecasts:

$$\hat{y}_{t+h} = \sum_{i=1}^n \pi_{i,t} y_{i,t+h} \quad (1)$$

for some potentially time-varying weights $\pi_{i,t}$, which are obtained using the information available at time t . Popular choices of the weights $\pi_{i,t}$ include the equal weight (i.e., $\pi_{i,t} = 1/n$) and the inverted forecast mean squared error (e.g., $\pi_{i,t} = \widehat{\text{mse}}_{i,t}^{-1} / \sum_{j=1}^n \widehat{\text{mse}}_{j,t}^{-1}$, where $\widehat{\text{mse}}_{i,t}$ is the sample forecast mean squared error of agency i using information at time t). In this chapter, we define the weights $\pi_{i,t}$ based on a novel idea, the forecast depth.

To define the forecast depth, we first let the $k \times 1$ vector of forecasts during the training period from $t-k+1$ to t ,¹

$$Y_{i,t} = (y_{i,t-k+1}, \dots, y_{i,t-1}, y_{i,t})' \text{ for } i=1, \dots, n,$$

which is the most recent k forecasts at time t , and the $k \times 1$ vector of the observed true values during this period,

$$Y_t^0 = (y_{t-k+1}^0, \dots, y_{t-1}^0, y_t^0)'.$$

We denote the $k \times 1$ forecast error vector of the agency i during the training period as

$$e_{i,t} = Y_{i,t} - Y_t^0 \text{ for } i=1, \dots, n.$$

We let $m = (m_1, \dots, m_k)'$ be a non-random $k \times 1$ discount vector with $\sum_{j=1}^k m_j = 1$, whose examples are to be discussed later. Given m , we define the normalized forecast error of $Y_{i,t}$ (or the forecast outlyingness) as

$$\mathcal{O}_{i,t} = \frac{|m'e_{i,t}|}{s_t} = \frac{|m'(Y_{i,t} - Y_t^0)|}{s_t}, \quad (2)$$

where $s_t \in (0, \infty)$ is some dispersion measure of $m'e_{i,t}$ that is affine invariant and measurable to the information set at t , say $\mathcal{I}_t = \sigma(\bigcup_{r=1}^n \{e_{i,r}\}_{r \leq t})$.² For instance, if

$\{e_{i,t}\}$ is a random sample from a common distribution across i , we can consider the conditional root mean squared error (RMSE),

$$s_t = \left(\mathbb{E} \left[(m' e_{i,t})^2 \mid \mathcal{I}_t \right] \right)^{1/2} \quad (3)$$

when the conditional mean of $m' e_{i,t}$ is assumed to be zero; or the conditional median absolute deviation (MAD),

$$s_t = \inf \left\{ v : \mathbb{P} \left(|m' e_{i,t}| \leq v \mid \mathcal{I}_t \right) \geq 1/2 \right\} \quad (4)$$

when the conditional median of $m' e_{i,t}$ is assumed to be zero.³ We define the *forecast depth* of agency i at time t as

$$\mathcal{D}_{i,t} = \frac{1}{1 + \mathcal{O}_{i,t}}, \quad (5)$$

where it is obtained using the data set in the k training period from $t-k+1$ to t and hence $\mathcal{D}_{i,t} \in \mathcal{I}_t$. By construction, the forecast depth $\mathcal{D}_{i,t}$ takes values between zero and one; it is one when the agency i yields perfect forecasts during the training period and hence $e_{i,t} = 0$.

Several examples of the discount vector m can be considered. In the forecast error vector

$$e_{i,t} = Y_{i,t} - Y_t^0 = \left((y_{i,t-k+1} - y_{t-k+1}^0), \dots, (y_{i,t-1} - y_{t-1}^0), (y_{i,t} - y_t^0) \right)',$$

since we typically consider that the forecast performance for the most recent observations are more important than the distant ones, we can let the j th element of $m = (m_1, \dots, m_k)'$ as

$$m_j = \frac{K(j/k)}{\sum_{\ell=1}^k K(\ell/k)} \text{ for } j=1, \dots, k, \quad (6)$$

where $K(\cdot)$ is some non-decreasing one-side kernel function. Examples include:

- the polynomial kernel, $K(j/k) = (j/k)^q$ for some $q \geq 1$;
- the discount factor approach by [Bates and Granger \(1969\)](#), $K(j/k) = \gamma^{k-j}$ for some $\gamma < 1$;
- the Box–Cox transform weights by [Diebold and Pauly \(1990\)](#).

If we treat all the forecast errors during the training period equally important, then we simply set $K(j/k) = 1$ or

$$m_j = \frac{1}{k} \text{ for all } j = 1, \dots, k.$$

At given t , we can also define m_j as the normalized inverse of cross-sectional MSE, by letting $K(j/k) = \left[n^{-1} \sum_{i=1}^n (y_{i,t-j+1} - y_{t-j+1}^0)^2 \right]^{-1}$. Note that all such choices of m do not require a balanced panel data structure in $Y_{i,t}$; that is all forecasts are not necessarily available over the given training period. Therefore, for each i , we can even define a $k_i \times 1$ heterogeneous discount vector $m_i = (m_{i1}, \dots, m_{ik_i})$ by letting $m_{ij} = K(j/k_i) / \sum_{\ell=1}^{k_i} K(\ell/k_i)$ for $j = 1, \dots, k_i$ with $k_i \leq k$.

We propose to define the weights $\pi_{i,t}$ for the forecast combination in (1) that is proportional to the individual forecast depth (5). The forecast depth $\mathcal{D}_{i,t}$ naturally provides a ranking of predictability through the entire training period among the (unrevealed) forecasting models from different agencies, since the better performing ones have higher levels of forecast depth. Therefore, we can also use the forecast depth as a tool to detect the under-performed forecasting agencies. To design a more robust forecast combination toward extremely poor forecasts, we trim out the forecasts by the agencies who belong to the lowest ranking group. More precisely, we set some trimming parameter $\tau \in (0,1)$ and let

$$\pi_{i,t} = \frac{1 \{ \hat{\mathcal{D}}_{i,t} \geq \tau \} W(\hat{\mathcal{D}}_{i,t})}{\sum_{j=1}^n 1 \{ \hat{\mathcal{D}}_{j,t} \geq \tau \} W(\hat{\mathcal{D}}_{j,t})}, \quad (7)$$

where $1\{\cdot\}$ is the binary indicator⁴ and $W(\cdot)$ is some scalar weight function $W: [0,1] \rightarrow [0,1]$. For most of the cases, we can simply choose $W(d) = d$. $\hat{\mathcal{D}}_{i,t}$ is the forecast depth estimator that is defined as

$$\hat{\mathcal{D}}_{i,t} = \frac{1}{1 + \hat{\mathcal{O}}_{i,t}} \text{ with } \hat{\mathcal{O}}_{i,t} = \frac{|m'_i(Y_{i,t} - Y_t^0)|}{\hat{s}_t} \quad (8)$$

for some consistent estimator \hat{s}_t . For each of the aforementioned examples in (3) and (4), we can use

$$\hat{s}_t = \left(\frac{1}{n} \sum_{i=1}^n m'(Y_{i,t} - Y_t^0)(Y_{i,t} - Y_t^0)' m \right)^{1/2} \quad (9)$$

and

$$\hat{s}_t = \text{med}_{1 \leq i \leq n} |m'(Y_{i,t} - Y_t^0)|, \quad (10)$$

respectively. Since we define the training period over a rolling window of the most recent k periods, the weight in (7) is naturally time-varying.

Using the weight (7) in the forecast combination (1), we define *the depth-weighted forecast combination* as a form of the trimmed depth-weighted mean given as

$$\hat{y}_{t+h} = \sum_{i=1}^n \pi_{i,t} y_{i,t+h} = \frac{\sum_{i=1}^n 1\{\hat{D}_{i,t} \geq \tau\} W(\hat{D}_{i,t}) y_{i,t+h}}{\sum_{i=1}^n 1\{\hat{D}_{i,t} \geq \tau\} W(\hat{D}_{i,t})}, \quad (11)$$

where $\{y_{i,t+h}\}_{i=1}^n$ are the individual h -step ahead forecasts for y_t^0 at the current time t .

The forecast combination in (11) assigns the weight on $y_{i,t+h}$ based on its forecast depth during the k training period. For this reason, we choose a small k when the time series of interest y_t^0 is very volatile and hence the forecasting performance of each agency highly fluctuates. If the i th agency's forecast error $m' e_{i,t}$ is near zero and hence the forecast depth estimator $\hat{D}_{i,t}$ is near maximum (i.e., near unity), then its forecast $y_{i,t+h}$ gets a high weight; if its forecast error is too large, on the other hand, it gets a low weight or even a zero weight by being trimmed. In this sense, unlike the forecast combination based on the equal weights or forecast mean squared errors, the depth-weighted forecast combination \hat{y}_{t+h} in (11) is robust toward very under-performed (or outlying) forecasts over the training period. Note that the trimming scheme is random as it depends on $\hat{D}_{i,t}$, so \hat{y}_{t+h} is a randomly trimmed forecast combination.

Based on the choice of τ and $W(\cdot)$, \hat{y}_{t+h} in (11) becomes other popular forecast combinations. For instance, when $W(\cdot) = 1$, \hat{y}_{t+h} is the trimmed equally-weighed combined forecast, which converges to the trimmed mean of $y_{i,t+h}$ as $n \rightarrow \infty$ if it exists; when $\tau = 0$ in addition, it is simply the (untrimmed) equally-weighed combined forecast. When $\tau = \max_{1 \leq i \leq n} \hat{D}_{i,t}$, \hat{y}_{t+h} is the same as $y_{i,t+h}$ whose forecast depth is the maximal; if $m_j = 1/k$ in addition, this maximal depth forecast corresponds to the forecast of agency i whose forecasts has been the most accurate during the entire training period (i.e., ex ante the best single forecast). When $y_{i,t+h}$ has a density function that is elliptically symmetric about its mode, this maximal depth forecast becomes the median combination forecast.

We finish this section by summarizing the steps of obtaining the forecast combination \hat{y}_{t+h} .

1. Determine the length of training period k and obtain the k -period forecast error $e_{i,t} = (Y_{i,t} - Y_t^0) = ((y_{i,t-k+1} - y_{t-k+1}^0), \dots, (y_{i,t-1} - y_{t-1}^0), (y_{i,t} - y_t^0))'$ for each i .
2. Determine the form of $k \times 1$ discount vector m in (6) and estimate the dispersion measure s_t of $m' e_{i,t}$ as in (9) or (10).
3. Estimate the forecast depth $\hat{D}_{i,t}$ as in (8).
4. Determine the trimming level $\tau \in (0, 1)$ and some scalar weight function $W(\cdot)$. (One can simply let $W(d) = d$.) Then, obtain the weights $\pi_{i,t}$ as in (7).
5. Obtain the forecast combination $\hat{y}_{t+h} = \sum_{i=1}^n \pi_{i,t} y_{i,t+h}$ as in (11).
6. The prediction interval can be obtained as (20) in Section 4.

3. COMPARISONS AND DISCUSSIONS

3.1. Projection Depth

The forecast depth $\mathcal{D}_{i,t}$ developed in the previous section is in a similar form as one of the popular data depths: the projection depth (e.g., Liu, 1992; Zuo & Serfling, 2000). The data depth measures the outlyingness of a given multivariate sample point with respect to its underlying joint distribution, which is formulated as an index between 0 and 1. If the data point is at the center of the distribution, then the depth value of the data point becomes unity. If a data point locates very far from the center, then the depth value of the point becomes near zero. Recall that, for a k -dimensional random sample $\{Z_i\}$, the projection depth of Z_i is given as $PD_i = 1/[1 + PO_i]$, where the projection-based outlyingness is defined as

$$PO_i = \sup_{\xi \in \mathbb{R}^k : \|\xi\|=1} \frac{|\xi' Z_i - \mu(\xi' Z_i)|}{\sigma(\xi' Z_i)} \quad (12)$$

for some univariate location and dispersion parameters, $\mu(\xi' Z_i)$ and $\sigma(\xi' Z_i)$, of the distribution of $\xi' Z_i$. PO_i is defined to be zero when $\xi' Z_i - \mu(\xi' Z_i) = \sigma(\xi' Z_i) = 0$.

Comparing the forecast depth $\mathcal{D}_{i,t}$ (or forecast outlyingness in (2)) and the projection depth PD_i (or projection-based outlyingness in (12)), we can point two important differences. First, the forecast depth considers the distance from each forecast $m' Y_{i,t}$ toward the observed true $m' Y_t^0$ (i.e., the forecast error $m' e_{i,t} = m' (Y_{i,t} - Y_t^0)$), whereas the projection depth considers the distance toward a central location parameter $\mu(\xi' Z_i)$ of the distribution of $\xi' Z_i$, such as the mean or the median. Since the true $m' Y_t^0$ is not necessarily the center of the distribution of the forecasts $m' Y_{i,t}$, the forecast depth has different implications from the projection depth. It should be emphasized that the original notion of depth is mainly motivated to define a robust central location of multi-dimensional variables. On the other hand, for the forecasting problem, the target location is not the centrality of the distribution of the forecasts $Y_{i,t}$; instead, the target is already given as the observed true value vector Y_t^0 . The forecast depth provides a normalized distance from a vector of forecasts $Y_{i,t}$ toward the vector of observed values Y_t^0 .

Second, we preset the discount vector m in defining the forecast depth that can be potentially heterogeneous, whereas the projection depth needs to search for the ξ vector as defined in (12) so that the outlyingness is maximized to the particular direction. This is possible in the forecasting problem because the researcher often has ordering of the importance among the forecast errors during the training period as we see several examples in the previous section. Though it seems unnecessary, one could find m in defining the forecast depth as ξ in the projection depth. It should be noted that, however, searching for such an m vector is computationally very costly when the dimension of m (i.e., the length of training period k) exceeds 2, which is also a well-known limitation of the projection depth.

Though the forecast depth is different from the standard data depths, it still satisfies the typical properties of the data depth (e.g., Zuo & Serfling, 2000).

In particular, for a given m , $\mathcal{D}_{i,t}$ does not change from any rescaling of the forecast error vector $e_{i,t}$ (*Affine Invariance*); $\mathcal{D}_{i,t}$ reaches the maximal value 1 if the model i makes perfect prediction (*Maximality at Center*) ; $\mathcal{D}_{i,t}$ decreases monotonically as it moves away from the maximal depth location, the deepest point (*Monotonicity Relative to the Deepest Point*); and $\mathcal{D}_{i,t}$ reaches to the minimal value 0 as the forecast error diverges (*Vanishing at Infinity*). The monotonicity yields a well-defined quantile function of $m'e_{i,t}$ since it excludes any quantile-crossing problem, which is to be the key to construct a depth-based trimming in the weight $\pi_{i,t}$. The last property is important for the forecast robustness against very under-performed forecasting agencies or, in other words, outliers.

The depth-weighted forecast combination in (11) is in the form of the depth-weighted trimmed mean (e.g., [Zuo, 2006](#)), and hence it shares the same robustness properties toward outliers (i.e., the very poor forecasts during the training period). In particular, unlike the equal weight combination or inversed forecast mean squared error combination, the depth-weighted forecast combination does not swing much or even stays unchanged when some agencies in the sample yield extremely poor performance during the training period. For instance, when we use the MAD for s_i as in (4), it can be shown that the breakdown point (i.e., the smallest fraction of contaminants in a sample that causes the forecast combination to break down) can reach to the maximal level. For more discussions of the depths and applications in panel data, see [Lee and Sul \(2022a, 2022b\)](#).

3.2. Inversed MSE

Forecast combination based on the inversed forecast mean squared errors (iMSE hereafter; e.g., [Stock & Watson, 2001](#)) is a popular approach in practice. For the $k \times 1$ forecast error vector $e_{i,t}$, the sample iMSE is obtained as $[e'_{i,t}e_{i,t} / k]^{-1}$ and the iMSE combination defines the weight $\pi_{i,t}$ in (1) as

$$\frac{[e'_{i,t}e_{i,t} / k]^{-1}}{\sum_{j=1}^n [e'_{j,t}e_{j,t} / k]^{-1}} = \frac{\left[(1/k) \sum_{\ell=1}^k (y_{i,t-\ell+1} - y_{t-\ell+1}^0)^2 \right]^{-1}}{\sum_{j=1}^n \left[(1/k) \sum_{\ell=1}^k (y_{j,t-\ell+1} - y_{t-\ell+1}^0)^2 \right]^{-1}}. \quad (13)$$

However, one limitation of this weight is that it could overly praise the perfect forecast during the training period, especially when the training period is short. As an extreme example, when we only have one training period (i.e., $k = 1$), if one agency i yields the perfect forecast and hence $e_{i,t} = 0$ or $(e_{i,t}^2)^{-1}$ is unbounded, then her weight $\pi_{i,t}$ in (13) for the forecast combination becomes 1 even when there are other agencies who produce reasonable or even near perfect forecasts. In comparison, the forecast depth is always bounded by unity by construction and it will distribute proper weights both to the agency with perfect forecast and to the other agencies who produce near perfect forecasts. The forecast-depth weight is a more sensible choice in this case, because the single best performer now is not necessarily the best performer in the next periods.

It is also worthy to note that $iMSE[e'_{i,t}e_{i,t}/k]^{-1}$ can be compared with the forecast depth based on the Mahalanobis distance of $e_{i,t}$:

$$\frac{1}{1 + e'_{i,t}\hat{\Sigma}_t^{-1}e_{i,t}}, \quad (14)$$

where $0 < \hat{\Sigma}_t < \infty$ is the $k \times k$ sample variance matrix of $e_{i,t}$. If we ignore the variance and simply let $\hat{\Sigma}_t$ be the identity matrix multiplied by k , then it becomes $(1 + [e'_{i,t}e_{i,t}/k])^{-1}$. It uses the k -dimensional vector of the forecast error $e_{i,t}$ without any discount vector m , but it counts the forecast performance of a specific time during the training period more heavily if the cross-sectional variance of $e_{i,t}$ (i.e., MSE at the specific time) is small.

As an illustration, we compare $iMSE$ with the following three forecast depths with $k = 1$, using 150 simulated forecasts generated from the standard normal when the true value is zero: FD_{mse} , the sample forecast depth in (8) using RMSE \hat{s}_t in (9); FD_{mad} , the sample forecast depth in (8) using MAD \hat{s}_t in (10); MD, the sample forecast depth based on the Mahalanobis forecast distance in (14). The graph on the left in Fig. 1 shows the weights $\pi_{i,t}$ based on $iMSE$ given in (13). We can see that the weight based on the $iMSE$ assigns a huge weight only on a particular agency, whose forecast error during the training period is near zero. In comparison, the graph on the right in Fig. 1 shows the weights based on three forecast depths. Unlike $iMSE$, we can see that the forecast depths well distribute the weights. This difference can be understood from the fact that the forecast depth in (8) considers the historic absolute deviations of each forecasting agency, whereas the $iMSE$ considers each historic squared errors. Compared with MD, the weights based on the forecast depths FD_{mse} and FD_{mad} are sharply concentrated at zero but levy less penalties on outliers. It hence advocates the usefulness of trimming on the forecast-depth based weights.

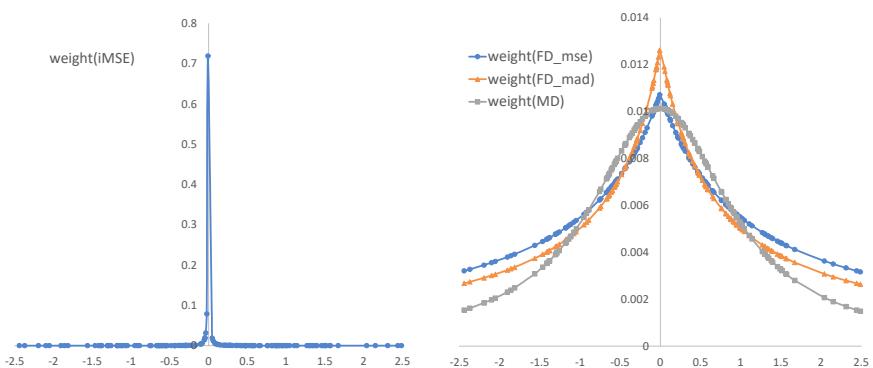


Fig. 1. Comparison of Weights.

4. LIMITING DISTRIBUTION OF COMBINED FORECAST

We now derive limiting distribution of the depth-weighted forecast combination (11), from which we can better understand the factors that affect the robustness of the depth-weighted forecast combination. We can also conduct further inferences using the limiting distribution result, such as constructing the prediction intervals, which is presented at the end of this section.

At time t , for the h -step ahead forecast,⁵ we define the bivariate forecast error vector as

$$\mathbf{x}_{i,t+h} = \begin{pmatrix} m' e_{i,t} \\ y_{i,t+h} - y_{t+h}^0 \end{pmatrix} \in \mathbb{R}^2, \quad (15)$$

which is assumed to be a random sample from an underlying joint distribution F_t for all i . We denote the marginal distributions as F_{1t} and F_{2t} .⁶ The depth estimator $\hat{\mathcal{D}}_{i,t}$ is only based on the forecast errors in the training set $e_{i,t}$ through the form of $m' e_{i,t} = m'(Y_{i,t} - Y_t^0)$, and hence we write $\hat{\mathcal{D}}_{i,t} = \mathcal{D}(m' e_{i,t}, \hat{F}_{1t})$ and $\mathcal{D}_{i,t} = \mathcal{D}(m' e_{i,t}, F_{1t})$, where \hat{F}_t and $(\hat{F}_{1t}, \hat{F}_{2t})$ respectively denote the joint and marginal empirical distributions of $x_{i,t+h}$.⁷ Similarly, we denote $\hat{\mathcal{O}}_{i,t} = \mathcal{O}(m' e_{i,t}, \hat{F}_{1t})$, $\mathcal{O}_{i,t} = \mathcal{O}(m' e_{i,t}, F_{1t})$, $s_t = s(F_{1t})$, and $\hat{s}_t = s(\hat{F}_{1t})$.

Using these notations, we can rewrite the sample forecast error of the depth-weighted forecast combination \hat{y}_{t+h} in (11) as

$$\begin{aligned} \theta^h(\hat{F}_t) &\equiv \hat{y}_{t+h} - y_{t+h}^0 \\ &= \frac{(1/n) \sum_{i=1}^n (y_{i,t+h} - y_{t+h}^0) \mathbf{1}\{\mathcal{D}(m' e_{i,t}, \hat{F}_{1t}) \geq \tau\} W(\mathcal{D}(m' e_{i,t}, \hat{F}_{1t}))}{(1/n) \sum_{i=1}^n \mathbf{1}\{\mathcal{D}(m' e_{i,t}, \hat{F}_{1t}) \geq \tau\} W(\mathcal{D}(m' e_{i,t}, \hat{F}_{1t}))} \\ &= \frac{\int u_2 \mathbf{1}\{\mathcal{D}(u_1, \hat{F}_{1t}) \geq \tau\} W(\mathcal{D}(u_1, \hat{F}_{1t})) d\hat{F}_t(u)}{\int \mathbf{1}\{\mathcal{D}(u_1, \hat{F}_{1t}) \geq \tau\} W(\mathcal{D}(u_1, \hat{F}_{1t})) d\hat{F}_{1t}(u_1)} \end{aligned} \quad (16)$$

for given t , where $u = (u_1, u_2)' \in \mathbb{R}^2$. As the number of forecasts n increases, $\theta^h(\hat{F}_t)$ will converge to a depth-weighted mean forecast error given by

$$\theta^h(F_t) = \frac{\int u_2 \mathbf{1}\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_t(u)}{\int \mathbf{1}\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u_1)} \quad (17)$$

provided $\sup_{m \in \mathbb{R}^k} |\hat{s}_t - s_t| = o_p(1)$ and $\sup_{u \in \mathbb{R}^2} |\hat{F}_t(u) - F_t(u)| = o_p(1)$, which holds in general from the standard results. We can rewrite the numerator of $\theta^h(F_t)$ in (17) as

$$\int \left\{ \int u_2 dF_{2|1,t}(du_2) \right\} \mathbb{1}\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u_1), \quad (18)$$

where $\int u_2 dF_{2|1,t}(u_2) = \mathbb{E}[y_{i,t+h} - y_{i,t+h}^0 | m'e_{i,t}]$. This expression implies that the depth-weighted mean forecast error $\theta^h(F_i)$ in (17) is a weighted average of the projection of $y_{i,t+h} - y_{i,t+h}^0$ on a linear combination of the past forecast errors $e_{i,t}$ during the training period, where the weights are given by $\mathbb{1}\{\mathcal{D}(\cdot, F_{1t}) \geq \tau\} W(\mathcal{D}(\cdot, F_{1t}))$.

To obtain the asymptotic representation of the depth-weighted forecast combination, we define the influence function of $\theta^h(F_t)$ in (17). We let δ_x be the point-mass distribution at $x \in \mathbb{R}^2$ and $F_t(\varepsilon, \delta_x) = (1-\varepsilon)F_t + \varepsilon\delta_x$ be a version of F_t that is contaminated by an ε amount of an arbitrary point-mass distribution at x , where $0 \leq \varepsilon \leq 1$. Then, the influence function of $\theta^h(F_t)$ is defined as

$$\phi(x; \theta^h(F_t)) = \lim_{\varepsilon \rightarrow 0+} \frac{1}{\varepsilon} \left\{ \theta^h(F_t(\varepsilon, \delta_x)) - \theta^h(F_t) \right\}$$

and the limiting distribution of $\theta^h(\hat{F}_t)$ can be obtained from

$$\sqrt{n} \left(\theta^h(\hat{F}_t) - \theta^h(F_t) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(x_{i,t+h}; \theta^h(F_t)) + o_p(1) \quad (19)$$

for given t . To this end, we first assume the following conditions, similarly as [Wu and Zuo \(2009\)](#) and [Lee and Sul \(2022a\)](#). For $\tau \in (0, 1)$, we let $L(\cdot) = -((1-\tau)/\tau)s(\cdot)$ and $U(\cdot) = ((1-\tau)/\tau)s(\cdot)$.

Assumption 1. (i) $W(\cdot)$ is continuously differentiable with a bounded derivative $\dot{W}(\cdot)$. For each t , it holds that (ii) $\int \mathbb{1}\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u_1) > 0$ and $\int |u_2| \mathbb{1}\{\mathcal{D}(u_1, F_{1t}) \geq \tau\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u) < \infty$; (iii) $s(\hat{F}_{1t}), s(F_{1t}) \in (0, \infty)$ with satisfying $\sup_{m \in \mathbb{R}^k} |s(\hat{F}_{1t}) - s(F_{1t})| = o_p(1)$; (iv) $\sup_{u \in \mathbb{R}^2} |\hat{F}_t(u) - F_t(u)| = o_p(1)$; (v) the joint density function $f_t(u_1, u_2)$ of $x_{i,t+h}$ exists and satisfies $\int (u_2 - \theta(F_t)) f_t(u_1, u_2) du_2 < \infty$ at $u_1 = U(F_{1t})$, $L(F_{1t})$.

The following theorem summarizes the asymptotic properties of $\theta^h(\hat{F}_t)$ in (16). The proof is in the Appendix.

Theorem 1. Suppose Assumption 1 holds. Then, for given m , k , h , and t , $\theta^h(\hat{F}_t) \rightarrow_p \theta^h(F_t)$ as $n \rightarrow \infty$. Furthermore, (19) holds, where

$$\phi(x_{i,t+h}; \theta^h(F_t)) = \frac{\phi_1^h(F_t) + \phi_2^h(F_t) + \phi_3^h(F_t)}{\int \mathbb{1}\{L(F_{1t}) \leq u_1 \leq U(F_{1t})\} W(\mathcal{D}(u_1, F_{1t})) dF_{1t}(u_1)}$$

with

$$\begin{aligned}
\phi_1^h(F_t) &= \left(y_{i,t+h} - y_{i,t}^0 - \theta^h(F_t) \right) \mathbf{1}\{L(F_{lt}) \leq m'e_{i,t} \leq U(F_{lt})\} W(\mathcal{D}(m'e_{i,t}, F_{lt})), \\
\phi_2^h(F_t) &= \int (u_2 - \theta^h(F_t)) \mathbf{1}\{L(F_{lt}) \leq u_1 \leq U(F_{lt})\} \dot{W}(\mathcal{D}(u_1, F_{lt})) \\
&\quad \phi_D(m'e_{i,t}; \mathcal{D}(u_1, F_{lt})) dF_t(u), \\
\phi_3^h(F_t) &= \frac{1-\tau}{\tau} W(\tau) \phi_s(m'e_{i,t}; s(F_{lt})) \int (u_2 - \theta^h(F_t)) \\
&\quad \{f_t(U(F_{lt}), u_2) - f_t(L(F_{lt}), u_2)\} du_2, \\
\phi_D(\cdot; \mathcal{D}(u_1, F_{lt})) &= \frac{\mathcal{O}(u_1, F_{lt}) \phi_s(\cdot; s(F_{lt}))}{s(F_{lt}) (1 + \mathcal{O}(u_1, F_{lt}))^2},
\end{aligned}$$

and $\phi_s(\cdot; s(F_{lt}))$ is the influence function of $s(F_{lt})$. Consequently,

$$\sqrt{n} \left(\theta^h(\hat{F}_t) - \theta^h(F_t) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{t,h}^2)$$

as $n \rightarrow \infty$, where $\sigma_{t,h}^2 = \mathbb{E}[\phi(x_{i,t}; \theta^h(F_t))^2]$

One important finding is that the specific form of the asymptotic variance $\sigma_{t,h}^2$ depends on the influence function of $s_t = s(F_{lt})$. Therefore, the choice of s_t heavily affect the robustness property of the depth-weighted forecast combination \hat{y}_{t+h} . For instance, for the RMSE s_t in (3), we can derive the influence function of s_t as

$$\phi_s(x_i; s(F_{lt})) = \frac{x_i^2 - \mathbb{E}[(m'e_{i,t})^2 | \mathcal{I}_t]}{2 \left(\mathbb{E}[(m'e_{i,t})^2 | \mathcal{I}_t] \right)^{1/2}}$$

when $0 < \mathbb{E}[(m'e_{i,t})^2 | \mathcal{I}_t] < \infty$ using the influence function of the mean. For the MAD s_t in (4), we let $\text{sgn}(c) = 1$ if $c > 0$; 0 if $c = 0$; -1 if $c < 0$. Then, using the influence functions of the median, we can derive

$$\phi_s(x_i; s(F_{lt})) = \frac{\text{sgn}(x_i - \lambda(m'e_{i,t}))}{2 f_{lt}(\lambda(m'e_{i,t}))},$$

where $\lambda(m'e_{i,t}) = \inf \{v : \mathbb{P}(|m'e_{i,t}| \leq v | \mathcal{I}_t) \geq 1/2\}$ is the conditional median of $m'e_{i,t}$, provided that the marginal density function $f_{lt}(\cdot)$ of $m'e_{i,t}$ at t satisfies $0 < f_{lt}(\lambda(m'e_{i,t})) < \infty$. Recall that the mean has an unbounded influence function when the support of F_{lt} is not bounded, whereas the median has a bounded influence function. It yields that the influence function of the RMSE is not necessarily

bounded, whereas that of the MAD is bounded. For this reason, the forecast combination using the MAD is more robust toward outliers or extremely underperformed forecasts.

From Theorem 1, we can conclude that the depth-weighted forecast combination \hat{y}_{t+h} satisfies $\hat{y}_{t+h} \xrightarrow{p} y_{t+h}^0 + \theta^h(F_t)$ and $\sqrt{n}(\hat{y}_{t+h} - \{y_{t+h}^0 + \theta^h(F_t)\}) \xrightarrow{d} \mathcal{N}(0, \sigma_{t+h}^2)$ as $n \rightarrow \infty$. Apparently, when $\theta^h(F_t) = 0$, \hat{y}_{t+h} becomes a consistent forecast. A sufficient condition for $\theta^h(F_t) = 0$ is that the forecasts $\{y_{i,t+h}\}$ are distributed symmetrically about the true value y_{t+h}^0 , which could be obtained once extreme forecasts are trimmed out. In this case, the $100(1-\alpha)\%$ prediction interval of y_{t+h}^0 can be obtained as

$$\left[\hat{y}_{t+h} \pm z_{\alpha/2} \frac{\hat{\sigma}_{t+h}}{\sqrt{n}} \right], \quad (20)$$

where $z_{\alpha/2}$ is the $(1-(\alpha/2))$ th quantile of the standard normal distribution and $\hat{\sigma}_{t+h}$ is a consistent estimator of σ_{t+h}^2 . Knowing that the weighted average of the individual interval forecasts does not necessarily provide a correct coverage rate (e.g., Timmermann, 2006), the prediction interval in (20) can be alternatively used, which does not rely on subsampling or bootstrap. σ_{t+h}^2 can be estimated as $\hat{\sigma}_{t+h}^2 = n^{-1} \sum_{i=1}^n \hat{\phi}(x_{i,t+h}; 0)^2$, where we replace y_{t+h}^0 by \hat{y}_{t+h} and use kernel density estimators and sample analogues of the terms in $\hat{\phi}(x_{i,t+h}; 0)$.⁸

5. FORECASTING NEW COVID-19 CASES

We apply the depth-weighted forecast combination (11) to predict weekly COVID-19 cases in the United States. The data set is collected from the CDC COVID Data Tracker (https://covid.cdc.gov/covid-data-tracker/#forecasting_weeklydeaths) as of August 7, 2021, which is updated on August 11, 2021. The data set consists of h -step ahead weekly forecast history from 40 individual modeling groups for $h = 1, 2, 3, 4$.⁹ It also includes the h -step ahead ensemble forecast that is reported in the weekly forecast digest by the CDC.¹⁰ We use the past 50 weekly forecasts from the week ending on August 29, 2020 (when all the h -step ahead forecasts became available) to the week ending on August 7, 2021.

We compare different forecast combination approaches, including the equally-weighted average of all the available forecasts at each t , ensemble forecast reported by the CDC (“Ensem”); inversed MSE based forecast combination (“iMSE”) as in (13); and the forecast-depth based combination (FD) developed in this chapter. For the forecast-depth based combination, for a given training period size $k = 2, 3, 4$, we consider two types of the discount vector $m = (m_1, \dots, m_k)'$:

- Type 1 (FD1): $m_j = (0.2)^{k-j} / \sum_{\ell=1}^k (0.2)^{k-\ell}$ for $j = 1, \dots, k$;
- Type 2 (FD2): $m_j = (j/k)^4 / \sum_{\ell=1}^k (\ell/k)^4$ for $j = 1, \dots, k$.

Both discount vectors have similar magnitudes. But Type 1 discounts the past information more heavily with large k ; Type 2 discounts the past information more heavily with small k . For each type, two forms of \hat{s}_t are considered:

- Sample RMSE in (9): “ $FD1_{mse}$ ” and “ $FD2_{mse}$ ”;
- Sample MAE in (10): “ $FD1_{mad}$ ” and “ $FD2_{mad}$ ”.

For $W(\cdot)$, we simply let $W(d) = d$. For the trimming parameter τ , we consider the thresholds at the 0% (no trimming), 10%, 20%, 30%, 40%, and 50% levels of the lowest estimated forecast depth in the sample. For instance, at the 20% threshold, we set τ such that $(1/n)\sum_{i=1}^n 1\{\hat{D}_{i,t} < \tau\} = 0.2$. We also report the trimmed iMSE forecast combinations, where the trimming thresholds are determined similarly as the forecast depth: 0% (no trimming), 10%, 20%, 30%, 40%, and 50% levels of the lowest estimated iMSE in the sample. For each horizons h , we obtain the weights $\pi_{i,t}$ based on the performance of the h -step ahead forecasts during the training period.

Tables 1–4 report the out-of-sample forecast mean squared error (FMSE) comparisons among different forecast combinations for h -step ahead forecasting for $h = 1, \dots, 4$, respectively. For each k , the values in the tables are the ratio

Table 1. 1-Step Ahead FMSE Ratio to the Equal-Weight Forecast Combination.

k	trim	\bar{n}	Ensem	iMSE	$FD1_{mse}$	$FD1_{mad}$	$FD2_{mse}$	$FD2_{mad}$
2	0.0	24.3	0.931	0.863	0.941	0.938	0.941	0.938
	0.1			0.866	0.871	0.873	0.873	0.873
	0.2			0.857	0.881	0.882	0.855	0.857
	0.3			0.880	0.902	0.901	0.884	0.885
	0.4			0.886	0.892	0.888	0.877	0.879
	0.5			0.899	0.855	*0.854	0.861	0.864
3	0.0	23.9	0.931	0.895	0.945	0.942	0.950	0.946
	0.1			0.900	0.872	0.873	0.895	0.894
	0.2			0.887	0.908	0.906	*0.870	0.872
	0.3			0.912	0.897	0.895	0.894	0.894
	0.4			0.926	0.882	0.877	0.887	0.888
	0.5			0.945	0.889	0.884	0.895	0.896
4	0.0	23.5	0.931	0.872	0.945	0.942	0.952	0.949
	0.1			*0.863	0.875	0.876	0.900	0.900
	0.2			0.869	0.913	0.911	0.907	0.908
	0.3			0.894	0.896	0.895	0.902	0.902
	0.4			0.913	0.889	0.885	0.903	0.903
	0.5			0.953	0.897	0.893	0.915	0.916

Notes: The table presents the ratio of the forecast mean squared error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 – 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; $FD1_{mse}$ is the forecast-depth combination with type-1 m and RMSE \hat{s}_t ; $FD1_{mad}$ is the forecast-depth combination with type-1 m and MAD \hat{s}_t ; $FD2_{mse}$ is the forecast-depth combination with type-2 m and RMSE \hat{s}_t ; $FD2_{mad}$ is the forecast-depth combination with type-2 m and MAE \hat{s}_t . The best performers are marked with * in each case.

Table 2. 2-Step Ahead FMSE Ratio to the Equal-Weight Forecast Combination.

<i>k</i>	trim	\bar{n}	Ensem	iMSE	FD1 _{mse}	FD1 _{mad}	FD2 _{mse}	FD2 _{mad}
2	0.0	22.8	0.943	1.120	0.962	0.963	0.940	0.933
	0.1			1.110	0.916	0.916	0.888	0.881
	0.2			1.100	0.891	0.893	0.875	0.865
	0.3			1.100	0.874	0.877	0.831	0.826
	0.4			1.090	0.866	0.869	0.834	0.829
	0.5			1.080	0.824	0.828	0.761	*0.759
3	0.0	22.4	0.943	0.949	0.971	0.973	0.950	0.943
	0.1			0.934	0.923	0.925	0.896	0.889
	0.2			0.944	0.897	0.900	0.884	0.875
	0.3			0.953	0.882	0.887	0.845	0.841
	0.4			0.971	0.867	0.872	0.848	0.844
	0.5			0.976	0.861	0.866	0.787	*0.785
4	0.0	22.0	0.943	0.934	0.978	0.981	0.959	0.954
	0.1			0.911	0.934	0.935	0.910	0.904
	0.2			0.905	0.908	0.911	0.903	0.896
	0.3			0.917	0.893	0.898	0.856	0.852
	0.4			0.959	0.882	0.887	0.855	0.852
	0.5			0.952	0.862	0.868	0.826	*0.823

Notes: The table presents the ratio of the forecast mean squared error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 – 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; FD1_{mse} is the forecast-depth combination with type-1 *m* and RMSE \hat{s}_p ; FD1_{mad} is the forecast-depth combination with type-1 *m* and MAD \hat{s}_p ; FD2_{mse} is the forecast-depth combination with type-2 *m* and RMSE \hat{s}_p ; FD2_{mad} is the forecast-depth combination with type-2 *m* and MAE \hat{s}_p . The best performers are marked with * in each case.

of the FMSE of each combined forecast to that of the equally-weighted forecast combination over the past $(50 - k)$ weeks of forecasting, where the weights for the forecast combinations of the iMSE and the forecast-depth at *t* are obtained using the training period of $(t-1, \dots, t-k)$. So, values less than 1 implies that the FMSE is smaller than that of the equally-weighted forecast combination; smaller values implies better performance.¹¹ Values marked with ** indicates the best performer in each case. In each table, *k* is the training period size and “trim” is the trimming proportion as we described above. The relative performance of different agencies can change over time, and we compute the weights over short rolling windows. For a given target forecasting week *t*, any agency *i* is dropped if it does not have at least *k* most recent forecasts to form the training period. We hence do not need a balanced panel over the entire period; we only use a short balanced panel of length *k* + 1 at each target forecasting week *t*, whose individual members can be different over *t*. In each table of the *h*-step ahead forecast, \bar{n} reports the average of the cross-sectional sample size n_t over the past $(50 - k)$ weeks (i.e., $\bar{n} = (50 - k)^{-1} \sum_{t=k+1}^{50} n_t$), where n_t is the number of individuals who report the *h*-step ahead forecast at *t* with *k* training periods and hence used for forecast combination.

Table 3. 3-Step Ahead FMSE Ratio to the Equal-Weight Forecast Combination.

k	trim	\bar{n}	Ensem	iMSE	$FD1_{mse}$	$FD1_{mad}$	$FD2_{mse}$	$FD2_{mad}$
2	0.0	21.0	0.989	0.832	0.924	0.915	0.920	0.911
	0.1			0.819	0.853	0.848	0.825	0.823
	0.2			0.792	0.791	0.789	0.796	0.794
	0.3			0.789	0.794	0.791	0.773	0.771
	0.4			0.769	0.764	0.761	0.731	0.730
	0.5			0.789	0.744	0.739	0.710	*0.709
3	0.0	20.6	0.989	0.869	0.932	0.923	0.925	0.916
	0.1			0.855	0.863	0.858	0.827	0.825
	0.2			0.835	0.788	0.787	0.801	0.800
	0.3			0.831	0.802	0.799	0.784	0.782
	0.4			0.826	0.782	0.777	0.734	0.734
	0.5			0.805	0.748	0.744	*0.695	0.696
4	0.0	20.1	0.989	0.853	0.940	0.931	0.934	0.924
	0.1			0.832	0.867	0.862	0.832	0.830
	0.2			0.815	0.790	0.789	0.809	0.807
	0.3			0.799	0.801	0.798	0.785	0.783
	0.4			0.795	0.791	0.786	0.744	0.744
	0.5			0.787	0.747	0.744	0.723	*0.722

Notes: The table presents the ratio of the forecast mean squared error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 - 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; $FD1_{mse}$ is the forecast-depth combination with type-1 m and RMSE \hat{s}_r ; $FD1_{mad}$ is the forecast-depth combination with type-1 m and MAD \hat{s}_r ; $FD2_{mse}$ is the forecast-depth combination with type-2 m and RMSE \hat{s}_r ; $FD2_{mad}$ is the forecast-depth combination with type-2 m and MAE \hat{s}_r . The best performers are marked with * in each case.

The main point of interest is how the forecast-depth based method performs even with a very short training period. It is evident that the forecast combinations based on the forecast-depth and the iMSE stand out; the forecast-depth based combination shows the best performance in general. It does not show the typical “forecast combination puzzle” – the equal weight outperforms the estimated weights in forecast combination. For some cases, the iMSE combination outperforms the forecast-depth based method, but it is mostly for the cases when we do not employ (enough) trimming. The trimming generally improves the FMSE though the change is not strictly monotonic to the trimming proportion. However, the benefit from trimming is much larger in FD than iMSE, and all the FD approaches eventually perform better than iMSE with proper trimming for all the cases. Since we compare the FMSE, we expected that the iMSE should perform better than FD as k increases, but the current results do not support it. This seems because the weekly COVID-19 cases fluctuates much and the forecasting agencies can hardly provide good predictions over an extended period consistently. For the choice of \hat{s}_r in FD, it does not make big difference, but MAE outperforms RMSE in general.

Table 4. 4-Step Ahead FMSE Ratio to the Equal-Weight Forecast Combination.

k	trim	\bar{n}	Ensem	iMSE	$FD1_{mse}$	$FD1_{mad}$	$FD2_{mse}$	$FD2_{mad}$
2	0.0	17.6	0.949	0.923	0.948	0.937	0.943	0.930
	0.1			0.900	0.876	0.871	0.876	0.870
	0.2			0.894	0.822	0.820	0.819	0.817
	0.3			0.893	0.793	0.789	0.780	0.776
	0.4			0.873	0.765	0.765	0.744	*0.741
	0.5			0.888	0.784	0.783	0.772	0.770
3	0.0	17.0	0.949	0.911	0.959	0.947	0.951	0.937
	0.1			0.897	0.880	0.875	0.879	0.873
	0.2			0.873	0.843	0.839	0.832	0.828
	0.3			0.875	0.796	0.792	0.772	*0.768
	0.4			0.845	0.778	0.778	0.775	0.771
	0.5			0.886	0.779	0.778	0.792	0.789
4	0.0	16.4	0.949	0.928	0.966	0.954	0.959	0.943
	0.1			0.905	0.883	0.878	0.881	0.875
	0.2			0.912	0.843	0.839	0.827	0.823
	0.3			0.908	0.792	0.789	0.761	0.758
	0.4			0.853	0.776	0.776	0.787	0.781
	0.5			0.858	0.732	0.733	*0.727	*0.727

Notes: The table presents the ratio of the forecast mean squared error of each combined forecast to that of the equally-weighted forecast combination, which are averaged over the past 50 weeks of forecasting, 8/29/2020 – 8/7/2021. Ensem is ensemble forecast by CDC; iMSE is inversed MSE forecast combination; $FD1_{mse}$ is the forecast-depth combination with type-1 m and RMSE \hat{s}_t ; $FD1_{mad}$ is the forecast-depth combination with type-1 m and MAD \hat{s}_t ; $FD2_{mse}$ is the forecast-depth combination with type-2 m and RMSE \hat{s}_t ; $FD2_{mad}$ is the forecast-depth combination with type-2 m and MAE \hat{s}_t . The best performers are marked with * in each case.

Table 5 compares the forecast combinations based on the forecast-depth and the iMSE in more details. In particular, it summarizes the bias ratios and the variance ratios between $FD2_{mad}$ and iMSE for 2-step ahead forecasts in **Table 2** above. It shows that the FMSE improvement of $FD2_{mad}$ mostly comes from the bias reduction.

In [Fig. 2](#), instead of averaging over the entire period as in the [Tables 1–4](#), we depict the weekly 1-step ahead forecast error ratio paths of Ensem, iMSE, $FD1_{mse}$, and $FD1_{mae}$ for $k = 1$ and with 30% trim. Since we consider the case with $k = 1$ period of training, $FD1_{mse} = FD2_{mse}$, and $FD1_{mae} = FD2_{mae}$. We can see that iMSE is quite volatile and some error ratios (to the error of equally-weighted forecast combination) are even off the chart. Ensem is still volatile though at a much smaller scale than iMSE. In comparison, the forecast error ratio paths of $FD1_{mse}$ and $FD1_{mae}$ show very little fluctuations and they mostly lie below 1, which implies that their forecast errors are smaller than that of the equally-weighted forecast combination.

[Fig. 3](#) depicts the weights based on $FD1_{mse}$, $FD1_{mad}$, and iMSE on a random date, October 10, 2020, with $k = 1$ and the 1-step ahead forecasts. Unlike the forecast-depth based forecast combinations, it shows that the weight based on iMSE assigns 99% of the total weight only on three agencies, which are 0.072, 0.171, 0.745, where the weight 0.745 is out of the chart.

Table 5. Comparison Between FD and iMSE.

k	trim	FD2 _{mad}		iMSE		Ratio (FD/ iMSE)	
		bias	stdv	bias	stdv	bias	stdv
2	0.0	-2497	204259	7520	223608	-0.332	0.913
	0.1	-1230	198483	8547	222433	-0.144	0.892
	0.2	-310	196746	8123	221738	-0.038	0.887
	0.3	714	192212	9046	221677	0.079	0.867
	0.4	4313	192573	10507	220445	0.410	0.874
	0.5	3492	184243	11998	219330	0.291	0.840
3	0.0	-2759	207473	3970	208199	-0.695	0.997
	0.1	-1538	201465	4979	206508	-0.309	0.976
	0.2	-902	199823	5240	207571	-0.172	0.963
	0.3	391	195373	6891	208545	0.057	0.937
	0.4	3754	196203	8156	210470	0.460	0.932
	0.5	3162	189271	11974	210770	0.264	0.898
4	0.0	-2675	210510	6363	208627	-0.420	1.009
	0.1	-1658	204900	6790	206073	-0.244	0.994
	0.2	-1165	203155	6180	205332	-0.189	0.989
	0.3	-12	198436	8219	206671	-0.001	0.960
	0.4	2985	198852	11271	211167	0.265	0.942
	0.5	4292	191383	16111	210139	0.266	0.911

Notes: “bias” is the average forecast bias; “stdv” is the average standard deviation of the forecast error; and “mse” is the average forecast mean squared error. Values in the table are based on the 2-step ahead forecasts using the same data set as in Table 2.

Finally, based on the available forecasts up to the week ending September 4, 2021, we report predictions for the next 4 weeks (ending on 8/14/2021, 8/21/2021, 8/28/2021, 9/4/2021) as of August 7, 2021. For each h -week ahead prediction

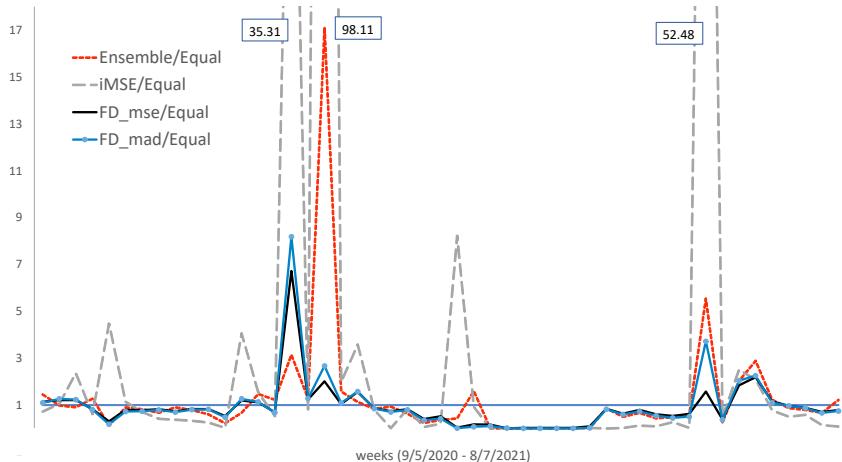


Fig. 2. Forecast Error Ratio to the Equal-Weight Forecast Combination. (Note: The dates on three peaks are 12/19/2020 (35.31), 1/2/2021 (98.11), and 6/12/2021 (52.48).)

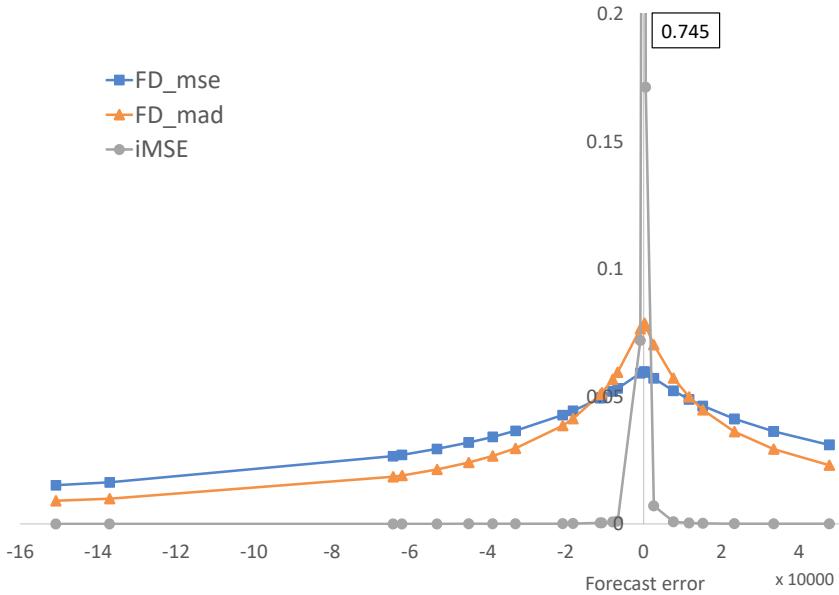


Fig. 3. Weights of Forecast Combinations.

($h = 1, 2, 3, 4$), we consider 25 different cases of $k = 1, \dots, 5$ and $\text{trim} = 0.1, \dots, 0.5$ for each of the four forecast combinations, FD1_{mse} , FD1_{mad} , FD2_{mse} , and FD2_{mad} . (Hence, 100 different forecast combinations for each h .) Fig. 4 reports the

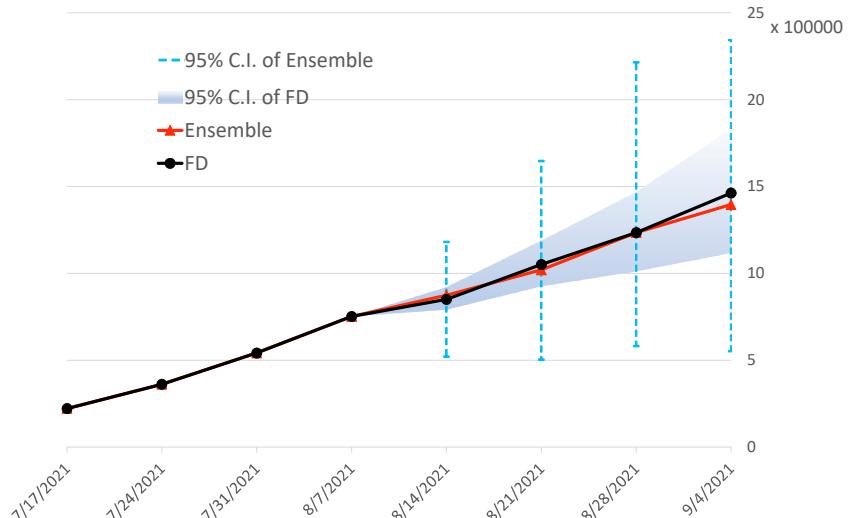


Fig. 4. COVID-19 New Cases Forecast (US National).

predictions of the average of these depth-weighted forecast combinations in black circle line (FD) and the ensemble forecast reported by the CDC in red triangle line (Ensemble). The predictions between these two methods are quite similar. However, the noticeable difference is in their prediction intervals. The pointwise 95% prediction interval of the FD forecast combination point is depicted by the shaded area, which is obtained as the maximum of the upper bound and the minimum of the lower bound points of the 95% prediction intervals of all the 100 depth-weighted forecast combinations using the normal approximation based on (20). Compared with the prediction interval of the ensemble forecast reported by CDC (in the dashed vertical lines), it shows much narrower bounds.

6. CONCLUDING REMARKS

In this chapter, we develop the forecast depth and a depth-weighted forecast combination with trimming. Since the weights are not obtained by minimizing a loss function, we do not discuss any optimality properties. However, the weights can be calculated even when we have many forecasts but the training period is as short as just one, and hence it can be practically very useful as complementing other forecast combinations. In comparison, when long training period is available, we can apply LASSO in estimating time-invariant weights by minimizing a L_2 loss function with L_1 penalty terms (e.g., Diebold & Shin, 2019), from which we can obtain weights on each forecasting individual with selection (i.e., trimming).

It should be noted that our approach ignores any estimation error in the forecast values, which is because we do not know the underlying model that yields each forecast value. However, we deal with random weights with endogenous trimming when deriving the limiting distribution of the forecast combination, which is the main analytical contribution of this chapter. In this regards, our approach should be distinguished from the model averaging approaches.

We can extend the forecast combination idea to multivariate forecasting. Since we can construct depth-based contour (i.e., multivariate quantile), it can provide a ranking among different models based on their forecast performance for multiple economic variables together. In addition, depending on the choice of the dispersion term s_p , the depth-weighted forecast combination does not necessarily require existence of the moments of the forecasts. Therefore, it can be applied for financial data with fat-tailed distributions.

NOTES

1. When the data set consists of h -step ahead forecasts for multiple values of h , we can use the h -step ahead forecasts during the training period for the h -step ahead forecast of y_{t+h} . In this cases, the weights $\pi_{i,t}$ can be different across h , say $\pi_{i,t}^h$. But the weights still use the information available at t and we omit the index h in $\pi_{i,t}$ for the sake of notational simplicity.

2. When $m'e_{i,t} = s_i = 0$, we define $\mathcal{O}_{i,t} = 0$.

3. If the distribution of $\{e_{i,t}\}$ is heterogeneous across i , we can consider $\mathcal{O}_{i,t} = |m_i' e_{i,t}| / s_t$, where $m_i = (m_{i1}, \dots, m_{ik})'$ be a $k \times 1$ vector with $\sum_{j=1}^k m_{ij} = 1$ for each i .

In such cases, s_t can be alternatively defined as $\left(\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbb{E}[(m_i' e_{i,t})^2 | \mathcal{I}_t] \right)^{1/2}$ and $\lim_{n \rightarrow \infty} \text{med}_{1 \leq i \leq n} \inf \{v_i : \mathbb{P}(|m_i' e_{i,t}| \leq v_i | \mathcal{I}_t) \geq 1/2\}$, respectively.

4. Instead of the hard-threshold trimming, we could consider a smooth transition function such as

$$\begin{cases} \frac{\exp(-c(1-(d/\tau))^2 - \exp(c))}{1-\exp(c)} & \text{if } d < \tau \\ 1 & \text{if } d \geq \tau \end{cases}$$

for some smoothing parameter $c > 0$. As $c \rightarrow \infty$, it approaches to a binary indicator function $1\{d \geq \tau\}$.

5. Note that we consider “direct” forecast because we do not observe each forecasting models nor impose any (autoregressive) dynamic structures.

6. If we further suppose that, for each i , the forecast error $y_{i,t} - y_t^0$ is stationary over t , we can drop the subscript t in the distribution notations. However, it is not required to derived the main results. It is important to note that, however, imposing stationarity of the forecast error does not exclude the potential nonstationarity of the observed series y_t^0 itself and the forecast series $y_{i,t}$ for each i .

7. By the affine invariance property of the forecast depth, we have $\mathcal{D}(m' e_{i,t}, F_{it}) = \mathcal{D}(e_{i,t}, \underline{F}_{it})$, where \underline{F}_{it} is the joint distribution of $e_{i,t} \in \mathbb{R}^k$.

8. Because \hat{y}_{t+h} is a cross-sectional weighted average, one could instead consider $\hat{\sigma}_{t,h}^2 = n \sum_{i=1}^n \pi_{i,t}^2 (y_{i,t+h} - \hat{y}_{t+h})^2$ or $\hat{\sigma}_{t,h}^2 = \sum_{i=1}^n \pi_{i,t} (y_{i,t+h} - \hat{y}_{t+h})^2$ in this case by ignoring the randomness in $\pi_{i,t}$, when serial dependence between $y_{i,t+h}$ and $Y_{i,t} = (y_{i,t-k+1}, \dots, y_{i,t-1}, y_{i,t})'$ is not too strong.

9. The list of the forecasting agencies can be found at <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html>.

10. From April 13 to July 21, 2020, the ensemble was created by the arithmetic average of each prediction quantile for all eligible models for a given location. The confidence interval at each prediction point is also calculated from the corresponding quantile ensembles ([Busetti, 2017](#)). However, starting on the week of July 28, 2020, which is the sample period of our analysis here, the median prediction was instead used across all eligible models at each quantile level. As of August 7, 2021, 23 models are included for the ensemble forecast. For further details about the ensemble forecast, see <https://covid19forecasthub.org/doc/ensemble/>.

11. Note that we drop one agency in the forecast pool, who reported very erroneous forecasts between 3/13/2021 and 4/17/2021. If we compare the FMSE of the equally-weighted forecast combinations with (“EQ₋₁”) and without (“EQ_{all}”) dropping this agency, we have

h	1	2	3	4
EQ_{all}/EQ_{-1}	173.19	194.49	229.15	245.05

which shows that the equally-weighted forecast combination is prone to swing much by only one outlier.

REFERENCES

- Aiolfi, M., & Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1–2), 31–53.
- Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20(4), 451–468.
- Busetti, F. (2017). Quantile aggregation of density forecasts. *Oxford Bulletin of Economics and Statistics*, 79(4), 495–512.
- Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics*, 214(1), 274–294.
- Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, 192(1), 152–167.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4), 559–583.
- Diebold, F. X., & Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting*, 6(4), 503–508.
- Diebold, F. X., & Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting*, 35(4), 1679–1691.
- Granger, C. W. J., & Jeon, Y. (2004). Thick modelling. *Economic Modelling*, 21(2), 323–343.
- Hu, B., Park, J. Y., & Qian, J. (2017). *Analysis of distributional dynamics for repeated cross-sectional and intra-period observations* [Working Paper]. Indiana University.
- Lee, Y., & Sul, D. (2022a). Depth-weighted means of noisy data: An application to estimating the average effect in heterogeneous panels. *Journal of Multivariate Analysis*, forthcoming.
- Lee, Y., & Sul, D. (2022b). Trimmed mean group estimation. *Advances in Econometrics*, 43B, 177–202.
- Liu, R. Y. (1992). Data depth and multivariate rank tests. In Y. Dodge (Ed.), *L1-Statistical analysis and related methods* (pp. 279–294). North-Holland.
- Stock, J. H., & Watson, M. (2001). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In R. F. Engle & H. White (Eds.), *Festschrift in honour of Clive Granger* (pp. 1–44). Cambridge University Press.
- Stock, J. H., & Watson, M. (2006). Forecasting with many predictors. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, pp. 515–554). North-Holland.
- Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting* (Vol. 1, pp. 135–196). North-Holland.
- Wu, M., & Zuo, Y. (2009). Trimmed and Winsorized means based on a scaled deviation. *Journal of Statistical Planning and Inference*, 139(2), 350–365.
- Zuo, Y. (2006). Multidimensional trimming based on projection depth. *Annals of Statistics*, 34(5), 2211–2251.
- Zuo, Y., & Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics*, 28(2), 461–482.

APPENDIX: PROOF OF THEOREM 1

To simplify the notations, we drop the subscript “ t ” in the distribution notations in the proof. So, we simply denote the joint distribution and density of $x_{i,t,h}$ in (15) as F and f respectively; the marginal distributions as F_1 and F_2 . The corresponding empirical distributions are denoted as \hat{F} , \hat{F}_1 , and \hat{F}_2 , respectively. Over the proof, $u = (u_1, u_2)' \in \mathbb{R}^2$ are used as generic variables of functions and in the integrals. We also define

$$\begin{aligned}\hat{\nu}(\cdot) &= \sqrt{n}(\hat{F}_1(\cdot) - F_1(\cdot)), \\ \hat{H}(\cdot) &= \sqrt{n}(\mathcal{D}(\cdot, \hat{F}_1) - \mathcal{D}(\cdot, F_1)).\end{aligned}$$

The consistency follows from Lemma A.5 and Theorem 6 of [Zuo \(2006\)](#) because we have

$$\begin{aligned}\sup_{u_1 \in \mathbb{R}} |\mathcal{D}(u_1, \hat{F}_1) - \mathcal{D}(u_1, F_1)| &= \sup_{m \in \mathbb{R}^k} \left| \frac{1}{1 + |u_1| / s(\hat{F}_1)} - \frac{1}{1 + |u_1| / s(F_1)} \right| \\ &\leq \sup_{m \in \mathbb{R}^k} \left\{ \frac{|s(\hat{F}_1) - s(F_1)|}{s(\hat{F}_1)} \frac{\mathcal{O}(u_1, F_1)}{1 + \mathcal{O}(u_1, F_1)} \right\} = o_p(1)\end{aligned}$$

as we assume $\sup_{m \in \mathbb{R}^k} |s(\hat{F}_1) - s(F_1)| = o_p(1)$ and $s(\cdot) \in (0, \infty)$. Note that $\mathcal{O}(u_1, F_1) = |u_1| / s(F_1)$ and $\mathcal{O}(u_1, F_1) / (1 + \mathcal{O}(u_1, F_1)) \in [0, 1]$ by construction. The asymptotic normality follows similarly as the proof of Theorem 4.1 in [Wu and Zuo \(2009\)](#), so we sketch the proof here. Recall that we define $L(\cdot) = -((1 - \tau) / \tau)s(\cdot)$ and $U(\cdot) = ((1 - \tau) / \tau)s(\cdot)$, where $\tau \in (0, 1)$ is the trimming parameter introduced in (7). Since $L(\cdot) < U(\cdot)$ by construction, we write

$$\begin{aligned}\sqrt{n}(\theta^h(\hat{F}) - \theta^h(F)) &= \frac{\sqrt{n} \int u_2^* \mathbf{1}\{\mathcal{D}(u_1, \hat{F}) \geq \tau\} W(\mathcal{D}(u_1, \hat{F})) d\hat{F}(u)}{\int \mathbf{1}\{\mathcal{D}(u_1, \hat{F}_1) \geq \tau\} W(\mathcal{D}(u_1, \hat{F}_1)) d\hat{F}_1(u_1)} \\ &= \frac{\sqrt{n} \int_{L(\hat{F}_1)}^{U(\hat{F}_1)} \left\{ \int u_2^* d\hat{F}_{2|1}(u_2 | u_1) \right\} W(\mathcal{D}(u_1, \hat{F}_1)) d\hat{F}_1(u_1)}{\int_{L(\hat{F}_1)}^{U(\hat{F}_1)} W(\mathcal{D}(u_1, \hat{F}_1)) d\hat{F}_1(u_1)},\end{aligned}$$

where $u_2^* = u_2 - \theta^h(F)$. We decompose the numerator into

$$\begin{aligned}N_{1n} &= \sqrt{n} \int_{L(F_1)}^{U(F_1)} \hat{g}(u_1) W(\mathcal{D}(u_1, F_1)) d\hat{F}_1(u_1), \\ N_{2n} &= \int_{L(F_1)}^{U(F_1)} \hat{g}(u_1) \sqrt{n} \{W(\mathcal{D}(u_1, \hat{F}_1)) - W(\mathcal{D}(u_1, F_1))\} d\hat{F}_1(u_1), \\ N_{3n} &= \sqrt{n} \left\{ \int_{L(\hat{F}_1)}^{U(\hat{F}_1)} - \int_{L(F_1)}^{U(F_1)} \right\} \hat{g}(u_1) W(\mathcal{D}(u_1, \hat{F}_1)) d\hat{F}_1(u_1),\end{aligned}$$

where $\hat{g}(u_1) = \int u_2^* d\hat{F}_{2|1}(u_2 | u_1)$.

For N_{1n} , we immediately have

$$N_{1n} = \sqrt{n} \int u_2^* 1\{\mathcal{D}(u_1, F_1) \geq \tau\} W(\mathcal{D}(u_1, F_1)) d\hat{F}_1(u_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_{1i},$$

where

$$\eta_{1i} = (y_{i,t+h} - y_{i,t+h}^0 - \theta^h(F)) 1\{L(F_1) \leq m' e_{i,t} \leq U(F_1)\} W(\mathcal{D}(m' e_{i,t}, F_1)).$$

For N_{2n} , we note that $\sup_{u_1 \in \mathbb{R}} |\hat{g}(u_1) - g(u_1)| = o_p(1)$ with $g(u_1) = \int u_2^* dF_{2|1}(u_2 | u_1)$ from the standard results of nonparametric conditional expectation estimators. We thus have

$$N_{2n} = \int_{L(F_1)}^{U(F_1)} g(u_1) \dot{W}(\hat{\Delta}(u_1)) \hat{H}(u_1) d\hat{F}_1(u_1) + o_p(1)$$

for some $\hat{\Delta}(u_1)$ between $\mathcal{D}(u_1, \hat{F}_1)$ and $\mathcal{D}(u_1, F_1)$, where $\sup_{u_1 \in [L(F_1), U(F_1)]} |\hat{\Delta}(u_1) - \mathcal{D}(u_1, F_1)| \leq \sup_{u_1 \in [L(F_1), U(F_1)]} |\mathcal{D}(u_1, \hat{F}_1) - \mathcal{D}(u_1, F_1)| \leq C \sup_{m \in \mathbb{R}^k} |s(\hat{F}_1) - s(F_1)| = o_p(1)$ for some positive $C < \infty$. By Lemma A.3 of Wu and Zuo (2009), $\sup_{u_1 \in [L(F_1), U(F_1)]} (1 + |u_1|) |\hat{H}(u_1)| = O_p(1)$ and there exists $\phi_D(x_1; \mathcal{D}(u_1, F_1))$ for $x_1 \in \mathbb{R}$ such that $\hat{H}(u_1) = \int \phi_D(x_1; \mathcal{D}(u_1, F_1)) d\hat{\nu}(x_1) + o_p(1)$ uniformly over $x_1 \in [L(F_1), U(F_1)]$. Similarly as the proof of Theorem 2 in Lee and Sul (2022a), therefore, we can verify that

$$\begin{aligned} N_{2n} &= \int_{L(F_1)}^{U(F_1)} g(u_1) \dot{W}(\mathcal{D}(u_1, F_1)) \left(\int \phi_D(x_1; \mathcal{D}(u_1, F_1)) d\hat{\nu}(x_1) \right) dF_1(u_1) + o_p(1) \\ &= \int \int u_2^* 1\{L(F_1) \leq u_1 \leq U(F_1)\} \dot{W}(\mathcal{D}(u_1, F_1)) \phi_D(x_1; \mathcal{D}(u_1, F_1)) dF(u) d\hat{\nu}(x_1) \\ &\quad + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_{2i} + o_p(1), \end{aligned}$$

where

$$\eta_{2i} = \int u_2^* 1\{L(F_1) \leq u_1 \leq U(F_1)\} \dot{W}(\mathcal{D}(u_1, F_1)) \phi_D(m' e_{i,t}; \mathcal{D}(u_1, F_1)) dF(u)$$

and $\phi_D(\cdot; \mathcal{D}(u_1, F_1))$ is the influence function of $\mathcal{D}(\cdot, F_1)$ given by

$$\phi_D(x_1; \mathcal{D}(u_1, F_1)) = \frac{\mathcal{O}(u_1, F_1) \phi_s(x_1; s(F_1))}{s(F_1) (1 + \mathcal{O}(u_1, F_1))^2}$$

with $\phi_s(\cdot; s(F_1))$ being the influence function of $s(F_1)$. For N_{3n} , we similarly have

$$\begin{aligned}
N_{3n} &= \sqrt{n} \int_{U(F_1)}^{U(\hat{F}_1)} g(u_1) W(D(u_1, F_1)) dF_1(u_1) \\
&\quad - \sqrt{n} \int_{L(F_1)}^{L(\hat{F}_1)} g(u_1) W(D(u_1, F_1)) dF_1(u_1) + o_p(1) \\
&= \sqrt{n} \int \int_{U(F_1)}^{U(\hat{F}_1)} u_2^* W(D(u_1, F_1)) f(u_1, u_2) du_1 du_2 \\
&\quad - \sqrt{n} \int \int_{L(F_1)}^{L(\hat{F}_1)} u_2^* W(D(u_1, F_1)) f(u_1, u_2) du_1 du_2 + o_p(1) \\
&= \int u_2^* \sqrt{n} \left\{ U(\hat{F}_1) - U(F_1) \right\} W(D(U(F_1), F_1)) f(U(F_1), u_2) du_2 \\
&\quad - \int u_2^* \sqrt{n} \left\{ L(\hat{F}_1) - L(F_1) \right\} W(D(L(F_1), F_1)) f(L(F_1), u_2) du_2 + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_{3i} + o_p(1),
\end{aligned}$$

where

$$\begin{aligned}
\eta_{3i} &= W(\tau) \phi_U(m' e_{i,t}; U(F_1)) \int u_2^* f(U(F_1), u_2) du_2 \\
&\quad - W(\tau) \phi_L(m' e_{i,t}; L(F_1)) \int u_2^* f(L(F_1), u_2) du_2
\end{aligned}$$

since $D(L(F_1), F_1) = D(U(F_1), F_1) = \tau$. Note that the influence functions of $U(F_1)$ and $L(F_1)$ are given as

$$\begin{aligned}
\phi_U(x_1; U(F_1)) &= ((1-\tau)/\tau) \phi_s(x_1; s(F_1)), \\
\phi_L(x_1; L(F_1)) &= -((1-\tau)/\tau) \phi_s(x_1; s(F_1)).
\end{aligned}$$

We can similarly verify $\int_{L(\hat{F}_1)}^{L(\hat{F}_1)} W(D(u_1, \hat{F}_1)) d\hat{F}_1(u_1) = \int_{L(F_1)}^{U(F_1)} W(D(u_1, F_1)) dF_1(u_1)$

$+o_p(1)$ in the denominator, and the desired result follows by combining the expressions η_{1i} , η_{2i} , and η_{3i} above. *Q.E.D.*

CHAPTER 10

IDENTIFICATION OF BELIEFS IN THE PRESENCE OF DISASTER RISK AND MISSPECIFICATION

Saraswata Chaudhuri^a, Eric Renault^b
and Oscar Wahlstrom^c

^a*Department of Economics, McGill University, Montreal, Canada*

^b*Department of Economics, University of Warwick, Coventry, United Kingdom*

^c*Department of Economics, Brown University, Providence, United States*

ABSTRACT

The authors discuss the econometric underpinnings of Barro (2006)'s defense of the rare disaster model as a way to bring back an asset pricing model "into the right ballpark for explaining the equity-premium and related asset-market puzzles." Arbitrarily low-probability economic disasters can restore the validity of model-implied moment conditions only if the amplitude of disasters may be arbitrary large in due proportion. The authors prove an impossibility theorem that in case of potentially unbounded disasters, there is no such thing as a population empirical likelihood (EL)-based model-implied probability distribution. That is, one cannot identify some belief distortions for which the EL-based implied probabilities in sample, as computed by Julliard and Ghosh (2012), could be a consistent estimator. This may lead to consider alternative statistical discrepancy measures to avoid the problem with EL. Indeed, the authors prove that, under sufficient integrability conditions, power divergence Cressie-Read measures with positive power coefficients properly define a unique population model-implied probability measure. However, when this

computation is useful because the reference asset pricing model is misspecified, each power divergence will deliver different model-implied beliefs distortion. One way to provide economic underpinnings to the choice of a particular belief distortion is to see it as the endogenous result of investor's choice when optimizing a recursive multiple-priors utility a la Chen and Epstein (2002). Jeong et al. (2015)'s econometric study confirms that this way of accommodating ambiguity aversion may help to address the Equity Premium puzzle.

Keywords: Asset pricing; beliefs distortion; disaster risk; misspecification; empirical likelihood; equity premium puzzle

1. INTRODUCTION

The absence of arbitrage opportunities implies the existence of a stochastic discount factor (SDF), such that the equilibrium price of a traded security can be represented as the conditional expectation of the future payoff discounted by the SDF. Thus, a typical asset pricing equation is:

$$E[SR - e_n | I] = 0$$

where R denotes an n -dimensional vector of gross returns corresponding to payoffs on financial assets over some investment horizon, S denotes the corresponding SDF for this horizon, and I stands for the information set of the representative investor. e_n is a n -dimensional vector of ones.

Hereafter, the interval $[t, t+1]$ is the period of the investment between date t (today) and date $(t+1)$ where gross returns, denoted by R_{t+1} , are observed. Besides the horizon time $(t+1)$, the SDF S_{t+1} may depend on unknown parameters θ , giving rise to the set of conditional moment restrictions:

$$E[g(X_{t+1}, \theta) | I(t)] = 0, \theta \in \Theta \subset \mathbb{R}^p \quad (1)$$

where the function $g(X_{t+1}, \theta) = S_{t+1}(\theta)R_{t+1} - e_n$ captures the parameter dependence of the SDF $S_{t+1}(\theta)$ along with random variables X_{t+1} observed by the econometrician and used to construct the payoffs, prices, and the SDF.

The unknown parameters θ typically describe the preferences of the representative investor and are identified by the fact that for a true unknown value θ^0 of the parameters, $g(X_{t+1}, \theta^0)$ should be a martingale difference sequence. However, standard asset pricing models lead to the so-called equity premium puzzle (EPP; see Mehra & Prescott, 1985), that is the failure of the representative agent model to fit historical averages of the equity premium and the risk free rate.

An alternative view put forward in our current paper is that part of the premium can be accommodated only by considering in (1) a distortion of the expectation operator that reflects the lack of investor confidence in the assignment of probabilities to future events. We set the focus on models that capture this

departure from rational expectations by one of the two following channels: distorted subjective beliefs or ambiguity aversion.

As clearly discussed by [Chen and Epstein \(2002\)](#) (see their Section 1.2), there is some observational equivalence between models with aversion for ambiguity captured by a multiple-priors recursive utility (and a minimax type of value function) and models of belief distortion which only relax the rational expectations hypothesis that the agent knows the true probability distribution. As a matter of fact, with the recursive multiple-priors utility, the investor's optimization ultimately delivers a distorted probability measure selected endogenously from the agent's set of priors.

This latter remark suggests an econometric procedure to look for a distorted probability distribution that is minimally distorted with respect to the Data Generating Process (DGP) while bringing a plausible solution to the EPP. This econometric issue has been addressed by [Jeong et al. \(2015\)](#) for the ambiguity aversion approach and by [Ghosh et al. \(2021\)](#) for the distorted subjective beliefs. In both cases, the goal is to state the asymptotic theory of a minimum distance approach to estimate both preference parameters and distorted beliefs.

Before developing the econometric methodology, it is worth understanding why we expect that these estimation procedures will deliver estimators of preference parameters and belief distortions that will improve upon the solution of the EPP.

In the case of a model of ambiguity aversion (the so-called κ -Ignorance model of [Chen and Epstein \(2002\)](#), [Jeong et al. \(2015\)](#)) end up with a three factor CAPM. The risk premium of a given asset is determined not only by the covariance of its return with consumption growth and with aggregate wealth (as in [Epstein & Zin, 1989](#)'s recursive utility model) but also by covariance with the density generator that defines the multiple-priors recursive utility model. The latter covariance adds some significant ambiguity compensation to the traditional risk premium, allowing the risk premium to be consistent with more realistic levels of relative risk aversion.

As far as the subjective beliefs distortion is concerned, [Barro \(2006\)](#) has revisited [Rietz \(1988\)](#)'s original way to address the EPP by bringing in low-probability economic disasters. According to [Barro \(2006\)](#), "the major reason for scepticism about Rietz's argument is the belief that it depends on counterfactually high probabilities and sizes of economic disasters." This has been the main motivation of [Barro \(2006\)](#)'s empirical analysis focused on "the measurement of the frequency and size of economic disasters that occurred during the twentieth century." The goal is to "calibrate the model using the observed probability distribution for economic disasters in the twentieth century" and the conclusion is that "the model's solution gets into the right ballpark for explaining the equity-premium and related asset market puzzles." Following this observation, [Julliard and Ghosh \(2012\)](#) have promoted an information theoretic empirical strategy that could reconcile a given asset pricing model with the observed data, and the asymptotic theory of this strategy has been settled by [Ghosh et al. \(2021\)](#).

In light of the above discussion, the contribution of our current chapter is threefold.

First, we provide some mathematical arguments to confirm the rather general validity of [Julliard and Ghosh \(2012\)](#)'s empirical observation that the estimated belief distortion makes economic sense because

a priori, we would expect that the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy (.) this is exactly what the estimated (probabilities) do.

Second, besides economic sense, we ask whether the estimated belief distortion also makes statistical sense because it may consistently estimate a well-defined population probability distribution. Since “higher weights to a few particular bad states” seems to give some support to the disaster risk theory as advocated by [Barro \(2006\)](#) (“counterfactually high probabilities and sizes of economic disasters”), one would expect to see positive probabilities assigned to unbounded disaster events. Unfortunately, we point out that when one does not maintain the rather restrictive assumption that the possible disasters are of bounded amplitude, a population distorted belief defined by minimization of a population statistical divergence function may not exist. In particular, we show that with a natural scheme of unbounded disasters, the population distorted beliefs do not exist when estimated by maximum EL, as in [Julliard and Ghosh \(2012\)](#). Of course, the non-existence issue is at stake only when there is a need to distort subjective beliefs because the historical probability distribution does not satisfy the moment conditions (1).

Finally, we prove that sufficient conditions for existence of distorted subjective population beliefs fulfilling moment restrictions (1) are given either by assuming that all possible disasters are of bounded amplitude, or by using a discrepancy function that, by contrast with EL, is an increasing convex function.

1.1. Relation to the Existing Literature

There is a vast literature on disaster risk and its implications for empirical asset pricing (see [Tsai & Wachter, 2015](#) and the references therein). In terms of link to the data, the current paper is particularly focused on the empirical results of [Julliard and Ghosh \(2012\)](#). Our goal is not to add to this important empirical evidence but rather to discuss its methodological underpinnings. As explained above, although we are able to confirm mathematically the main intuition of these authors that the subjective empirical beliefs deliver a rare events distribution that assigns relatively higher weights to a few particular bad states, we question the statistical meaning of this observation by proving an impossibility theorem.

This theorem which puts forward rather realistic circumstances in which the possibility of disasters of unbounded amplitude precludes the existence of population distorted beliefs is a minor extension of a result first proved by [Chen et al. \(2021\)](#). It confirms the problematic asymptotic behavior of the EL estimator in the presence of misspecification as documented by [Schennach \(2007\)](#). The latter paper shows that, even when postulating the existence of a pseudo-true value, there does not exist a root- T consistent estimator of it. Our impossibility theorem even stresses that a unique pseudo-true value may not exist.

By contrast, we apply [Csizar \(1995\)](#)'s “generalized projections for non-negative functions” to provide sufficient conditions for the existence of a population distorted belief solution of minimization of a general ϕ -divergence function. While boundedness of disasters amplitude is a sufficient condition (under standard regularity conditions), the boundedness assumption can be relaxed if we consider only increasing ϕ -divergence functions. Recent work by [Cerreia-Vioglio et al. \(2021\)](#) provides a first step to extend the min-max analysis under model ambiguity by also considering ϕ -divergence functions to acknowledge that the model used in decision-making is a simplified approximation. Both our impossibility theorem and our existence theorems set the focus on unconditional mean restrictions obtained by integrating out the conditional moment restrictions (1). This is conformable to the empirical strategy of [Julliard and Ghosh \(2012\)](#) and not restrictive as explained by [Hansen and Jagannathan \(1997\)](#) through the concept of actively managed portfolios (see Section 2.1 below). An alternative approach would be to refer to [Komunjer and Ragusa \(2016\)](#)'s “conditional density projections” to define directly conditional distorted subjective beliefs.

The extant literature also suggests some interesting connections to make between the ambiguity approach as developed by [Jeong et al. \(2015\)](#) and asset pricing under disaster risk as discussed above. First we note that, similarly to disasters of bounded amplitude, the multiple-priors recursive utility model with κ -ignorance only considers a bounded set of possible scenarios (κ is an upper bound for the density generator of different priors). The model identifies the true unknown value of the parameters by imposing the martingale condition for pricing error. As noted by [Jeong et al. \(2015\)](#), “the spirit of the methodology is therefore somewhat similar to the GMM estimation for the nonlinear Euler equation models,” or more generally to the minimization of ϕ -divergence subject to the conditional moment restrictions (1).

In [Jeong et al. \(2015\)](#), the main trick for estimation, following the general method of “Martingales Regressions for Conditional Mean Models” developed by [Park \(2021\)](#), is based on the theorem of Dambis, Dubins and Schwarz, that allows to convert (by a well-suited time change) any continuous martingale into Brownian motion. The actual martingale estimator is defined as a minimum distance estimator based on the discrepancy between the empirical distributions of normalized pricing errors after time change and the standard normal distribution. The boundedness assumption on the conditional mean in the κ -ignorance model makes easy the application of the martingale regression for estimation of the ambiguity model similarly to the estimation of population subjective beliefs in the case of bounded disasters. There is an obvious analogy between considering only disasters of bounded amplitude and only bounded worst case scenarios in the κ -ignorance model of ambiguity.

This connection between asset pricing with disaster risk and ambiguity aversion, as captured by multiple-priors recursive utility, paves the way for a potentially unified framework. For instance, it would be worth checking that, as in the distorted beliefs framework of [Julliard and Ghosh \(2012\)](#), the min-max approach to multiple priors also leads to distorted beliefs that make economic sense because “the rare events distribution (...) assigns relatively higher weights to a few particular bad states of the economy.”

1.2. Outline of the Chapter

Section 2 defines the so-called model-implied probabilities, that are empirical probabilities computed by the minimization of a ϕ -divergence with respect to the empirical distribution, as characterizing our empirical distorted subjective beliefs. We briefly discuss the choice of a specific ϕ -divergence function. While [Chaudhuri and Renault \(2020\)](#) had shown an asymptotic equivalence result between implied probability distributions corresponding to different ϕ -divergence functions in case of a well-specified asset pricing model (1), we show that it cannot be the case when the asset pricing model is misspecified. Therefore, it is only in the case of misspecified models that it is worth considering distorted subjective beliefs. For two of the most popular ϕ -divergence functions, EL and Euclidean Empirical Likelihood (EEL), we show that, if we assume that the distorted subjective beliefs converge in distribution toward a population distribution, then this distribution should confirm the main intuition of the rare event hypothesis, namely that “the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy.”

In Section 3, we argue that in the case of possibly unbounded risk, the model-implied empirical probability distribution cannot be safely interpreted as estimator of a meaningful population probability distribution because such a model-implied probability distribution may not exist if the asset pricing model is misspecified. The non-existence of a model-implied population probability distribution is caused by the conjunction of two effects:

- (i) Unbounded vector $g(X_{t+1}, \theta)$ of pricing errors,
- (ii) Decreasing divergence function ϕ .

We then conclude that the researcher is faced with the following vicious circle:

- (i) Either the asset pricing model is well-specified, and then the model-implied population probability distribution coincides with the naive historical distribution.
- (ii) Or the asset pricing model is misspecified, and then the model-implied population probability distribution depends on the choice of a ϕ -divergence function. Since, following [Chen et al. \(2021\)](#)’s terminology, EL is a problematic ϕ -divergence function (the associated model-implied population probability distribution may not exist), there is no such thing as a natural choice to replace the naive historical distribution by a well-suited population distribution of distorted subjective beliefs.

We prove in Section 4 two theorems of existence based on the assumption that either (i) or (ii) above does not hold, i.e., based respectively on the assumption of bounded pricing errors $g(X_{t+1}, \theta)$ or on the assumption of increasing divergence function ϕ .

Section 5 concludes and discusses possible alternative strategies because the results of the current paper lead us to share with [Chen et al. \(2021\)](#) the opinion that

“we do not see why the subjective beliefs of market participants must appear to the econometrician to have minimal divergence relative to rational expectations.”

2. WHY MODEL-IMPLIED PROBABILITIES MAY CONFIRM THE RARE EVENTS HYPOTHESIS?

We want to allow for the beliefs that are revealed by the market to differ from the rational expectations beliefs (the historical distribution) implied by infinite histories of the data, assuming that the observed process $X_t, t = 1, 2, \dots$, is strictly stationary and ergodic for the historical distribution. We will do so for any given possible value $\theta \in \Theta$ of the parameters.

2.1. Model-implied Empirical Probabilities

The information theoretic approaches to inference in moment condition models have become popular in econometrics since the seminal paper of [Imbens et al. \(1998\)](#), by applying [Corcoran \(1998\)](#)’s minimum contrast inference strategy to the Cressie-Read family of power divergences. For a given observed sample $X_{t+1}, t = 1, 2, \dots, T$, and a divergence function ϕ , we follow [Corcoran \(1998\)](#) by considering the minimization program over T -dimensional vectors $\pi_T = (\pi_{t,T})_{1 \leq t \leq T}$:

$$\min_{\pi_T \in \mathbb{R}^T} \sum_{t=1}^T \phi(T\pi_{t,T}) \quad (2)$$

subject to:

$$\sum_{t=1}^T \pi_{t,T} = 1, \sum_{t=1}^T \pi_{t,T} g(X_{t+1}, \theta) = 0 \quad (3)$$

where the function ϕ is a given strictly convex function for which $\phi(1) = 0$. The strict convexity of the function ϕ allows us to apply Jensen’s inequality to conclude that:

$$\frac{1}{T} \sum_{t=1}^T \phi(T\pi_{t,T}) \geq \phi\left[\sum_{t=1}^T \frac{1}{T} T\pi_{t,T}\right] = \phi(1) = 0$$

with a strict inequality except if $\pi_{t,T}$ is independent of t , that is if and only if:

$$\pi_{t,T} = \frac{1}{T}, \forall t = 1, \dots, T. \quad (4)$$

Therefore, the solution $\hat{\pi}_T(\theta) = (\hat{\pi}_{t,T}(\theta))_{1 \leq t \leq T}$ of (2)/(3) is given by:

$$\hat{\pi}_{t,T}(\theta) = \frac{1}{T}, \forall t = 1, \dots, T \quad (5)$$

if and only if the moment restrictions are fulfilled in sample:

$$\frac{1}{T} \sum_{t=1}^T g(X_{t+1}, \theta) = 0.$$

More generally, the vector $\hat{\pi}_T(\theta) = (\hat{\pi}_{t,T}(\theta))_{1 \leq t \leq T}$ can be interpreted as a distorted empirical probability distribution which, by construction, ensures the validity of the moment restrictions:

$$\sum_{t=1}^T \hat{\pi}_{t,T}(\theta) g(X_{t+1}, \theta) = 0 \quad (6)$$

while being the closest possible (in the sense of the divergence function ϕ) to the empirical distribution (4). Note that we interpret as an expectation operator denoted $\hat{E}_\phi[\cdot]$ the operator which associates to any numerical function $\psi(\cdot)$ the real number

$$\hat{E}_T^{\theta, \phi} [\psi(X_{t+1})] = \sum_{t=1}^T \hat{\pi}_{t,T}(\theta) \psi(X_{t+1}).$$

This interpretation may actually be an abuse of notation since no non-negativity constraint for $\hat{\pi}_{t,T}(\theta)$ is maintained in the minimization program (2)/(3). This abuse of notation will be at stake throughout the paper without explicit mention of it.

The bottom line is that the T numbers $\hat{\pi}_{t,T}(\theta), t = 1, \dots, T$ are model-implied empirical probabilities which ensure in sample the validity of the asset pricing equation (1) for a given value θ of the parameters:

$$\hat{E}_T^{\theta, \phi} [g(X_{t+1}, \theta)] = 0. \quad (7)$$

Several remarks are in order:

First, empirical expectation (7) takes into account conditional moment restrictions (1) only through its unconditional implication:

$$E_0 [g(X_{t+1}, \theta)] = 0 \quad (8)$$

where $E_0[\cdot]$ stands for the expectation operator of the true (historical) unknown distribution of the stationary process (X_t) .

We have known since Hansen and Jagannathan (1997) that this is not restrictive since the vector R_{t+1} of gross returns can include not only some primitive assets but also actively managed portfolio returns built on these primitive assets. Since the shares of investment in primitive assets to define the actively managed portfolios can include any function of the conditioning information set $I(t)$, the unconditional moment conditions (8) may arguably summarize all the conditional information provided by the asset pricing model of interest and observation of asset returns.

Second, having conditional moment restrictions (1) in the background, we do not consider the possibility to pre-average the consecutive values $g(X_{t+h}, \theta)$ for $h = \pm 1, 2, \dots, H_T$ with a convenient bandwidth H_T to take care of serial dependence in moment functions (see, e.g., Kitamura & Stutzer, 1997). While (1) tells us

that $g(X_{t+1}, \theta^0)$ is a martingale difference sequence for the true unknown value θ^0 of θ (if it exists), it is not the case for other values of $\theta \in \Theta$ and then, pre-averaging may be relevant. However, as explained later, it would not change the substance of our empirical discussions.

Third, it is precisely because $\hat{\pi}_{t,T}(\theta)$ may differ from the sample distribution (5) that some belief distortions are at stake. Revisiting the analysis of [Julliard and Ghosh \(2012\)](#) we will discuss later how these distortions can be interpreted.

Fourth, when the asset pricing model (1) is well-specified and $\theta = \theta^0$, the true value of the parameters, we expect that even the distorted empirical distribution converges weakly toward the true unknown probability distribution, meaning that for any bounded function $\psi(X_t)$, we have:

$$p \lim_{T \rightarrow \infty} \hat{E}_T^{\theta, \phi} [\psi(X_{t+1})] = p \lim_{T \rightarrow \infty} \left[\frac{1}{T} \sum_{t=1}^T \psi(X_{t+1}) \right] = E_0 [\psi(X_{t+1})].$$

Fifth, when the moment conditions (8) are not fulfilled for the given value θ , we may hope that the model-implied empirical probabilities $\hat{\pi}_{t,T}(\theta)$ still define asymptotically a population distribution, but it will be a distribution of distorted subjective beliefs, that will differ from the true one. We will write:

$$p \lim_{T \rightarrow \infty} \hat{E}_T^{\theta, \phi} [\psi(X_{t+1})] = \tilde{E}^{\theta, \phi} [\psi(X_{t+1})] \quad (9)$$

where \tilde{E}_ϕ^θ is an expectation operator different from E_0 . The discussion of existence and properties of these population distorted beliefs conformable to (9) is one of the main focuses of interest of this paper. We first discuss in the next subsection what would be a population analog of the minimization program (2)/(3).

2.2. Model-Implied Population Probability Distribution

We consider throughout this subsection a given value θ of the parameters and, taking advantage of the stationarity of the historical distribution of X_{t+1} , we simplify the notations by writing for any function ξ such that $\xi[g(X_{t+1}, \theta)]$ is integrable:

$$E_0 [\xi[g(X_{t+1}, \theta)]] = E_0 [\xi[Y(\theta)]].$$

The population analog of the minimization program (2)/(3) is written as:

$$\min_M E_0 [\phi(M[Y(\theta)])] \quad (10)$$

subject to:

$$E_0 [M[Y(\theta)]] = 1, E_0 [M(Y(\theta))Y(\theta)] = 0. \quad (11)$$

In particular, if the historical distribution of $Y(\theta)$ is characterized by a probability density function $f_Y(\cdot | \theta)$ with respect to some σ -finite measure λ , then:

$$E_0[M(Y(\theta))Y(\theta)] = E_M[Y(\theta)] = \int y f_Y^M(y|\theta) d\lambda(y)$$

where the distorted probability density function $f_Y^M(y|\theta)$ of $Y(\theta)$ is defined by its Radon-Nikodym derivative $M(y)$ with respect to the historical distribution:

$$f_Y^M(y|\theta) = M(y)f_Y(y|\theta).$$

The rationale for this minimization is an obvious implication of Jensen's inequality (jointly with $\phi(1) = 0$), telling us that on the one hand:

$$E_0[M[Y(\theta)]] = 1 \Rightarrow E_0[\phi(M[Y(\theta)])] \geq 0$$

and on the other hand when not only $E_0[M[Y(\theta)]] = 1$ but also $E_0[M(Y(\theta))Y(\theta)] = 0$, then:

$$E_0[\phi(M)] = 0 \Leftrightarrow E_0[Y(\theta)] = 0.$$

In particular, the value of the minimization program (10)/(11) is zero if and only if $E_0[Y(\theta)] = 0$, that is when the pricing model (1) is well-specified and θ is a true value of the parameters. In this case, the minimum is reached at only one (up to an almost sure equality) change of measure $M[Y(\theta)]$ which is identical to the constant 1.

Otherwise, if it exists, a solution $M_\phi^\theta[Y(\theta)]$ is a non-degenerate random variable, and when the historical probability distribution is defined by a density function $f_Y(y|\theta)$, a density function for distorted beliefs is given by:

$$f_Y^{M_\phi^\theta}(y|\theta) = M_\phi^\theta(y)f_Y^0(y|\theta).$$

In any case, when it exists, the change of measure given by $M_\phi^\theta[Y(\theta)]$ defines distorted expectations of bounded functions $\xi[Y(\theta)]$:

$$E^{\theta,\phi}[\xi[Y(\theta)]] = E_0[M_\phi^\theta[Y(\theta)]\xi[Y(\theta)]].$$

In particular, a maintained assumption is that distorted beliefs are absolutely continuous with respect to historical ones. They cannot assign positive probabilities to events that would almost surely not happen historically.

Obviously, since (10)/(11) is the population analog of (2)/(3) we expect that the solutions of the programs, when existing, should be related asymptotically. More precisely, we expect that for any bounded function $\xi[g(X_{t+1}, \theta)]$:

$$P \lim_{T \rightarrow \infty} \hat{E}_T^{\theta,\phi}[\xi[g(X_{t+1}, \theta)]] = \tilde{E}^{\theta,\phi}[\xi[g(X_{t+1}, \theta)]] = E^{\theta,\phi}[\xi[g(X_{t+1}, \theta)]]. \quad (12)$$

However, it is worth keeping in mind that while the finite sample solution $\hat{E}_T^{\theta,\phi}[\xi[g(X_{t+1}, \theta)]]$ always exists, while its limit $\tilde{E}^{\theta,\phi}[\xi[g(X_{t+1}, \theta)]]$ may exist under convenient regularity conditions, we do not know yet whether the population solution $E^{\theta,\phi}[\xi[g(X_{t+1}, \theta)]]$ may exist. As already announced, we will present in Sections 3 and 4 respectively, situations where it does not (resp. it does) exist.

2.3. The Cases of EL

EL is a ϕ -divergence corresponding to:

$$\phi_L(m) = -\log(m).$$

Owen (2001) has dubbed EEL the quadratic approximation of EL in the neighborhood of $m=1$:

$$\phi_Q(m) = \frac{m^2 - 1}{2}.$$

This approach is underpinned by the fact that with EEL, one is led to minimize $E_0[\phi_Q(m)]$, which is nothing but the first term in the Taylor series expansion of $E_0[\phi_L(m)]$ in the neighborhood of $m = 1$. The fact that $\phi_Q(m)$ is a quadratic function allows us to see the program (2)/(3) as a quadratic program subject to linear restrictions so that we get a solution in closed form that can be written (see, e.g., Chaudhuri & Renault, 2020) as a function of the first two empirical moments:

$$\hat{\pi}_{i,T}^Q(\theta) = \frac{1}{T} - \bar{Y}_T(\theta)' [V_T(Y_i(\theta))]^{-1} \frac{1}{T} [Y_i(\theta) - \bar{Y}_T(\theta)] \quad (13)$$

where:

$$Y_i(\theta) = g(X_{i+1}, \theta), \bar{Y}_T(\theta) = \frac{1}{T} \sum_{i=1}^T Y_i(\theta), \\ V_T(Y_i(\theta)) = \frac{1}{T} \sum_{i=1}^T Y_i(\theta) [Y_i(\theta) - \bar{Y}_T(\theta)]'.$$

Even though no closed form formula is available in the case of genuine EL, Chaudhuri and Renault (2020) have shown that an equation formally similar to (13) is available for the solution of the EL program:

$$\hat{\pi}_{i,T}^L(\theta) = \frac{1}{T} - \bar{Y}_T(\theta)' [V_T^{\theta,L}(Y_i(\theta))]^{-1} \hat{\pi}_{i,T}^L(\theta) Y_i(\theta) \quad (14)$$

where:

$$V_T^{\theta,L}(Y_i(\theta)) = \sum_{i=1}^T \hat{\pi}_{i,T}^L(\theta) Y_i(\theta) Y_i(\theta)'. \quad (15)$$

Obviously, (15) can be interpreted as an empirical variance but where the sample distribution (4) has been replaced by the model-implied probabilities $\hat{\pi}_{i,T}^L(\theta)$. It is worth noting that in spite of the similarities of formulas, (14) does not

deliver, by contrast with (13) for EEL implied probabilities, closed form formulas for EL implied probabilities $\hat{\pi}_{t,T}^L(\theta)$. Not only the implied probability $\hat{\pi}_{t,T}^L(\theta)$ is explicitly on the RHS of (14), but even solving for it would not give it in closed form since all the implied probabilities $\hat{\pi}_{\tau,T}^L(\theta), \tau = 1, 2, \dots, T$, are hidden within the matrix $[V_T^L(Y_t(\theta))]$.

The distorted subjective belief distributions obtained by maximization of EL and minimization of quadratic divergence EEL respectively are equivalently defined by associated expectation operators for any bounded function $\xi[Y_t(\theta)]$:

$$\hat{E}_T^{\theta,A}[\xi[Y_t(\theta)]] = \sum_{t=1}^T \hat{\pi}_{t,T}^A(\theta) \xi[Y_t(\theta)], A \in \{Q, L\}.$$

We deduce from (13) and (14) that:

$$\hat{E}_T^{\theta,Q}[\xi[Y_t(\theta)]] = \frac{1}{T} \sum_{t=1}^T \xi[Y_t(\theta)] - \bar{Y}_T(\theta)' [V_T[Y_t(\theta)]]^{-1} Cov_T[Y_t(\theta), \xi[Y_t(\theta)]] \quad (16)$$

and:

$$\begin{aligned} \hat{E}_T^{\theta,L}[\xi[Y_t(\theta)]] &= \frac{1}{T} \sum_{t=1}^T \xi[Y_t(\theta)] - \bar{Y}_T(\theta)' [V_T^{\theta,L}[Y_t(\theta)]]^{-1} Cov_T^{\theta,L}[Y_t(\theta), \xi[Y_t(\theta)]] \\ &\quad (17) \end{aligned}$$

where:

$$\begin{aligned} Cov_T[Y_t(\theta), \xi[Y_t(\theta)]] &= \frac{1}{T} \sum_{t=1}^T Y_t(\theta) \xi[Y_t(\theta)]' - \bar{Y}_T(\theta) \left(\frac{1}{T} \sum_{t=1}^T \xi[Y_t(\theta)] \right)' \\ Cov_T^{\theta,L}[Y_t(\theta), \xi[Y_t(\theta)]] &= \sum_{t=1}^T \hat{\pi}_{t,T}^L(\theta) Y_t(\theta) \xi[Y_t(\theta)]'. \end{aligned}$$

Note that the last formula for covariance is justified by the fact that, by definition:

$$E_T^{\theta,L}[Y_t(\theta)] = 0.$$

Assuming at this stage, for sake of simplified interpretation, that all the quantities considered in (12) exist for both ϕ -divergences ϕ_Q and ϕ_L , we are led to the definition of population probability distributions for which the expectation operators are defined (with obvious notations) as:

$$E^{\theta,Q}[\xi(Y_t(\theta))] = E[\xi(Y_t(\theta))] - E[Y_t(\theta)]' [Var[Y_t(\theta)]]^{-1} Cov[Y_t(\theta), \xi[Y_t(\theta)]] \quad (18)$$

and:

$$\begin{aligned} E^{\theta,L}[\xi(Y_t(\theta))] &= E[\xi(Y_t(\theta))] - E[Y_t(\theta)]' [Var^{\theta,L}[Y_t(\theta)]]^{-1} Cov^{\theta,L}[Y_t(\theta), \xi[Y_t(\theta)]] \\ &\quad (19) \end{aligned}$$

where:

$$\begin{aligned} Var^{\theta,L}[Y_t(\theta)] &= E^{\theta,L}\left[Y_t(\theta)Y_t(\theta)'\right] \\ Cov^{\theta,L}[Y_t(\theta), \xi[Y_t(\theta)]] &= E^{\theta,L}\left[Y_t(\theta)\xi[Y_t(\theta)']\right]. \end{aligned}$$

Note that by contrast with (18), (19) does not give an explicit definition of a probability distribution. The operator $E^{\theta,L}[\cdot]$ is defined as implicit solution of the equation (19), while it shows up not only on the LHS but also twice on the RHS (for the definition of $Var^{\theta,L}$ and $Cov^{\theta,L}$). Assuming that both the historical, the EEL and the EEL probability distributions are absolutely continuous with respect to the same measure on \mathbb{R}^n , we deduce from (18) and (19) that their respective density functions $f_Y(\cdot|\theta)$, $f_Y^{\theta,Q}(\cdot|\theta)$ and $f_Y^{\theta,L}(\cdot|\theta)$ are related as follows:

$$f_Y^{\theta,Q}(y|\theta) = f_Y(y|\theta) - E[Y(\theta)]' [Var[Y(\theta)]]^{-1} [y - E[Y(\theta)]] f_Y(y|\theta) \quad (20)$$

and:

$$f_Y^{\theta,L}(y|\theta) = f_Y(y|\theta) - E[Y(\theta)]' [Var^{\theta,L}[Y_t(\theta)]]^{-1} y f_Y^{\theta,L}(y|\theta). \quad (21)$$

Note that (21) defines implicitly the density function $f_Y^{\theta,L}(\cdot|\theta)$ as solution of an equation that contains it once in the LHS and twice in the RHS (one of them to compute $Var^{\theta,L}$).

2.4. The Rare Disaster Interpretation

Both, the explicit formula (20) and the implicit formula (21) allow us to find some theoretical underpinnings for the common intuition that disaster risk may help to rationalize the Equity Premium Puzzle. To follow the counterfactual analysis in [Julliard and Ghosh \(2012\)](#) means: (i) fixing the risk aversion parameter or more generally our vector θ of parameters to a “reasonable” value (relative risk aversion parameter fixed to 10 in their case), and (ii) then asking the EEL and EL “estimation procedures to identify the distribution of the data that would solve the EPP in the historical sample.” According to [Julliard and Ghosh \(2012\)](#),

the first question to ask is whether the implied state probabilities make economic sense. A priori, we would expect that the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particularly bad states of the economy. Figure 4 suggests that this is exactly what the estimated $P_j(y)$ (probabilities) do.

To address this issue, we first note that:

$$E[Y(\theta)] = Cov[S(\theta), R] + E[S(\theta)]E[R] - e_n.$$

Therefore, if for instance, for the sake of expositional simplicity, we consider that the risk free asset is properly priced:

$$E[S(\theta)] = \frac{1}{R_F}$$

where R_F stands for the risk-free return in any unit period (interest rate risk is overlooked for expositional simplicity), then we will have:

$$E[Y(\theta)] = \text{Cov}[S(\theta), R] + \frac{1}{R_F} EPR$$

where:

$$EPR = E[R] - R_F e_n$$

is the equity premium vector for the vector of n risky assets under consideration. Therefore, if the equity premium of a given asset $R_{i,t+1}$ is larger than the covariance with the counterfactual SDF $S_{t+1}(\theta)$ can explain, we will have:

$$E[Y_{i,t+1}(\theta)] > 0.$$

Let us then consider a collection of asset returns $R_{i,t+1}, i = 1, \dots, n$, some of the assets being exactly priced ($E[Y_{i,t+1}(\theta)] = 0$) while the other ones displaying a risk premium higher than the model prediction ($E[Y_{i,t+1}(\theta)] > 0$). We assume in addition that the joint precision matrix $C(\theta) = [Var[Y(\theta)]]^{-1}$ of the pricing errors $Y_{i,t+1}(\theta), i = 1, \dots, n$ has only non-negative coefficients $c_{ij}(\theta), i, j = 1, \dots, n$. Note that this is in particular true if the n pricing errors $Y_{i,t+1}(\theta), i = 1, \dots, n$ are pairwise non-correlated. By revisiting the popular concept of multivariate total positivity of order 2 (see, e.g., Chapter 2, [Joe, 1997](#)), we could say that we assume more generally a kind of multivariate total negativity of the pricing errors (see Example 2.2 and Exercise 2.19, [Joe, 1997](#)). It turns out that without this assumption, the comparison between the historical distribution density $f_Y(y|\theta)$ and the EEL distribution density $f_Y^{\theta,Q}(y|\theta)$ would be ambiguous. As shown in Example 1 below, positive dependence between different asset returns will give rise to more likely disasters through contagion, even though their likelihood is not captured by a high value of $f_Y^{\theta,Q}(y|\theta)$.

To figure out the comparison between $f_Y(y|\theta)$ and $f_Y^{\theta,Q}(y|\theta)$ in a bad state y of the economy, we first deduce from (20) that:

$$f_Y^{\theta,Q}(y|\theta) = f_Y(y|\theta) - \left\{ \sum_{1 \leq i, j \leq n} c_{ij}(\theta) E[Y_{j,t+1}(\theta)] \{y_i - E[Y_{i,t+1}(\theta)]\} \right\} f_Y(y|\theta).$$

This formula shows that if $y = (y_i)_{1 \leq i \leq n}$ is a “bad state” because:

$$y_i < E[Y_{i,t+1}(\theta)], \forall i = 1, \dots, n$$

then:

$$f_Y^{\theta,Q}(y|\theta) > f_Y(y|\theta).$$

Let us consider a simple example to further illustrate the underlying ideas. For notational brevity, we will in general suppress in this example the dependence of the concerned quantities on the given θ .

Example 1. Assume for sake of notational simplicity that $n = 2$ and let ρ stand for the correlation between the two asset pricing errors with $-1 < \rho < 1$, so that the variance matrix can be inverted. The analysis could be easily extended to more than two assets by considering instead partial correlations between two pricing errors given the other ones. Let us assume that the first asset is accurately priced while the second one displays some overly high risk premium:

$$E(Y_1) = 0, E(Y_2) = \mu_2 > 0.$$

Then, recalling that $C = [Var(Y_1, Y_2)]$, with elements denoted by c_{ij} , is the joint precision matrix of the pricing errors, we obtain that:

$$\begin{aligned} & \sum_{1 \leq i, j \leq 2} c_{ij}(\theta) E[Y_{j,t+1}(\theta)] \{y_i - E[Y_{i,t+1}(\theta)]\} \\ &= c_{12}\mu_2 y_1 + c_{22}\mu_2(y_2 - \mu_2) \\ &= \frac{\mu_2}{(1-\rho^2)Var(Y_2)} \left\{ y_2 - \mu_2 - \rho \left[\frac{Var(Y_2)}{Var(Y_1)} \right]^{1/2} y_1 \right\}. \end{aligned}$$

By abuse of notation, let us denote the affine regression of Y_2 on Y_1 as a conditional expectation. It is correct in case of joint normality and all statements below remain valid in case of affine regression. Then the above formula shows that the relative difference between the historical density $f_Y(y|\theta)$ and the EEL implied one $f_Y^{\theta,Q}(y|\theta)$ is:

$$\frac{f_Y^{\theta,Q}(y|\theta) - f_Y(y|\theta)}{f_Y(y|\theta)} = \frac{-\mu_2}{(1-\rho^2)Var(Y_2)} \{y_2 - E[Y_2|Y_1 = y_1]\}. \quad (22)$$

While $c_{22} = [(1-\rho^2)Var(Y_1)]^{-1}$ is positive by definition, the sign of $c_{12} = -\rho(1-\rho^2)^{-1} [Var(Y_1)Var(Y_2)]^{-1/2}$ will be crucial to figure out the impact on implied probabilities of a bad state, such that:

$$y_1 < E(Y_1) = 0, y_2 < E(Y_2) = \mu_2.$$

We see with (22) that the sign of the relative difference between densities is the opposite of the sign of the conditional surprise on Y_2 :

$$v_2 = y_2 - E[Y_2 | Y_1 = y_1].$$

Then, if $c_{12} \geq 0$, meaning that $\rho \leq 0$ and $v_2 \leq y_2 - \mu_2 < 0$ since:

$$E[Y_2 | Y_1 = y_1] = \mu_2 + \rho \left[\frac{Var(Y_2)}{Var((Y_1))} \right]^{1/2} y_1 \geq \mu_2.$$

By contrast, if $\rho > 0$, then $E[Y_2 | Y_1 = y_1] < E(Y_2)$ so that, even though (y_1, y_2) is a bad state, it is still possible that $v_2 = y_2 - E[Y_2 | Y_1 = y_1] > 0$ so that:

$$f_Y^{\theta, Q}(y|\theta) < f_Y(y|\theta).$$

The contagion (or correlation) effect makes the conditionally pricing error $E[Y_2 | Y_1 = y_1]$ worse than the unconditionally expected pricing error $E(Y_2)$, so that it may be that y_2 is not conditionally a bad state (it exceeds the conditional expectation given other assets). Hence, while (y_1, y_2) had been defined unconditionally as a bad state, it is not true anymore that the rare events distribution needed to rationalize the EPP assigns relatively higher weight to this state. ■

Overall, at least for EEL, it is true that, when precluding overly influential contagion effects, “the rare events distribution needed to rationalize the EPP assigns relatively higher weights to a few particular bad states of the economy.”

However, in their empirical study, [Julliard and Ghosh \(2012\)](#) do not use EEL but EL. Formula (21) shows that EL should lead to the same kind of conclusion as EEL, at least if we define “bad state” by:

$$y_i < 0, \forall i = 1, \dots, n$$

and the maintained assumption about implied precision matrix (assumption of non-negative coefficients) is about the EL implied precision matrix $[Var^{\theta, L}[Y(\theta)]]^{-1}$.

3. DISASTER RISK AND PROBLEMATIC CHARACTERIZATION OF DISTORTED BELIEFS

3.1. Unidentified Beliefs Under Disaster Risk

We show in this subsection that when an excess of equity premium (or more generally $E[Y(\theta)] \neq 0$) not explained by the asset pricing model (under historical distribution) is matched by a disaster risk, it may lead to the non-existence of a well-defined model-implied probability distribution.

To see that, we first introduce the notation $E_\lambda(Z|B)$ for any Borel set B such that $0 < \lambda(B) < +\infty$:

$$E_\lambda(Z|B) = \frac{1}{\lambda(B)} \int_B z d\lambda(z).$$

In other words, $E_\lambda(Z|B)$ is the expectation of the probability distribution on \mathbb{R}^n defined by:

$$P_\lambda(A|B) = \frac{\lambda(A \cap B)}{\lambda(B)}.$$

For instance, if λ is the Lebesgue measure on \mathbb{R}^n , $P_\lambda(\cdot|B)$ is the uniform probability distribution on B .

We maintain in this subsection the following assumption:

Assumption “Undetected Misfit” (mispricing matched by unbounded disaster risk):

For a given $\theta \in \Theta$ such that $E[Y(\theta)] \neq 0$, there exists a sequence B_j of Borel sets of \mathbb{R}^n and a sequence α_j of positive real numbers such that for all $j = 1, 2, \dots$:

$$\begin{aligned} 0 &< \lambda(B_j) < +\infty \\ E_\lambda(Z|B_j) &= -\alpha_j E[Y(\theta)], \lim_{j \rightarrow \infty} \alpha_j = +\infty \\ f_Y(y|\theta) &> 0, \forall y \in B_j. \end{aligned}$$

For instance, if λ is the Lebesgue measure on \mathbb{R}^n , B_j may be a ball with center $[-\alpha_j E[Y(\theta)]]$. We can then prove the following theorem:

Theorem 1. If ϕ is a decreasing function on \mathbb{R}^+ , continuous at $m = 1$ (with $\phi(1) = 0$), under assumption “Undetected Misfit,” there exists a sequence of changes of measure $M^{(j)}(y|\theta)$ such that for all $j = 1, 2, \dots$:

$$E[M^{(j)}(Y(\theta)|\theta)] = 1, E[M^{(j)}(Y(\theta)|\theta)Y(\theta)] = 0$$

and:

$$\lim_{j \rightarrow \infty} E\{\phi[M^{(j)}(Y(\theta)|\theta)]\} = 0.$$

The interpretation of Theorem 1 is clear. It proves that under assumption “Undetected Misfit,” with a decreasing contrast function ϕ , the minimization problem (10) subject to (11) does not admit a solution. We can find changes of measure M , fulfilling the constraints (11) and making $E[\phi(M)]$ arbitrarily close to zero but the lower bound zero cannot be reached since the pricing model is misspecified. This impossibility theorem is important since it can be applied in particular to EL that corresponds to a divergence function:

$$\phi_L(M) = -\log(M).$$

It means in particular that the empirical exercise performed by [Julliard and Ghosh \(2012\)](#), as described in Section 2 above, may not be meaningful because there would be no such thing as a model-implied population probability distribution $f_Y^{\theta,L}(y|\theta)$ for which the empirical model-implied distribution would be a consistent estimator.

It is worth looking at the proof of Theorem 1 to get convinced that, insofar as we reckon the possibility of unbounded disasters, the case of impossibility put forward by Theorem 1 is not unrealistic. This proof is a slight generalization of a proof first proposed in [Chen et al. \(2021\)](#). The trick is to define a sequence of changes of measure as follows:

$$M^{(j)}(Y(\theta)|\theta) = 1 - \pi_j + \frac{\pi_j}{\lambda(B_j)} \frac{1_{B_j}[Y(\theta)]}{f_Y(Y(\theta)|\theta)}.$$

This variable is well-defined since by assumption “Undetected Misfit”:

$$Y(\theta) \in B_j \Rightarrow f_Y(Y(\theta)|\theta) > 0.$$

By construction:

$$E[M^{(j)}(Y(\theta)|\theta)] = 1 - \pi_j + \frac{\pi_j}{\lambda(B_j)} \int_{B_j} d\lambda(y) = 1$$

while:

$$\begin{aligned} E[M^{(j)}(Y(\theta)|\theta)Y(\theta)] &= (1 - \pi_j)E[Y(\theta)] + \frac{\pi_j}{\lambda(B_j)} \int_{B_j} y d\lambda(y) \\ &= (1 - \pi_j)E[Y(\theta)] + \pi_j E_\lambda(Z|B_j) \\ &= (1 - \pi_j)E[Y(\theta)] - \pi_j \alpha_j E[Y(\theta)] \\ &= 0 \Leftrightarrow \pi_j = \frac{1}{1 + \alpha_j}. \end{aligned}$$

Thus, by applying assumption “Undetected Misfit,” we conclude that the moment matching implies that:

$$\lim_{j \rightarrow \theta} \pi_j = 0.$$

In other words, it is precisely because the disaster risk is unbounded that we have been able to match moments with a sequence of changes of measure that converge to the unit constant, which implies since the divergence measure ϕ is decreasing:

$$0 \leq E\{\phi[M^{(j)}(Y(\theta)|\theta)]\} \leq E\{\phi(1 - \pi_j)\} = \phi(1 - \pi_j)$$

which converges to zero since $\phi(1) = 0$ and ϕ is continuous at $m = 1$. QED

3.2. About the Choice of a ϕ -Divergence

We will argue that this issue of the choice of a ϕ -divergence is dramatically different depending on whether the asset pricing model is well-specified and studied at the true unknown value θ^0 of the parameters (or at a consistent estimator of θ^0) or the asset pricing model is misspecified (or studied at a calibrated value of the parameters that is not a consistent estimator of the true one).

3.2.1. Case of a Well-specified Asset Pricing Model

We assume in this subsection that the asset pricing model is well-specified and identified, such that there is a unique true unknown value θ^0 such that:

$$E[Y_i(\theta^0)] = 0.$$

Moreover, we assume that $Y_t(\theta^0)$ is a stationary martingale difference sequence. This allows us to apply the results of Chaudhuri and Renault (2020) and Antoine et al. (2007). Even though their results have been proved for i.i.d. data, we can be sure that, as usual for inference based on moment conditions, procedures that are valid for i.i.d. data remain valid when the vector of moments is, at the true unknown value of the parameters, a stationary martingale difference sequence.

For the choice of a ϕ -divergence, we set the focus on the family of power divergence functions introduced by Cressie and Read (1984) that we define as follows for any given real number γ :

$$\phi_\gamma(m) = \begin{cases} \frac{1}{\gamma(\gamma+1)}[m^{\gamma+1} - 1], & \gamma < 0 \\ \frac{1}{\gamma(\gamma+1)}[m^{\gamma+1} - m], & \gamma \geq 0 \end{cases}. \quad (23)$$

Note that we adopt the trick of Chen et al. (2021) to modify the classical definition of power divergence functions by replacing the term (-1) (used when $\gamma < 0$) by $(-m)$ (used when $\gamma \geq 0$). This change is immaterial when minimizing $E[\phi_\gamma(M)]$ subject to restriction $E(M) = 1$ but it helps to figure out the two limit cases $\gamma \rightarrow (-1)$ and $\gamma \rightarrow 0$, just by application of L'Hopital's rule to obtain two special cases of interest:

$$\begin{aligned} \lim_{\gamma \rightarrow (-1)} \phi_\gamma(m) &= -\log(m) = \phi_L(m) \\ \lim_{\gamma \rightarrow 0} \phi_\gamma(m) &= m \log(m) = \phi_E(m) \end{aligned}$$

where we recognize the negative log-likelihood $\phi_L(\cdot)$ and $\phi_E(\cdot)$ is by definition the relative entropy. For $\gamma = 1$, we get:

$$\phi_1(m) = \frac{m^2 - m}{2}$$

which, as already mentioned, leads to the same minimization program as EEL:

$$\phi_Q(m) = \frac{m^2 - 1}{2}.$$

More generally, for any real number γ , $\phi_\gamma(\cdot)$ is obviously a valid ϕ -divergence, that is a strictly convex function such that $\phi(1) = 0$.

Therefore, we can define model-implied empirical probabilities $\hat{\pi}_T^{(\gamma)}(\theta) = (\hat{\pi}_{t,T}^{(\gamma)}(\theta))_{1 \leq t \leq T}$, $\gamma \in \mathbb{R}$, as solution of the minimization program:

$$\min_{\pi_T \in \mathbb{R}^T} \sum_{t=1}^T \phi_\gamma(T\pi_{t,T})$$

subject to:

$$\sum_{t=1}^T \pi_{t,T} = 1, \sum_{t=1}^T \pi_{t,T} g(X_{t+1}, \theta) = 0.$$

By doing so, we define a vast family of model-implied probabilities indexed by $\gamma \in \mathbb{R}$, with particular cases:

$$\hat{\pi}_T^{(1)}(\theta) = \hat{\pi}_T^O(\theta), \hat{\pi}_T^{(-1)}(\theta) = \hat{\pi}_T^L(\theta).$$

[Chaudhuri and Renault \(2020\)](#) prove that implied probabilities, irrespective of the Cressie-Read divergence that one uses, are asymptotically equivalent at the convenient order, meaning that for all $t = 1, \dots, T$:

$$|\hat{\pi}_{t,T}^{(\gamma)}(\theta^0) - \hat{\pi}_{t,T}^O(\theta^0)| = o_p\left(\frac{1}{T\sqrt{T}}\right), \forall \gamma \in \mathbb{R}. \quad (24)$$

Note that we dub “convenient” the order of magnitude in the upper bound (24) because, even though it is not uniform over $t = 1, \dots, T$, it allows [Chaudhuri and Renault \(2020\)](#) to prove that it makes immaterial the choice of a discrepancy measure for estimating a population expectation. For any integrable real function $\xi[Y(\theta)]$:

$$\sqrt{T} \left\{ \sum_{t=1}^T \hat{\pi}_{t,T}^{(\gamma)}(\theta^0) \xi(Y_t(\theta^0)) - \sum_{t=1}^T \hat{\pi}_{t,T}^O(\theta^0) \xi(Y_t(\theta^0)) \right\} = o_p(1), \forall \gamma \neq 0. \quad (25)$$

As far as inference on the population expectation is concerned, (25) implies that we have the same asymptotic normal distribution for any estimator (for all γ):

$$\sqrt{T} \left\{ \sum_{t=1}^T \hat{\pi}_{t,T}^{(\gamma)}(\theta^0) \xi(Y_t(\theta^0)) - E[\xi(Y_t(\theta^0))] \right\} \xrightarrow{d} \mathcal{N}(0, \Sigma^0)$$

Since inference is not the focus of interest of this paper, we let the reader to check in [Antoine et al. \(2007\)](#) what is the value of the asymptotic variance Σ^0 and how it must be modified when, for the purpose of feasible inference, we replace θ^0 in $\sum_{t=1}^T \hat{\pi}_{t,T}^{(\gamma)}(\theta^0) \xi(Y_t(\theta^0))$ by an efficient GMM or any GEL estimator $\hat{\theta}_T$. The key point is that the information provided by the well-specified moment conditions allows to get an asymptotic variance smaller than the one of the naive estimator based on the sample mean, i.e.,

$$Var[\xi(Y_t(\theta^0))] - \Sigma^0 \text{ is positive semi-definite.}$$

It is worth realizing that, since the moment conditions are well-specified, the model-implied empirical probabilities are asymptotically matching the sample distribution (same weight $(1/T)$ for all observations $t = 1, \dots, T$), but we only have for $t = 1, \dots, T$:

$$\left| \hat{\pi}_{t,T}^{(\gamma)}(\theta^0) - \frac{1}{T} \right| = o_p\left(\frac{1}{T}\right), \forall \gamma \neq 0. \quad (26)$$

Obviously the upper bound in (26) is less tight than the one in (24), precisely because model-implied empirical probabilities are not equivalent to the naive sample frequencies: they provide asymptotically more accurate estimators of a population expectation than the naive sample mean. However, we always have consistent estimators:

$$p \lim_{T \rightarrow \infty} \sum_{t=1}^T \hat{\pi}_{t,T}^{(\gamma)}(\theta^0) \xi(Y_t(\theta^0)) = p \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \xi(Y_t(\theta^0)) = E[\xi(Y_t(\theta^0))].$$

Hence, there is no point for the purpose of economic interpretation to compute model-implied probabilities $\hat{\pi}_{t,T}^{(\gamma)}(\theta^0)$ (or feasible counterparts $\hat{\pi}_{t,T}^{(\gamma)}(\hat{\theta}_T)$). Up to statistical accuracy of estimators of population expectations, they do not provide any economically meaningful economic information that is not carried by sample frequencies.

Finally, it is worth knowing that while positivity is not granted for some of the implied probabilities, it is always possible to shrink the probabilities to preclude their possible negativity. More precisely, following [Antoine et al. \(2007\)](#), we define nonnegative shrunk probabilities as follows:

$$\hat{\pi}_{t,T}^{(\gamma),sh}(\hat{\theta}_T) = \frac{1}{1 + \varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)} \hat{\pi}_{t,T}^{(\gamma)}(\hat{\theta}_T) + \frac{\varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)}{1 + \varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)} \frac{1}{T}$$

where:

$$\varepsilon_{t,T}^{(\gamma)}(\theta) = -T \min \left[\min_{1 \leq t \leq T} \hat{\pi}_{t,T}^{(\gamma)}(\theta), 0 \right].$$

By virtue of (26), $\min_{1 \leq t \leq T} \hat{\pi}_{t,T}^{(\gamma)}(\hat{\theta}_T)$ is asymptotically nonnegative with probability one ($\varepsilon_{t,T}^{(\gamma)}(\hat{\theta}_T)$ is asymptotically nil with probability one) so that, as proved by [Antoine et al. \(2007\)](#), there is no harm for the asymptotic efficiency of estimators to shrink the model-implied empirical probabilities:

$$\sqrt{T} \left\{ \sum_{t=1}^T \hat{\pi}_{t,T}^{(\gamma),sh}(\theta^0) \xi(Y_t(\theta^0)) - \sum_{t=1}^T \hat{\pi}_{t,T}^Q(\theta^0) \xi(Y_t(\theta^0)) \right\} = o_p(1).$$

3.2.2. Case of a Misspecified Asset Pricing Model (Or Counterfactual Analysis)
Let us now consider the case of counterfactual analysis based on a given parameter value θ that is not local to the true unknown value (that may even not exist):

$$E[Y_t(\theta)] \neq 0. \quad (27)$$

While we have argued that there is no point computing model-implied empirical probabilities in the case of a well-specified model with a consistent estimator of θ^0 (since they all estimate the same object as naive sample frequencies), it is obviously different in the case of biased moments as (27). For instance following (18) and (19):

$$p \lim_{T \rightarrow \infty} \hat{E}_T^{\theta,Q} [\xi(Y_t(\theta))] = E[\xi(Y_t(\theta))] - E[Y_t(\theta)]' [Var[Y_t(\theta)]]^{-1} Cov[Y_t(\theta), \xi[Y_t(\theta)]]$$

while, if the limit exists:

$$p \lim_{T \rightarrow \infty} \hat{E}_T^{\theta,L} [\xi(Y_t(\theta))] = E[\xi(Y_t(\theta))] - E[Y_t(\theta)]' [Var^{\theta,L}[Y_t(\theta)]]^{-1} Cov^{\theta,L}[Y_t(\theta), \xi[Y_t(\theta)]].$$

We deduce from the comparison of these two formulas that it is hard to believe that there is something economically meaningful to learn from “allowing the probabilities of the states of the economy to differ from their sample frequencies” ([Julliard & Ghosh, 2012](#)).

First, there is no reason to believe that the Cressie-Read discrepancies will pick the same “pseudo-true value” of population expectations provided asymptotically by implied probabilities $\hat{\pi}_{t,T}^{(\gamma)}(\theta)$ and $\hat{\pi}_{t,T}^{(\gamma^*)}(\theta)$, $\gamma \neq \gamma^*$. The almost closed form formulas (18) and (19) clearly show the difference. Different probabilities will lead to compute modified variance and covariance which in turn imply that probabilities are different. Of course, this difference is present even asymptotically, precisely because we have (27).

Second the definite proof that implied probabilities depend (even asymptotically) on the discrepancy measure has been given by the impossibility theorem of Section 3.1. And the impossibility is precisely in the case of rare disasters whose potential impact is unbounded, thereby causing the model-implied distribution to not exist for decreasing ϕ -divergences, in particular for Cressie-Read power divergences ϕ_γ for negative powers γ and in particular for the EL case.

The bottom line is that we have no clear argument to claim that the population model-implied probability distribution associated to some particular power value γ has a better economic interpretation than others. The deep reason why there is not much economics in this discussion is that ϕ -divergences are purely statistical objects and there is no compelling argument to relate them to investor’s preferences. For instance, [Cressie and Read \(1984\)](#) have introduced power divergences for the purpose of goodness-of-fit tests. In this respect, it makes sense to consider as target of estimation not the minimal divergence beliefs (as in [Ghosh et al., 2021](#)) but rather a set of

plausible beliefs and model parameters consistent with certain levels of divergence from rational expectations (i.e. misspecification sets) and perform sensitivity analysis with respect to the level of divergence. (see [Chen et al., 2021](#))

However, it must be acknowledged that even this coherent approach focused on set identification rests upon the somewhat arbitrary choice of a ϕ -divergence, at least in the set of those that are not “problematic.” This choice issue may arguably justify the ambiguity approach in which, as in [Jeong et al. \(2015\)](#), the investor’s optimization ultimately delivers a distorted probability measure selected endogenously from the investor’s set of priors. These distorted subjective beliefs are not the result of a statistical artifact, but produced instead by investor’s concern for robustness. A similar comment applies to robust control as promoted by [Hansen and Sargent \(2011\)](#). Some additional work may be still needed to relate these results to the economic literature on disaster risk. The ultimate goal would be to reconcile “uncertainty outside and inside economic models” ([Hansen, 2014](#)).

4. SUFFICIENT CONDITIONS FOR THE EXISTENCE OF MODEL-IMPLIED POPULATION PROBABILITIES

We have seen in Section 3 that non-existence of a model-implied population probability distribution was caused by the conjunction of two effects:

- (i) Unbounded vector $Y(\theta)$ of pricing errors;
- (ii) Decreasing divergence function ϕ .

We prove in this section two results (resp. in Subsections 4.1 and 4.2) of existence based on the assumption that either (i) or (ii) does not hold, i.e., based respectively on the assumption of bounded pricing errors $Y(\theta)$ or on the assumption of increasing divergence function ϕ .

4.1. The Case of Bounded Variables

For a given value $\theta \in \Theta$, the boundedness of the n pricing errors $Y_i(\theta)$ for $i = 1, \dots, n$ allows us to consider $2n$ non-negative random variables a_i :

$$a_i(Y(\theta)) = L - Y_i(\theta), \quad a_{i+n}(Y(\theta)) = Y_i(\theta) - l, \quad \forall i = 1, \dots, n$$

where it is assumed that we have with probability one, for all $i = 1, \dots, n$:

$$l \leq Y_i(\theta) \leq L.$$

Then, for any change of measure $M(\cdot)$, we can characterize the moment restrictions $\{E[MY(\theta)] = 0\}$ by the following system of $2n$ inequalities:

$$\begin{aligned} \int a_i(y)M(y)dP_\theta(y) &\leq L, \quad \forall i = 1, \dots, n \\ \int a_i(y)M(y)dP_\theta(y) &\leq -l, \quad \forall i = n+1, \dots, 2n. \end{aligned} \tag{28}$$

In order to apply the results of [Csiszar \(1995\)](#), we will maintain the following assumption.

Assumption A1. The probability distribution P_θ of the random vector $Y(\theta)$ is absolutely continuous with respect to some σ -finite measure λ :

$$\frac{dP_\theta}{d\lambda}(y) = f_Y(y | \theta)$$

and $f_Y(y | \theta) > 0$ λ – almost everywhere. ■

Note that, the strict positivity of $f_Y(y | \theta)$ λ – almost everywhere is hardly restrictive since by definition:

$$[f_Y(y | \theta) = 0, \forall y \in B] \Rightarrow P_\theta(B) = 0$$

and then, the dominating measure λ can always be chosen such that $\lambda(B) = 0$. In this context, [Csiszar \(1995\)](#) studies the linear inverse problem (28) by looking for a probability density function $s(y)$ with respect to the measure λ solution of a minimization problem:

$$\min_s \int f_Y(y|\theta)G\left(\frac{s(y)}{f_Y(y|\theta)}\right)d\lambda(y) \quad (29)$$

subject to:

$$\int a_i(y)s(y)d\lambda(y) \leq L, \forall i = 1, \dots, n, \text{ and } \int a_i(y)s(y)d\lambda(y) \leq -l, \forall i = n+1, \dots, 2n$$

where G is a given differentiable strictly convex function on \mathbb{R}_*^+ , satisfying:

$$G(1) = G'(1) = 0.$$

The objective function of (29) is well-defined precisely because $f_Y(y|\theta) > 0$ λ -almost everywhere. Note that, if we consider a differentiable divergence function ϕ then we get a well-suited function G by considering:

$$G(u) = \phi(u) - \phi'(1)[u - 1].$$

We are then able to apply Theorem 3(i) in [Csiszar \(1995\)](#) by noting that with our notation $M(\cdot)$ for the change of measure, the program (29) can be rewritten as:

$$\min_M \int G[M(y)]dP_\theta(y)$$

subject to:

$$\begin{aligned} \int a_i(y)M(y)dP_\theta(y) &\leq L, \forall i = 1, \dots, n, \text{ and} \\ \int a_i(y)M(y)dP_\theta(y) &\leq -l, \forall i = n+1, \dots, 2n, \end{aligned}$$

which can be rewritten as:

$$\min_M E[G[M(Y(\theta))]] \text{ subject to: } E[M(Y(\theta))Y(\theta)] = 0.$$

Moreover, it is worth noting that:

$$E[G[M(Y(\theta))]] = E[\phi[M(Y(\theta))]] - \phi'(1)[E[M(Y(\theta))] - 1] = E[\phi[M(Y(\theta))]]$$

since by definition:

$$E[M(Y(\theta))] = 1.$$

In other words, the minimization problem (29) is nothing but the minimization problem of interest with the change of variable $M \mapsto s$.

Regarding the minimization problem (29), Theorem 3(i), p177, in [Csiszar \(1995\)](#) tells us that, thanks to the non-negativity of the random variables $a_i(Y_\theta)$ for all $i = 1, \dots, 2n$, and to Assumption A1, a solution s_θ to the problem (29) (the so-called D-projection problem in Csiszar's terminology) always exists. Then, from the above discussion, we do have a solution:

$$M(y|\theta) = \frac{s_\theta(y)}{f_Y(y|\theta)}, \quad \lambda - ae$$

to our problem of interest. The function s_θ is the D-projection of $f_Y(y|\theta)$ on the set of functions s defined by inequalities (28) (with $M(y)$ replaced by $s(y)/f_Y(y|\theta)$).

4.2. How to Deal with Unbounded Pricing Errors?

To figure out why the unboundedness of moment conditions may be harmful for our minimization problem and what conditions on contrast functions may provide a hedge, it is worth looking at the first order conditions of our minimization program. The Lagrangian function can be written as:

$$\begin{aligned} \mathcal{L} = & \int \phi[M(y)] f_Y(y|\theta) d\lambda(y) - a \left\{ \int M(y) f_Y(y|\theta) d\lambda(y) - 1 \right\} \\ & - b' \left\{ \int M(y) y f_Y(y|\theta) d\lambda(y) \right\} \end{aligned}$$

where $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ are the Lagrange multipliers. Then, under very general conditions with a differentiable contrast function, the first order conditions can be written (for $\lambda - a.e.$ value of y) after differentiation with respect to $M(y)$:

$$\phi'[M(y)] f_Y(y|\theta) - af_Y(y|\theta) - b'y f_Y(y|\theta) = 0. \quad (30)$$

By right-multiplying by $M(y)$ (resp. $M(y)y'$), integrating with respect to y , and using the constraints of the program, we get 1 equation (resp. n equations) to determine the Lagrange multipliers a and b respectively as:

$$\begin{aligned} a^* &= E[M(Y(\theta))\phi[M(Y(\theta))]] \\ E[M(Y(\theta))Y(\theta)Y(\theta)']b^* &= E[M(Y(\theta))Y(\theta)\phi'[M(Y(\theta))]]. \end{aligned}$$

By plugging these values of a and b in (30), we get the optimal value of $M(y)$ for all y (up to $\lambda - a.e.$ equality) by inverting the strictly increasing function ϕ' . Thanks to Assumption A1, we can deduce from (30) that:

$$\phi'[M(y)] = a^* + b^*y, \quad \lambda - a.e.$$

In particular if a solution $M(\cdot|\theta)$ exists, we will have almost surely:

$$\phi' [M(Y(\theta)|\theta)] = a^* + b^* Y(\theta). \quad (31)$$

The identity (31) displays clearly the challenge we are facing for the existence of $M(\cdot|\theta)$. If the random variable $Y(\theta)$ is not bounded, the linear function $[a^* + b^* Y(\theta)]$ is not bounded either. Since the function ϕ' is strictly increasing, the divergence of $Y(\theta)$ must be coupled with a divergence of the density function $M(\cdot|\theta)$ of the model-implied population probabilities, leading to the divergence of $\phi' [M(Y(\theta)|\theta)]$ thanks to the maintained Assumption A2.

Assumption A2:

$$\lim_{m \rightarrow \infty} \phi'(m) = +\infty. \quad \blacksquare$$

In the context of the Cressie-Read family (23) of contrasts:

$$\begin{aligned} \phi'_\gamma(m) &= \frac{m^\gamma}{\gamma}, \forall \gamma \neq 0 \\ \phi'_0(m) &= \log(m) + 1, \end{aligned}$$

we note that Assumption A2 is fulfilled if and only if $\gamma \geq 0$, while by contrast for all EL-like contrasts ($\gamma < 0$):

$$\lim_{m \rightarrow \infty} \phi'_\gamma(m) = 0.$$

While Assumption A2 appears to be necessary for the existence of $M(\cdot|\theta)$ in case of an unbounded variable $Y(\theta)$ (as confirmed by the pretty general construction in Section 3 of a counter-example for all decreasing power divergence functions), we can again use [Csiszar \(1995\)](#) to show to what extent it is sufficient.

If we want to relax the boundedness assumption about $Y(\theta)$, we can simply consider the system (28) of inequalities with arbitrary values of numbers l and L (that are not bounds anymore), for instance $l = L = 0$, and variables $a_i(Y), j = 1, \dots, 2n$, which are not assumed anymore to be non-negative. As in the former subsection, we still note the equivalence between [Csiszar \(1995\)](#)'s projection problem (29) and our problem of interest, through the change of variable $M \mapsto s$.

Regarding the minimization problem (29), Theorem 3(iii), p177, in [Csiszar \(1995\)](#) tells us that, thanks to Assumptions A1 and A2, and in spite of the fact that the functions $a_i(Y), i = 1, \dots, 2n$ may take both positive and negative values, a solution s_θ to the problem (29) always exists as soon as:

$$\int G^* [\alpha a_i^-(y)] f_Y(y|\theta) d\lambda(y) < \infty, \forall \alpha > 0, \forall i = 1, \dots, 2n$$

where:

$$a_i^-(y) = \max(0, -a_i(y))$$

and G^* denotes the convex conjugate of G :

$$G^*(v) = \sup_u [uv - G(u)].$$

As reminded in [Csizar \(1995\)](#) (see formula (3.3), p177), our Assumption A2 allows us to characterize the convex conjugate of $G(u) = \phi(u) - \phi'(1)[u-1]$ as follows:

$$G^*(v) = \int_0^v (G')^{-1}(z) dz = \int_0^v (\phi')^{-1}[z + \phi'(1)] dz.$$

With our definition:

$$a_i(Y) = -Y_i(\theta), a_{i+n}(Y) = Y_i(\theta), \forall i = 1, \dots, n,$$

we are then led to maintain the following assumption.

Assumption A3. For all $\alpha > 0$ and all $i = 1, \dots, n$:

$$E\{G^*[\alpha Y_i^+(\theta)]\} < \infty, E\{G^*[\alpha Y_i^-(\theta)]\} < \infty$$

where:

$$y^+ = \max(y, 0), y^- = \max(-y, 0)$$

$$G^*(v) = \int_0^v (\phi')^{-1}[z + \phi'(1)] dz. \blacksquare$$

Then, from the above discussion, under Assumptions A1, A2 and A3, we do have a solution:

$$M(y|\theta) = \frac{s_\theta(y)}{f_Y(y|\theta)}, \lambda - ae$$

to our problem of interest.

This result ensures very generally the existence of the model-implied population probabilities for any Cressie-Read power divergence ϕ_γ , with $\gamma \geq 0$ (as in particular EEL, $\gamma = 1$), since we can now show the following result in Lemma 1.

Lemma 1. Let us consider a Cressie-Read contrast function ϕ_γ with $\gamma \geq 0$. Then a necessary and sufficient condition for Assumption A3 with $\phi = \phi_\gamma$ is:

For $\gamma > 0$, $|Y_i(\theta)|^{\frac{\gamma+1}{\gamma}}$ is integrable for all $j = 1, \dots, n$.

For $\gamma = 0$, $Y(\theta)$ has a finite Laplace transform $E[\exp(t'Y(\theta))]$ for all $t \in \mathbb{R}^n$. \blacksquare

Not surprisingly, the smaller the index γ , the more restrictive is the integrability assumption about $Y(\theta)$ that is needed for the existence of the model-implied population probabilities. The condition for $\gamma = 0$ is tantamount to assuming the integrability at any order, which is as expected the limit case (when $\gamma \rightarrow 0$) of the

assumption needed in the case $\gamma > 0$. However, it is worth noting that the necessary and sufficient condition put forward by Lemma 1 is very natural. To see that, we first note that when using the contrast function ϕ_γ , we work with changes of measure $M \geq 0$ such that $\phi_\gamma(M)$ is integrable, meaning (with standard notations) that $M \in L^{\gamma+1}$. Thus, we want that:

$$M \in L^{\gamma+1} \Rightarrow MY_i(\theta) \in L^1, \forall i = 1, \dots, n$$

in order to be able to impose the constraint $E[MY(\theta)] = 0$. By virtue of the Holder-inequality, this assumption will be fulfilled if:

$$Y_i(\theta) \in L^q, \forall i = 1, \dots, n$$

such that:

$$\frac{1}{q} + \frac{1}{\gamma+1} = 1, \text{ that is } q = \frac{\gamma+1}{\gamma},$$

and this is exactly the condition put forward for Lemma 1. For instance, with EEL ($\gamma = 1$), we need to use changes of measure M with finite variance and the corresponding moment functions, i.e., the components of $Y(\theta)$, must have finite variance as well.

5. CONCLUSION

In this paper, we address the issue of econometric analysis of a structural dynamic model that is defined by a finite-dimensional set of unconditional moment restrictions. These restrictions are misspecified under rational expectations but valid under agent's subjective beliefs. Our point of view is more general than misspecification of the asset pricing model. It also may be motivated by the willingness to perform a counterfactual analysis for a value of preference parameters that we impose on the basis of prior knowledge (like limited value of risk aversion) while it does not match the rational expectation restrictions.

This empirical strategy has been pervasive in the extant literature on calibration of disaster risk for the purpose of asset pricing. [Julliard and Ghosh \(2012\)](#) have provided arguably the most appealing variant of this strategy by minimizing a ϕ -divergence between the historical distribution and the set of candidate distortions of subjective beliefs. However, we conclude that the theoretical underpinnings of this appealing empirical strategy are problematic for many reasons (see also [Chen et al. \(2021\)](#) for related arguments):

First, the minimally divergent belief may not even exist when disaster risk is unbounded, in particular in the case of an EL approach, which is rather worrying.

Second, even for ϕ -divergences for which a minimizer exists to get a well-defined subjective beliefs distortion, the approach is problematic since different choices of ϕ -divergences will lead to different minimally divergent beliefs. Since

maximization of EL does not work in general misspecified models, there is no such thing as a natural choice of the ϕ -divergence.

Third, we stress that ϕ -divergences have been first introduced in statistics, in particular for issues of goodness-of-fit testing but have arguably no economic interpretation. We put forward the work of [Chen and Epstein \(2002\)](#) (and [Jeong et al. \(2015\)](#) for econometric implementation) to suggest an alternative way of eliciting a subjective beliefs distortion as the endogenous result of the investor's optimization among a set of priors. Moreover, [Chen and Epstein \(2002\)](#) clearly explain why the model is such that ambiguity will "not disappear eventually as the agent learns about her environment." By eliciting "conditional one-step-ahead beliefs that are independent of history of times," the κ -ignorance model solved by [Jeong et al. \(2015\)](#) belongs to what [Chen and Epstein \(2002\)](#) dub IID ambiguity.

Fourth, even though the underlying rational expectations model comes as a set of conditional moment restrictions that ensures that pricing errors are martingale difference sequences, this property is by definition violated in the case of counterfactual analysis. Therefore, an efficient statistical approach should take into account serial dependence in the sequence of pricing errors. That would at least lead to revise the model-implied probabilities computed in the current paper, by replacing the stationary marginal variance and covariances by long run quantities (with HAC estimators) taking account the likely infinite order of moving average dynamics of pricing errors.

Fifth, as far as statistical efficiency is concerned, one should prefer to perform a conditional density projection of the distribution of pricing errors on the set of distributions characterized by the conditional moment restrictions that define the asset pricing model. While the statistical characterization of these projections has been thoroughly tackled by [Komunjer and Ragusa \(2016\)](#), this is arguably a statistical challenge without a compelling economic interpretation.

Overall, it seems to us that, in spite of its statistical appeal, the concept of ϕ -divergence may not be so appealing if one looks for an economically meaningful interpretation of the investors' beliefs regarding disaster risk. It may be more relevant to contemplate a direct matching of the observed data to the data simulated according to an asset pricing model including some concern for disaster risk. This would take an extension of Indirect Inference to misspecified models as sketched in [Dridi et al. \(2007\)](#) and developed in the case of disaster risk in the recent work of [Sonksen and Grammig \(2021\)](#). By putting forward a novel strategy to estimate and empirically assess pricing models that allow for multi-period disaster risk, the latter paper is able to increase the likelihood of concern for disaster risk because "the total contraction can pan out over subsequent quarters," avoiding to impose counterfactual severity of the disaster events.

ACKNOWLEDGMENT

We thank the coeditor Simon Lee, an anonymous referee, Alastair Hall, Lars Peter Hansen and Essie Maasoumi for their very helpful comments.

REFERENCES

- Antoine, B., Bonnal, H., & Renault, E. (2007). On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics*, 138, 461–487.
- Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121, 823–866.
- Cerreia-Vioglio, S., Hansen, L. P., Maccheroni, F., & Marinacci, M. (2021). *Making decisions under model misspecification* [Working paper]. University of Chicago.
- Chaudhuri, S., & Renault, E. (2020). Score tests in GMM: Why use implied probabilities. *Journal of Econometrics*, 219, 260–280.
- Chen, Z., & Epstein, L. (2002). Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70, 1403–1444.
- Chen, X., Hansen, L. P., & Hansen, P. G. (2021). *Robust inference for moment condition models without rational expectations* [Working paper]. University of Chicago.
- Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*, 85, 967–972.
- Cressie, N., & Read, T. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46, 440–464.
- Csiszar, I. (1995). Generalized projections for non-negative functions. *Acta Mathematica Hungarica*, 68, 161–186.
- Dridi, R., Guay, A., & Renault, E. (2007). Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics*, 136, 397–430.
- Epstein, L., & Zin, S. (1989). Substitution, risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57, 937–969.
- Ghosh, A., Otsu, T., & Roussellet, G. (2021). *Subjective beliefs estimators and their properties* [Working paper]. McGill University.
- Hansen, L. P. (2014). Nobel lecture: Uncertainty outside and inside economic models. *Journal of Political Economy*, 122, 954–987.
- Hansen, L. P., & Jagannathan, R. (1997). Assessing specification errors in stochastic discount factors models. *Journal of Finance*, 52, 557–590.
- Hansen, L. P., & Sargent, T. J. (2011). Robustness and ambiguity in continuous time. *Journal of Economic Theory*, 146, 1195–1223.
- Imbens, G. W., Spady, R. H., & Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*, 66, 333–357.
- Jeong, J., Kim, H., & Park, J. (2015). Does ambiguity matter? Estimating asset pricing models with a multiple-priors recursive utility. *Journal of Financial Economics*, 115, 361–382.
- Joe, H. (1997). *Multivariate models and dependence concepts* (Vol. 73). Chapman and Hall/CRC.
- Julliard, C., & Ghosh, A. (2012). Can rare events explain the equity premium puzzle? *The Review of Financial Studies*, 25, 3037–3076.
- Kitamura, Y., & Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65, 861–874.
- Komunjer, I., & Ragusa, G. (2016). Existence and characterization of conditional density projections. *Econometric Theory*, 32, 947–987.
- Mehra, R., & Prescott, E. C. (1985). The equity premium: A puzzle. *Journal of Monetary Economics*, 15, 145–161.
- Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall.
- Park, J. (2021). *Martingale regressions for conditional mean models in continuous time* [Working paper], Indiana University.
- Rietz, T. A. (1988). The equity risk premium: A solution. *Journal of Monetary Economics*, XXII, 117–131.
- Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35, 634–672.
- Sonksen, J., & Grammig, J. (2021). Empirical asset pricing with multi-period disaster risk: A simulation-based approach. *Journal of Econometrics*, 222, 805–832.
- Tsai, J., & Wachter, J. (2015). Disaster risk and its implications for asset pricing. *Annual Review of Financial Economics*, 7, 219–252.

CHAPTER 11

A NEW MODEL FOR AGRICULTURAL LAND-USE MODELING AND PREDICTION IN ENGLAND USING SPATIALLY HIGH-RESOLUTION DATA

Namhyun Kim^a Patrick Wongsa-art^b and
Ian J. Bateman^a

^a*University of Exeter Business School, University of Exeter, United Kingdom*
^b*Cardiff Business School, Cardiff University, Cardiff, United Kingdom*

ABSTRACT

In this chapter, the authors contribute toward building a better understanding of farmers' responses to behavioral drivers of land-use decision by establishing an alternative analytical procedure, which can overcome various drawbacks suffered by methods currently used in existing studies. Firstly, our procedure makes use of spatially high-resolution data, so that idiosyncratic effects of physical environment drivers, e.g., soil textures, can be explicitly modeled. Secondly, we address the well-known censored data problem, which often hinders a successful analysis of land-use shares. Thirdly, we incorporate spatial error dependence (SED) and heterogeneity in order to obtain efficiency gain and a more accurate formulation of variances for the parameter estimates. Finally, the authors reduce the computational burden and improve estimation accuracy by introducing an alternative generalized method of moments (GMM)-quasi maximum likelihood (QML) hybrid estimation procedure.

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 291–317

Copyright © 2023 by Namhyun Kim, Patrick Wongsa-art and Ian J. Bateman

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045B013

The authors apply the newly proposed procedure to spatially high-resolution data in England and found that, by taking these features into consideration, the authors are able to formulate conclusions about causal effects of climatic and physical environment, and environmental policy on land-use shares that differ significantly from those made based on methods that are currently used in the literature. Moreover, the authors show that our method enables derivation of a more effective predictor of the land-use shares, which is utterly useful from the policy-making point of view.

Keywords: Agro-environmental policy; land-use; multivariate Tobit; system of censored equation; spatial model; error component model

JEL Codes: C13; C21; C23; C34; Q15; Q53

1. INTRODUCTION

Land is the most critical natural assets which provides us with various fundamentals of life, from clean water and food to the natural regulation of hazards, such as flooding. In the literature, the term “land-use” is used when describing socio-economic use of land, such as agricultural, recreational, or residential use. In the current chapter, we focus on agricultural land-use, e.g., arable land and pasture land (or grassland). In many countries, the biggest land-use category is agriculture. For example, agriculture is the biggest land-use category in England at 63% compared to transport/utilities and residential at 4% and 1%, respectively ([Ministry of Housing, Communities and Local Government, 2017](#)). Hence, it is a common knowledge that agricultural land-use decision (e.g., to grow wheat or barley instead of oilseed rape) significantly affects the environment (e.g., biodiversity) and socio-economic welfare (see also [Mattison & Norris, 2005](#); [Reidsma et al., 2006](#)). Moreover, there is a prevalent suggestion that introducing changes to agricultural land-use, i.e., how agricultural land is cultivated, can help to achieve deep emission reductions and prepare for climate change. In the UK, Committee on Climate Change (CCC) proposes various changes to the way we cultivate land to help to achieve net-zero emission target ([CCC, 2018](#)). These are (a) to reduce land-use for grasslands by 26 to 36%, (b) to introduce new woodlands by 1.5 million hectares, and (c) to increase land-use for bio-energy crops, e.g., oilseed rape, by up to 1.2 million hectares. It is believed that these changes should lead to between 35% and 80% overall reductions in Metric tons of carbon dioxide equivalent by 2050.

Although the above paragraph only briefly provides an insight into the importance of agricultural land-use to the environment and socio-economic welfare, from the policy-making point of view, this is sufficient to highlight the necessity to manage how land is allocated between its alternative uses. To manage agricultural land requires a good understanding of farmers’ responses to behavioral drivers of land-use decision. Previous studies have suggested various behavioral drivers. These can be categorized into: (i) Climatic drivers, e.g., rain and temperature, (ii) Economic drivers, e.g., input/output prices, and (iii) Environmental policies

and schemes, e.g., greenbelts and environmentally sensitive areas (ESAs) in the UK. A good understanding of how these drivers influence land-use decisions over time and spatial space should help the UK government to both evaluate existing practices and formulate new environmental policies, especially after Brexit.

In the current chapter, we aim to contribute toward improving the ability to formulate a better understanding of farmers' responses to the behavioral drivers of land-use decision. We achieve this objective by establishing an analytical procedure, which can handle complex data structures and is able to overcome various methodological drawbacks suffered by existing methods, and applying such a method to investigate how the climatic, economic and policy drivers influence agricultural land-use patterns in England. In this chapter, these agricultural land-use patterns are depicted by "land-use share," which is defined hereafter as a proportion of a given plot of land used for cultivating a given crop. Moving beyond methods usually used in existing studies ([Ay et al., 2017](#); [Chakir & Le Gallo, 2013](#); [Fezzi & Bateman, 2011](#); [Marcos-Martinez et al., 2017](#), for example), our analytical framework explores various directions. These are (i) to examine use of spatially disaggregated data, (ii) to model the land-use share as a censored response, (iii) to allow for potential SED, (iv) to model unobserved heterogeneity in an error component structure, and (v) to reduce computational burden by introducing a hybrid estimation procedure.

We discuss statistical and empirical underpinnings of these proposals in Section 2.1. Incorporating these features gives rise to a system of two-limit (TL) random-effect (RE) Tobit models with SED (TL-RE-SED-Tobit hereafter) for spatially high-resolution panel data. We thoroughly explain the construction of such a system in Sections 2.2 and 2.3, while introducing a new hybrid QML/GMM estimation procedure in Section 3. We explain each of the necessary steps in detail in Sections 3.1 to 3.4. In Section 4, we apply our method to spatially high resolution data for England and formulate conclusions about causal effects of climatic and physical environment, and environmental policy on land-use shares. We note that these conclusions differ significantly from those made based on deficient methods that are currently used in the literature. Moreover, we show that our method enables derivation of a more effective predictor of the land-use shares, which is utterly useful from the policy making point of view. Section 5 draws some important conclusions. Finally, mathematical proof and other technical details are delegated to an Online Appendix.

2. SYSTEM OF TL-RE-SED TOBIT EQUATIONS

We begin with a set of analytical considerations that lead to the need to formulate the system of TL-RE-SED-Tobit.

2.1. Methodological Explorations

These are intended to address drawbacks in the empirical methods used by previous studies for agricultural land-use modeling and prediction.

2.1.1. Exploring the Use of Spatially Disaggregated Data

In an analysis of agricultural land-use, different techniques are required for different data resolutions. At the extreme ends of the spectrum, we have individual data (e.g., parcel-level data) and aggregated data on a larger geographical region (e.g., national level). Regarding the former, an analysis is often conducted within a discrete-choice modeling framework (e.g., [Li et al., 2013](#)). An analysis of the latter involves tools in panel-data regression models and seemingly unrelated regressions (e.g., [Ay et al., 2017](#); [Baltagi & Pirotte, 2011](#); [Chakir & Le Gallo, 2013](#); [Marcos-Martinez et al., 2017](#)). We believe that there are benefits to be gained by exploring spatially-disaggregated data, i.e., a convenient middle ground. In this regard, individual choices are aggregated to construct the land-use shares. But, unlike the case for national level data, aggregation is done only on a scale that is small enough to capture the spatial variation in the environmental and climatic drivers of farmers' behavior, and proportionate with the scale of the decision-making unit. The first characteristic suggests that an important benefit is the ability to explicitly model the idiosyncratic effects of policies and other physical environmental drivers (e.g., mean elevation, land slope and altitude). Furthermore, to satisfy the second characteristic, we assume that each of the spatial units considered is a decision-making unit (see also the discussion in Remark 4.2 for details).

2.1.2. Modeling Land-Use Shares as Censored Responses

A difficulty of modeling spatially disaggregated data resides in an issue often referred to as censoring problem, i.e., we are likely to see a wide range of land-use share values between 0 and 1 with pile-ups at the two endpoints. The failure to account for these features leads to numerous methodological shortfalls, especially the biasness and inconsistency of the parameter estimates (see e.g., [Greene, 2008](#); [Wooldridge, 2010](#)). In this chapter, we address the problem by modeling land-use share equations, which are based on farmers' profit maximization, as a system of simultaneous Tobit equations. Hence, we have drawn upon a set of tools recently developed for estimating censored household demand systems (see e.g., [Dong et al., 2004](#); [Yen et al., 2003](#)). These are explained in detail in Section 3 below.

2.1.3. Allowing for Potential SED

Often the use of the spatially disaggregated data involves some degree of spatial dependence. This may be brought about by endogenous interaction effects, which indicates a spatial lag specification, or by the Durbin effects, which is an exogenous interaction counterpart. Nonetheless, these effects seem to be secondary in the context of a land-use share model. A more relevant type of dependence is the SED. Measurement errors that spill across grid boundaries, for example, can lead to the SED. Otherwise, there may exist unobservable latent variables that might be unaccounted for in the model. For instance, some specific land characteristics, which cannot be accounted for in the model due to unavailability of the data, may lead to the SED if they are spatially correlated (see also [Chakir & Le Gallo, 2013](#); [Moscone et al., 2007](#)).

When the problem is not properly addressed, usual maximum likelihood and QML methods can be severely affected. In the current chapter, we first construct a panel-data Tobit model with error components, which allow both spatial and time-wise correlations. This model forms the basis for the development of our system of land-use shares (see Sections 2.2 and 2.3 for details).

2.1.4. Modeling Unobserved Heterogeneity in an Error Component Structure

In an econometric point of view, heterogeneity can be handled via either a RE or a fixed-effect model. The choice between these two alternatives is complex and depends on the model and data. In a spatial setting, using individual fixed effects might induce an incidental parameter problem as the asymptotics in the cross-sectional dimension is necessary. Some researchers, e.g., [Lee and Yu \(2010\)](#), suggested methods to overcome this problem. However, none of these papers deals with a system of equations with inter-equation correlation as in our case. Moreover, in a fixed-effect model, land quality, which is a time-invariant variable, is swept away by the within estimator and the associated coefficient is not identified. Unlike previous studies, the current paper explores the use of spatially disaggregated data. An important advantage for the use of the disaggregated data is the ability to explicitly model idiosyncratic effects of policies and other physical environmental drivers (e.g., land slope/altitude). In the other words, by using the spatially disaggregated data, we can almost completely capture the heterogeneity of spatial units, whose existence is due to the differences in geographical conditions of land. However, since data limitations can hinder a complete assessment of the influence of inter-regional biophysical and socio-economic differences on land-use dynamics, here we model such leftover individual-effects via a RE model. Taking into consideration the above discussion, an additional assumption that the unobserved variables are uncorrelated with the regressors seems to be less problematic than opting for a fixed-effect model.

2.1.5. Reducing the Computational Burden via a Hybrid Estimation Procedure

In the literature, the SED is often modeled on the basis of one of many variants of the [Cliff and Ord \(1973, 1981\)](#) formulations. An estimation of the Cliff-Ord specifications can be computationally burdensome. This is so even for spatial panel data models of uncensored responses (e.g., [Kapoor et al., 2007](#); [Yang, 2013](#)). To lighten the computational burden, [Liu and Yang \(2015\)](#) suggested an alternative QML procedure that involves concentrating out a subset of the parameters and maximizing a concentrated log-likelihood function. Nonetheless, it is not straightforward to apply such a tool to our case of censored responses. Hence, in this chapter, we formulate a hybrid method that is a combination of the QML and the GMM techniques for estimating our system of simultaneous Tobit equations of land-use shares. Even though [Kelejian and Prucha \(1999\)](#) and [Kapoor et al. \(2007\)](#) have presented some key asymptotic results for the GMM procedure, in Section 3, we discuss additional properties that are crucial to the statistical validity of our hybrid framework.

2.2. Constructing the TL-RE-SED-Tobit Model for Panel Data

The censoring problem explained in Section 2.1.2 suggests that we model land-use shares based on the TL Tobit model of the form

$$y_{k,it}^* = x_{k,it}\beta_k + u_{k,it} \quad (2.1)$$

$$y_{k,it} = \begin{cases} 0 & \text{if } y_{k,it}^* \leq 0 \\ y_{k,it}^* & \text{if } 0 < y_{k,it}^* < 1, \\ 1 & \text{if } y_{k,it}^* \geq 1 \end{cases} \quad (2.2)$$

where k signifies the k th alternative crop grown on the land, e.g., arable, $x_{k,it} = [1, x_{k,2,it}, \dots, x_{k,J,it}]$, J denotes the number of land-use determinants included in the model, i and t signify the i th grid of land and t th time period, respectively. Let $k = 1, \dots, K$, $i = 1, \dots, N$ and $t = 1, \dots, T$. The models in (2.1) is well founded since it can be viewed as a reduced form of a well-known structural profit-maximization problem discussed in Chambers and Just (1989), and extended to the context of agricultural land-use by Fezzi and Bateman (2011). A general form of the model is obtained by replacing 0 and 1 in (2.2) with a and b , where $a, b \in \mathbb{R}$ and $a < b$.

We now incorporate the RE-SED component into the TL-Tobit model by specifying the disturbance process in each time period as following the first-order spatial autoregressive (SAR) process

$$u_k(t) = \rho_k W_k u_k(t) + \varepsilon_k(t), \quad (2.3)$$

where $u_k(t) = (u_{k,1t}, u_{k,2t}, \dots, u_{k,Nt})^\top$ (i.e., an $N \times 1$ vector of disturbances), ρ_k is a scalar autoregressive parameter, $\varepsilon_k(t)$ is an $N \times 1$ vector of innovations in period t , W_k is an $N \times N$ weighting matrix of known constants. We also assume that innovation vector $\varepsilon_k(t)$ follows the error component structure

$$\varepsilon_k(t) = \mu_k + v_k(t), \quad (2.4)$$

where μ_k denotes a vector of the unit-specific error component, which suggests that the disturbances are auto-correlated both spatially and time-wise.

With regard to the above model and specifications, we maintain the following assumptions throughout this chapter.

Assumption 2.1. (a) Let T be a fixed positive integer. (b) For all $1 \leq t \leq T$ and $1 \leq i \leq N$, where $N \geq 1$, $v_{k,it}$ are identically and independently distributed (iid) with zero mean, variance of $0 < \sigma_{v,y}^2 < b_v < \infty$, and finite fourth moment. Also, $E(v_{k,it} | x_{k,it}) = 0$ almost surely. (c) For all $1 \leq i \leq N$, where $N \geq 1$, the unit-specific error components $\mu_{k,i}$ are iid with zero mean, the variance of $0 < \sigma_{\mu,k}^2 < b_\mu < \infty$, and finite fourth moment. Also, $E(\mu_{k,i} | x_{k,it}) = 0$ almost surely. (d) The processes $\{v_{k,it}\}$ and $\{\mu_{k,i}\}$ are independent. \square

Assumption 2.1(a) suggests that our analysis concerns the case where T is fixed and T Assumptions 2.1(b) and (c) imply $E(\varepsilon_{k,it}) = 0$ and

$$E(\varepsilon_{k,it}\varepsilon_{k,js}) = \begin{cases} \sigma_{k,\mu}^2 + \sigma_{k,v}^2 & \text{if } i = j; t = s \\ \sigma_{k,\mu}^2 & \text{if } i = j; t \neq s \\ 0 & \text{otherwise} \end{cases} \quad (2.5)$$

In the other words, the innovations $\varepsilon_{k,it}$ are temporally correlated within a unit, but are not spatially correlated across units.

Moreover, concatenation of the innovation vector with respect to time $t = 1, \dots, T$ leads to $\varepsilon_k = (e_T \otimes I_N)\mu_k + v_k$, where \otimes denotes the Kronecker product, e_T is a $T \times 1$ vector of 1s, I_N is an identity matrix of size N , and $v_k = (v_k^T(1), v_k^T(2), \dots, v_k^T(T))^T = (v_{k,11}, v_{k,21}, \dots, v_{k,N1}, v_{k,12}, \dots, v_{k,NT})^T$. This suggests that $E(\varepsilon_k) = 0$ and covariance matrix $E(\varepsilon_k \varepsilon_k')$ of the form

$$\Omega_{k,\varepsilon} = \sigma_{k,v}^2 I_{NT} + \sigma_{k,\mu}^2 (J_T \otimes I_N) = \sigma_{k,v}^2 Q_0 + \sigma_{k,\mu}^2 Q_1, \quad (2.6)$$

where I_{NT} denotes an identity matrix of size NT , $\sigma_{k,\mu}^2 = \sigma_{k,v}^2 + T\sigma_{k,\mu}^2$, $Q_0 = \left(I_T - \frac{J_T}{T}\right) \otimes I_N$ and $Q_1 = \frac{J_T}{T} \otimes I_N$ in which $J_T = e_T e_T'$ is a $T \times T$ matrix of unit elements. In (2.6), Q_0 and Q_1 are transformation matrices often used in the error component literature (see e.g., Baltagi, 2008). These matrices are symmetric, idempotent, orthogonal to each other and satisfy the following properties: (i) $Q_0 + Q_1 = I_{NT}$, (ii) $TR(Q_0) = N(T-1)$ and $TR(Q_1) = N$, and (iii) $Q_0 Q_1 = 0$. In the light of these properties,

$$\Omega_{k,\varepsilon}^{-1} = \sigma_{k,v}^{-2} Q_0 + \sigma_{k,\mu}^{-2} Q_1 \quad \text{and} \quad \Omega_{k,\varepsilon}^{-1/2} = \sigma_{k,v}^{-1} Q_0 + \sigma_{k,\mu}^{-1} Q_1. \quad (2.7)$$

A similar concatenation to (2.3) also leads to

$$u_k = \rho_k (I_T \otimes W_k) u_k + \varepsilon_k = [I_T \otimes (I_N - \rho_k W_k)^{-1}] \varepsilon_k, \quad (2.8)$$

which we maintain the following assumptions throughout this chapter.

Assumption 2.2. (a) The matrix $I_N - \rho_k W_k$ is nonsingular. (b) $|\rho_k| < 1$. (c) All diagonal elements of W_k are zero. \square

Assumption 2.2(a) ensures that the model is closed, in the sense that it can be uniquely solved for the disturbance u_k in terms of the innovation ε_k , whereas Assumption 2.2(c) is a normalization, which implies that no unit is related in a meaningful way or being a neighbor to itself. Although the elements of W_k are assumed to be nonvarying over t , they are allowed to depend on the cross-sectional dimension N (i.e., they are allowed to form a triangular array). This corresponds to models in which the weighting matrix is row-normalized and the number of neighbors for a given unit depends on the sample size. In this respect, we also assume:

Assumption 2.3. Row and column sums of W_k and $H_k = (I_N - \rho_k W_k)^{-1}$ are bounded in absolute values by $c_w < \infty$ and $c_h < \infty$, respectively. \square

Accordingly, $E(u_k) = 0$ and covariance matrix $E(u_k u_k^T)$ is of the form

$$\Omega_{k,u} = [I_T \otimes (I_N - \rho_k W_k)^{-1}] \Omega_{k,\varepsilon} [I_T \otimes (I_N - \rho_k W_k^T)^{-1}]. \quad (2.9)$$

It is useful to also note $\Omega_{k,u}^{-1} = [I_T \otimes (I_N - \rho_k W_k^T)] \Omega_{k,\varepsilon}^{-1} [I_T \otimes (I_N - \rho_k W_k)]$. From (2.9), it is clear that the variance-covariance matrix of the disturbance vector u_k is proportional to H_k . Since this property is preserved under matrix multiplication, Assumption 2.3 implies that the row/column sums of this matrix are bounded uniformly in absolute values, which restricts the degree of cross-sectional correlation between the model disturbances.

2.3. System Variance-Covariance Structure

We now specify the system variance-covariance structure. Consider first the disturbance $u = (u_1^T, \dots, u_K^T)^T$. In accordance with the definition in (2.8), covariance matrix of the system disturbance $E(uu^T)$ is

$$\Omega_u = A\Omega_\varepsilon A^T, \quad (2.10)$$

where $A = \text{diag}(A_{11}, \dots, A_{KK})$ and $A_{kk} = I_T \otimes H_k$. In the other words, covariance matrix $E[u_k u_l^T]$ can be expressed as

$$\Omega_{kl,u} = E[\varepsilon_k \varepsilon_l^T] \{I_T \otimes H_k H_l^T\}, \quad (2.11)$$

where $H_k = B_k^{-1}$ and $B_k = (I_N - \rho_k W_k)$. Now we discuss the covariance of the innovations $\varepsilon = (\varepsilon_1^T, \dots, \varepsilon_K^T)^T$. We assume that the cross-equation correlation of the innovations is driven by

$$E \begin{pmatrix} \mu_k \\ v_k \end{pmatrix} \left(\begin{matrix} \mu_l^T & v_l^T \end{matrix} \right) = \begin{pmatrix} \sigma_{kl,\mu}^2 (J_T \otimes I_N) & 0 \\ 0 & \sigma_{kl,v}^2 I_{NT} \end{pmatrix}, \quad (2.12)$$

where $k, l = 1, \dots, K$. Accordingly, covariance matrix of the innovations, i.e., $E(\varepsilon \varepsilon^T)$, is

$$\Omega_\varepsilon = \Omega_v \otimes Q_0 + \Omega_l \otimes Q_l = [\Omega_{kl,\varepsilon}], \quad (2.13)$$

where $\Omega_\mu = [\sigma_{kl,\mu}^2]$ and $\Omega_v = [\sigma_{kl,v}^2]$ both with dimension $K \times K$, such that $\Omega_{kl,\varepsilon}$ is $E(\varepsilon_k \varepsilon_l^T)$ defined as

$$\Omega_{kl,\varepsilon} = \sigma_{kl,\mu}^2 (J_T \otimes I_N) + \sigma_{kl,v}^2 I_{NT}, \quad (2.14)$$

which is in line with (2.6), where $\sigma_{kl,v}^2 = E(v_k v_l^T)$ and $\sigma_{kl,\mu}^2 = E(\mu_k \mu_l^T)$. Alternatively,

$$\Omega_{kl,\varepsilon} = \sigma_{kl,v}^2 Q_0 + \sigma_{kl,l}^2 Q_l \quad (2.15)$$

obtained by defining $\sigma_{kl,l}^2 = \sigma_{kl,v}^2 + T \sigma_{kl,\mu}^2$. In relation to (2.7), we also write

$$\Omega_{\varepsilon}^{-1/2} = \Omega_{v}^{-1/2} \otimes Q_0 + \Omega_{l}^{-1/2} \otimes Q_l. \quad (2.16)$$

Finally, the use of (2.14) in (2.11) leads to

$$\Omega_{kl,u} = \{\sigma_{kl,l}^2 \bar{J}_T + \sigma_{kl,v}^2 (I_T - \bar{J}_T)\} \otimes H_k H_l^T, \quad (2.17)$$

where $\bar{J}_T = J_T / T$.

2.4. Other Useful Results and Transformations

We finish this section by presenting a set of results that will be useful for the discussion that follows. Firstly, let $\omega_{k,i}$ signify covariance between future and the current disturbances, $E[u_{k,i,T+\tau} u_k^T]$. Deriving $\omega_{k,i}$ in the context of the TL-RE-SED model requires first noting that $u_k(t) = B_k^{-1}(\mu_k + v_k(t))$ and $u_k = (e_T \otimes H_k) \mu_k + (I_T \otimes H_k) v_k$. In this regard,

$$\begin{aligned} E[u_k(T+\tau) u_k^T] &= E[B_k^{-1}(\mu_k + v_k(T+\tau))((e_T \otimes H_k) \mu_k + (I_T \otimes H_k) v_k)^T] \\ &= \sigma_{kk,\mu}^2 H_k (e_T^T \otimes H_k^T), \end{aligned}$$

where $\sigma_{kk,\mu}^2 = E[\mu_k \mu_k^T]$, which is $N \times TN$. As the results,

$$E[u_{k,i,T+\tau} u_k^T] = \sigma_{kk,\mu}^2 h_{k,i} (e_T^T \otimes H_k^T), \quad (2.18)$$

where $h_{k,i}$ is the i th row of $H_k = B_k^{-1}$, for an individual i at time $T+\tau$. Moreover, recall

$$\begin{aligned} \Omega_{kk,u} &= E[\varepsilon_k \varepsilon_k^T] \{I_T \otimes H_k H_k^T\} \\ &= \{\sigma_{kk,l}^2 \bar{J}_T + \sigma_{kk,v}^2 (I_T - \bar{J}_T)\} \otimes H_k H_k^T, \end{aligned} \quad (2.19)$$

which were presented previously in (2.11) and (2.17). In this regard, (2.19) and (2.18) suggest collectively that

$$\begin{aligned} \omega_{k,i}^T \Omega_{kk,u}^{-1} &= \frac{\sigma_{kk,\mu}^2}{\sigma_{kk,v}^2} h_{k,i} (e_T^T \otimes H_k^T) \left[I_T \otimes B_k B_k^T - \frac{T \sigma_{kk,\mu}^2}{\sigma_{kk,l}^2} \bar{J}_T \otimes B_k B_k^T \right] \\ &= \frac{\sigma_{kk,\mu}^2}{\sigma_{kk,l}^2} h_{k,i} (e_T^T \otimes B_k) \end{aligned} \quad (2.20)$$

since $e_T^T = e_T^T \bar{J}_T$ and $\frac{\sigma_{kk,\mu}^2}{\sigma_{kk,v}^2} - \left(\frac{\sigma_{kk,\mu}^2}{\sigma_{kk,v}^2} \times \frac{T \sigma_{kk,\mu}^2}{\sigma_{kk,l}^2} \right) = \frac{\sigma_{kk,\mu}^2}{\sigma_{kk,l}^2}$. Since $h_{k,i}$ is the i th row of $H_k = B_k^{-1}$ and $B_k^{-1} B_k = I_N$, $h_{k,i} B_k = l_{k,i}^T$, where $l_{k,i}^T$ is the i th row of I_N ,

$h_{k,i} (e_T^T \otimes B_k) = (1 \otimes h_{k,i}) (e_T^T \otimes B_k) = (e_T^T \otimes l_{k,i}^T)$, which is $(1 \times TN)$. Hence,

$$\omega_{k,i}^T \Omega_{kk,u}^{-1} = \frac{\sigma_{kk,\mu}^2}{\sigma_{kk,l}^2} (e_T^T \otimes l_{k,i}^T). \quad (2.21)$$

We shall revisit this result in Section 3.4 when we discuss prediction under the TL-RE-SED Tobit specification.

In addition, results in the previous section allow derivation of a set of transformations that are essential for the discussion in the next section. To this end, let $Y = [Y_1^T, \dots, Y_K^T]^T$, where $Y_k = [Y_k^T(1), \dots, Y_k^T(T)]^T$ and $Y_k(t) = [y_{k,1t}, \dots, y_{k,Nt}]^T$, and let $X = \text{diag}[x_1, x_2, \dots, x_K]$, where $x_k = [x_k^T(1), \dots, x_k^T(T)]^T$ and $x_k(t) = [x_{k,1t}, \dots, x_{k,Nt}]^T$. Firstly, it is the Cochrane–Orcutt-type transformation

$$\dot{X} = A^{-1}X \text{ and } \dot{Y} = A^{-1}Y. \quad (2.22)$$

Guided by the classical error component literature, we can also include the RE-GLS-type transformation to obtain the “*Cochrane-Orcutt plus RE-GLS transformations*” of the form

$$\ddot{X} = \Omega_\varepsilon^{-1/2} \dot{X} \text{ and } \ddot{Y} = \Omega_\varepsilon^{-1/2} \dot{Y}. \quad (2.23)$$

In relation to (2.22) and (2.23), let us also define

$$\dot{u} = A^{-1}u \text{ and } \ddot{u} = \Omega_\varepsilon^{-1/2} \dot{u}. \quad (2.24)$$

While the former is equivalent by definition to ε , we assume that the latter has a contemporaneous error correlation matrix $[r_{kl}]$ for $k, l = 1, \dots, K$.

3. HYBRID QML/GMM ESTIMATION PROCEDURE

In the current section, we propose a hybrid QML–GMM procedure for estimating the above–discussed system of simultaneous Tobit equations for land-use shares. Overall, our procedure consists of four key steps, namely: (3.1) Estimating the TL-Tobit panel data model of land-use shares for all the K categories. Our objective is to obtain consistent estimates of the disturbances $u_{k,u}$. (3.2) Performing a GMM estimation to obtain consistent estimates of the spatial parameters ρ_k for all $k = 1, \dots, K$ and constructing the Cochrane–Orcutt transformations \dot{Y} and \dot{X} . (3.3) Estimating system of TL-RE-SED Tobit equations of the land-use shares under consideration with QML in order to obtain estimates of the parameters $\beta = [\beta_1^T, \dots, \beta_K^T]^T$, and the elements of $\Omega_v = [\sigma_{kl,v}^2]$ and $\Omega_u = [\sigma_{kl,u}^2]$. (3.4) Constructing land-use prediction. Performing Steps 3.1 to 3.3 should provide sufficient information for analyzing causal effects of the various drivers on agricultural land-use shares. Therefore, the final step can be viewed as a supplemental step, which is useful for policy making. Below we will discuss these steps in turn.

3.1. Estimating the TL-Tobit Panel Data Models

The current step involves estimating the TL–Tobit panel data model for the k th category of land-use using QML estimation. Our objective is to obtain a consistent estimate of the disturbance, $u_{k,u}$. A number of issues must be taken into consideration to this end.

3.1.1. Heteroscedasticity

With regard to the QML estimation, it is well known that presence of heteroscedasticity is likely to lead to inconsistent estimates. However, consistent estimation is possible by specifying a model for heteroscedasticity. Particularly, let

$$\sigma_{k,u,it} = \exp(z_{k,i} \alpha_k), \quad (3.1)$$

where $z_{k,i} = [1, z_{k,1,it}, \dots, z_{k,J,it}]$ and “J” is used with a slight abuse of notation since it may not be the same as the number of determinants in (2.1). Here, we assume a multiplicative error specification as is often done in the autoregressive heteroskedasticity literature (see e.g., [Tsay, 2005](#)).

3.1.2. Pooled QML Estimation

Following the popular pooled method, pooled QML estimators maximize the quasi-log-likelihood function

$$\mathcal{L}_{k,N} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \ell_{k,it}(\bar{\beta}_k, \bar{\alpha}_k), \quad (3.2)$$

where $\ell_{k,it}(\bar{\beta}_k, \bar{\alpha}_k)$ is log-likelihood function for the it th observation, i.e.,

$$\begin{aligned} \ell_{k,it}(\bar{\beta}_k, \bar{\alpha}_k) = & 1[y_{k,it} = 0] \log[\Phi((-x_{k,it}\bar{\beta}_k)/\sigma_{k,u,it}(\bar{\alpha}_k))] \\ & + 1[0 < y_{k,it} < 1] \log[(1/\sigma_{k,u,it}(\bar{\alpha}_k))\phi((y_{k,it} - x_{k,it}\bar{\beta}_k)/\sigma_{k,u,it}(\bar{\alpha}_k))] \\ & + 1[y_{k,it} = 1] \log[\Phi(-(1-x_{k,it}\bar{\beta}_k)/\sigma_{k,u,it}(\bar{\alpha}_k))], \end{aligned}$$

and $1[\cdot]$ signifies an indicator function. Lemma 3.1 below confirms that consistent estimates of the disturbances can be obtained using the above QML estimation. The proof of this lemma requires some additional assumptions.

Assumption 3.1. (a) x_k has a full column rank, i.e. $\text{rank}(x_k) = J$, where $J < \infty$.
(b) For a column of x_k , i.e. $x_{k,l}$, $\lim_{N \rightarrow \infty} x_{k,l}^T x_{k,l} \rightarrow \infty$, $\lim_{N \rightarrow \infty} x_{k,l,it}^2 / x_{k,l}^T x_{k,l} \rightarrow 0$ and $E(x_{k,l,it}^4) < \infty$ for all $l = 1, \dots, J$ and $it = 1, 21, \dots, N1, 12, \dots, NT$. (c) The empirical distribution function $G_{k,N}$ (defined by $G_{k,N}(x_k) = j/NT$ where j is the number of points $x_{k,it} \leq x_k$) converges to a distribution function G_k for all $it = 1, 21, \dots, N1, 12, \dots, NT$ and $k = 1, \dots, K$. \square

With the exception of the finite fourth moment condition on $x_{k,l,it}$, which is necessary for the proof of Theorem 1 below, Assumption 3.1 is standard in the Tobit model literature (see e.g., [Amemiya, 1973](#)).

Lemma 3.1. Let \mathbb{B}_1 denote the vector of true parameters $(\beta_k^T, \alpha_k^T)^T$ and $\hat{\mathbb{B}}_1$ be the QML estimator of \mathbb{B}_1 . Under Assumptions 2.1 to 2.3 and 3.1, \mathbb{B}_1 is uniquely identifiable and $\hat{\mathbb{B}}_1 = \mathbb{B}_1 + O_p((NT)^{-1/2})$ as $N \rightarrow \infty$. \square

3.1.3. Standardized Residuals

Upon completion of the above estimation, the required standardized residuals for uncensored observations are constructed as

$$\tilde{u}_{k,it} / \tilde{\sigma}_{k,u,it} = (y_{k,it} - x_{k,it} \tilde{\beta}_k) / \tilde{\sigma}_{k,u,it}, \quad (3.3)$$

where $\tilde{\sigma}_{k,u,it} = \exp(z_{k,it} \tilde{\alpha}_k)$ as in equation (3.1), and $\tilde{\beta}_k$ and $\tilde{\alpha}_k$ are the QML parameter estimates from Step 3.1.2. Otherwise, generalized residuals for censored observations are computed via the inverse Mills ratio

$$\lambda_{k,it} = \phi(x_{k,it} \beta_k / \sigma_{k,u,it}) / \{\Phi(x_{k,it} \beta_k / \sigma_{k,u,it})\}, \quad (3.4)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote normal density function and the cumulative distribution function, respectively. For instance, the generalized residuals can be computed as $\tilde{u}_{k,it} = -\tilde{\lambda}_{k,it}$ for observations left-censored at 0, where $\tilde{\lambda}_{k,it}$ is obtained using (3.4) and by plugging in the parameter estimates from Section 3.1.2.

3.2. Estimating ρ_k & Constructing Cochrane–Orcutt Transformations

Now we use $\tilde{u}_k = (\tilde{u}_{k,11}, \tilde{u}_{k,21}, \dots, \tilde{u}_{k,N1}, \tilde{u}_{k,12}, \dots, \tilde{u}_{k,NT})^T$ in place of the true disturbances in order to obtain estimates for ρ_k by using the GMM procedure introduced in Kapoor et al. (2007). To be accustomed to such a practice, one only has to note that generalized residuals are commonly used for performing diagnostic tests in a standard Tobit model literature (see e.g., Cameron & Trivedi, 2005). Although the QML estimation is used in Step 3.1 unlike Kapoor et al. (2007), who employed the ordinary least squares, consistency of our GMM estimators for ρ_k , $\sigma_{k,v}^2$ and $\sigma_{k,l}^2$ can be shown in a similar fashion.

Lemma 3.2. *Let \mathbb{B}_2 denote the vector of true parameters $(\rho_k, \sigma_{k,v}^2, \sigma_{k,l}^2)^T$ and $\widehat{\mathbb{B}}_2$ is the GMM estimator of \mathbb{B}_2 . Under Assumptions 1 to 3 and 1, \mathbb{B}_2 is uniquely identifiable and $\widehat{\mathbb{B}}_2 = \mathbb{B}_2 + O_p((NT)^{-1/2})$ as $N \rightarrow \infty$.* \square

The underlying moment conditions and weight matrices for the above GMM estimators, and the proof of Lemma 3.2 are discussed in detail in the Online Appendix. Once the above estimation is completed for all K categories of land-use, GMM estimates $\hat{\rho}_1, \dots, \hat{\rho}_K$ of the autoregressive parameters are readily available. The remainder of the current step focuses on estimating the matrix A_{kk} via expression (2.10). Particularly, $\widehat{A}_{kk} = I_T \otimes \widehat{H}_k$, where $\widehat{H}_k = (I_N - \hat{\rho}_k W_k)^{-1}$. These are then used for computing the Cochrane–Orcutt transformations of Y and X , $\dot{X} = \widehat{A}^{-1} \widetilde{X}$ and $\dot{Y} = \widehat{A}^{-1} \widetilde{Y}$ where $\widetilde{Y} = [\widetilde{Y}_1^T, \widetilde{Y}_2^T, \dots, \widetilde{Y}_K^T]^T$ and $\widetilde{X} = \text{diag}[\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_K]$ are the standardized versions of Y and X with respect to $\tilde{\sigma}_{k,u,it}$, respectively.

In order to discuss the estimation in the next step, it is useful to clarify the following notational issues. Firstly, it is a rewriting of the space–time subscripts it to $i = 1, 2, \dots, NT$. Particularly let,

$$\dot{Y}_k = (\dot{y}_{k,11}, \dots, \dot{y}_{k,N1}, \dot{y}_{k,12}, \dots, \dot{y}_{k,NT})^T \equiv (\dot{y}_{k,1}, \dot{y}_{k,2}, \dots, \dot{y}_{k,NT})^T$$

and $\dot{X}_{k,i} = [\dot{x}_{k,1,i}, \dots, \dot{x}_{k,J,i}]$, which is the i th row of \dot{X}_k . Secondly, it is the notational distinction between these transformations, and

$$\dot{Y}_k = (\dot{y}_{k,1}, \dot{y}_{k,2}, \dots, \dot{y}_{k,NT})^T \text{ and } \dot{X}_{k,\iota} = [\dot{x}_{k,1,\iota}, \dots, \dot{x}_{k,J,\iota}], \quad (3.5)$$

which are the k th element of \dot{Y} and ι th row of \dot{X}_k , respectively. We recall that \dot{Y} and \dot{X}_k denote the conceptual Cochrane–Orcutt transformations in which the autoregressive parameters are known.

3.3. Estimating System of TL–RE–SED Tobit Equations

We first note an important drawback of the traditional Amemiya–Tobin mechanism. This resides in the fact that the adding-up restriction, which was discussed in Section 2, holds only for the latent equations (i.e., equation (2.1)), but not for the observed land-use shares. Here, we address such an issue by treating the K th use of land as a residual category with no specific land-use demand of its own. Therefore, the current step focuses on estimating the system of TL–RE–SED Tobit models of land-use shares for the total of $\mathcal{K}=K-1$ categorizes (see e.g., Chakir and Le Gallo, 2013; Fezzi & Bateman, 2011, who have also followed this approach).

Moreover, an estimation of a Tobit system requires evaluating multiple Gaussian integrals, which is computationally expensive when there are more than three equations. Recent studies on the estimation of consumer demand system suggested a few approaches to alleviate this problem. In this paper, we follow a suggestion made by Yen et al. (2003) and specify the likelihood function based on a sequence of bivariate Tobit likelihoods. To elaborate, let $\theta=[\beta^T, S_v^T, S_1^T, S_u^T]^T$ be vector of true parameters in a system of \mathcal{K} TL–RE–SED Tobit equations, where $\beta=[\beta_1^T, \dots, \beta_{\mathcal{K}}^T]^T$, $S_m=[\sigma_{11,m}^2, \dots, \sigma_{\mathcal{K}\mathcal{K},m}^2, \sigma_{12,m}^2, \dots, \sigma_{\mathcal{K}-1,\mathcal{K},m}^2]^T$, $S_u=[\sigma_1, \dots, \sigma_{\mathcal{K}}, r_{11}, \dots, r_{\mathcal{K},\mathcal{K}-1}]^T$, and $\mathcal{K}=K-1$. QML estimators of the vector of true parameters θ can be obtained by maximizing the quasi-likelihood

$$L = \prod_{\iota=1}^{NT} \left(L_{1,\mathcal{K},\iota} \prod_{k=2}^{\mathcal{K}} \prod_{j=1}^{k-1} L_{k,j,\iota} \right) \quad (3.6)$$

in which

$$\begin{aligned} L_{k,j,\iota} = & \left\{ \Psi(\tilde{h}_{k,\iota}, \tilde{h}_{j,\iota}; \bar{r}_{kj}) \right\}^{1[y_{k,\iota}=0, y_{j,\iota}=0]} \\ & \times \left\{ \bar{\sigma}_k^{-1} \bar{\sigma}_j^{-1} (1 - \bar{r}_{kj}^2)^{-1/2} \psi(\tilde{h}_{k,\iota}, \tilde{h}_{j,\iota}; \bar{r}_{kj}) \right\}^{1[0 < y_{k,\iota} < 1, 0 < y_{j,\iota} < 1]} \\ & \times \left\{ \Psi(-\tilde{h}_{k,\iota}, \tilde{h}_{j,\iota}; -\bar{r}_{kj}) \right\}^{1[y_{k,\iota}=1, y_{j,\iota}=0]} \\ & \times \left\{ \Psi(\tilde{h}_{k,\iota}, -\tilde{h}_{j,\iota}; \bar{r}_{kj}) \right\}^{1[y_{k,\iota}=0, y_{j,\iota}=1]} \\ & \times \left\{ \bar{\sigma}_k^{-1} \phi(\tilde{h}_{k,\iota}) \Phi[(\tilde{h}_{j,\iota} - \bar{r}_{kj} \tilde{h}_{k,\iota}) / (1 - \bar{r}_{kj}^2)^{1/2}] \right\}^{1[0 < y_{k,\iota} < 1, y_{j,\iota}=0]} \\ & \times \left\{ \bar{\sigma}_j^{-1} \phi(\tilde{h}_{j,\iota}) \Phi[(\tilde{h}_{k,\iota} - \bar{r}_{kj} \tilde{h}_{j,\iota}) / (1 - \bar{r}_{kj}^2)^{1/2}] \right\}^{1[y_{k,\iota}=0, 0 < y_{j,\iota} < 1]}, \end{aligned} \quad (3.7)$$

where $\tilde{h}_{k,\iota} = [\tilde{y}_{k,\iota} - \tilde{x}_{k,\iota} \bar{\beta}_k] / \bar{\sigma}_k$, $\tilde{h}_{j,\iota} = [\tilde{y}_{j,\iota} - \tilde{x}_{j,\iota} \bar{\beta}_j] / \bar{\sigma}_j$, $1[y_{k,\iota}=0, y_{j,\iota}=0]$ is a dichotomous indicator which equals 1 when $y_{k,\iota}=0$ and $y_{j,\iota}=0$, and $\psi(\cdot, \cdot, \cdot)$ and $\Psi(\cdot, \cdot, \cdot)$ are the bivariate standard normal probability density function and

corresponding cumulative distribution, respectively. Furthermore, $\ddot{y}_{k,t}$ and $\ddot{x}_{k,t}$ are elements of $\ddot{\mathcal{Y}} = \bar{\Omega}_\varepsilon^{-1/2} \dot{\mathcal{Y}}$ and $\ddot{\mathcal{X}} = \bar{\Omega}_\varepsilon^{-1/2} \dot{\mathcal{X}}$ in which

$$\bar{\Omega}_\varepsilon^{-1/2} = \bar{\Omega}_v^{-1/2} \otimes Q_0 + \bar{\Omega}_l^{-1/2} \otimes Q_l,$$

where $\bar{\Omega}_v^{-1/2} = [\bar{\sigma}_{kl,v}^2]$ and $\bar{\Omega}_l^{-1/2} = [\bar{\sigma}_{kl,l}^2]$.

To establish consistency of the proposed estimation requires first defining the following counterpart of (3.6)

$$L^0 = \prod_{i=1}^{NT} \left(L_{l,K,i}^0 \prod_{k=2}^K \prod_{j=1}^{k-1} L_{k,j,i}^0 \right), \quad (3.8)$$

which is constructed under an assumption that the spatial parameter ρ_k is known. In this regard,

$$\begin{aligned} L_{k,j,i}^0 &= \left\{ \Psi(h_{k,i}, h_{j,i}; \bar{r}_{kj}) \right\}^{I[y_{k,i}=0, y_{j,i}=0]} \\ &\times \left\{ \bar{\sigma}_k^{-1} \bar{\sigma}_j^{-1} (1 - \bar{r}_{kj}^2)^{-1/2} \psi(h_{k,i}, h_{j,i}; \bar{r}_{kj}) \right\}^{I[0 < y_{k,i} < 1, 0 < y_{j,i} < 1]} \\ &\times \left\{ \Psi(-h_{k,i}, h_{j,i}; -\bar{r}_{kj}) \right\}^{I[y_{k,i}=1, y_{j,i}=0]} \\ &\times \left\{ \Psi(h_{k,i}, -h_{j,i}; \bar{r}_{kj}) \right\}^{I[y_{k,i}=0, y_{j,i}=1]} \\ &\times \left\{ \bar{\sigma}_k^{-1} \phi(h_{k,i}) \Phi[(h_{k,i} - \bar{r}_{kj} h_{k,i}) / (1 - \bar{r}_{kj}^2)^{1/2}] \right\}^{I[0 < y_{k,i} < 1, y_{j,i}=0]} \\ &\times \left\{ \bar{\sigma}_j^{-1} \phi(h_{j,i}) \Phi[(h_{k,i} - \bar{r}_{kj} h_{j,i}) / (1 - \bar{r}_{kj}^2)^{1/2}] \right\}^{I[y_{k,i}=0, 0 < y_{j,i} < 1]}, \end{aligned} \quad (3.9)$$

where $h_{k,i} = [\ddot{y}_{k,i} - \ddot{x}_{k,i} \bar{\beta}_k] / \bar{\sigma}_k$ and $h_{j,i} = [\ddot{y}_{j,i} - \ddot{x}_{j,i} \bar{\beta}_j] / \bar{\sigma}_j$. Furthermore, $\ddot{y}_{k,i}$ and $\ddot{x}_{k,i}$ are elements of $\ddot{\mathcal{Y}} = \bar{\Omega}_\varepsilon^{-1/2} \dot{\mathcal{Y}}$ and $\ddot{\mathcal{X}} = \bar{\Omega}_\varepsilon^{-1/2} \dot{\mathcal{X}}$, where $\dot{\mathcal{Y}}$ and $\dot{\mathcal{X}}$ are defined in (2.22), i.e., under the assumption that the spatial parameter ρ_k is known (see also (3.5)). In this regard, establishing the consistency of the proposed QML estimation requires showing that

$$\mathcal{L}(\bar{\theta}) = \mathcal{L}^0(\bar{\theta}) + O_p((NT)^{-1/2}) \quad (3.10)$$

uniformly over a compact parameter space Θ , where $\bar{\theta} \in \Theta$, and \mathcal{L} and \mathcal{L}^0 represent $\frac{1}{NT} \ln L$ and $\frac{1}{NT} \ln L^0$, respectively. Theorem 3.1 below presents the consistency of the proposed QML estimation.

Theorem 3.1. Let $\inf_{u_{k,i}, u_{j,i} \in \mathbb{R}^2} \psi(u_{k,i}, u_{j,i}) = \delta_1$ and $\inf_{u_{k,i} \in \mathbb{R}} \phi(u_{k,i}) = \delta_2$, where $\delta_l > 0$ is an arbitrary small value for $l = 1$ or 2 . In addition, let

$$\limsup_{N \rightarrow \infty} \left\{ \max_{\bar{\theta} \in \bar{D}_\delta(\theta) \cap \Theta} E \mathcal{L}^0(\bar{\theta}) \right\} \neq \limsup_{N \rightarrow \infty} E \mathcal{L}^0(\theta)$$

for any $\bar{\theta}$, where $\bar{D}_\delta(\theta)$ is the complement of the δ -neighborhood of θ . Then, under the conditions of Lemma 3.2, θ is uniquely identified and $\hat{\theta} = \theta + O_p((NT)^{-1/2})$ as $N \rightarrow \infty$. \square

By taking into consideration the consistency of the GMM estimation, i.e., Lemma 3.2, and that presented in Theorem 3.1, the asymptotic normality and variance formula of our estimators are in line with those of a standard QML for system of Tobit equations. This is in conformity with standard results in the literature, e.g., Theorem 4 of Kapoor et al. (2007) who based their claim of asymptotic normality of their feasible GLS estimators on a similar set of consistency. To discuss asymptotic normality of a standard QML for system of Tobit equations, let us begin with Amemiya (1973) who showed such a result under mild regularity conditions for a univariate Tobit model with normal disturbances. With respect to our model, since there are no irregularities for our multivariate generalizations, Amemiya's analysis can be generalized to establish asymptotic normality of the QML estimators for the multivariate Tobit models. Such a generalization was previously discussed in e.g. Lee (1993), Wooldridge (2010), and Deng and Xue (2014). An important point to note, however, is the fact that, by specifying the likelihood function based on a sequence of bivariate Tobit likelihoods (e.g. (3.8)), the QML estimators provide the most efficient parameter estimates if and only if the quasi-likelihood function is the true likelihood function of the data. However, it is not possible to theoretically derive such loss of efficiency without imposing further assumptions on the data generating process.

Remark 3.1. *In practice, we may perform the estimation discussed in Step 3.3 by using a similar iterative steps to that in Wang and Kockelman (2007), and Baltagi and Pirotte (2011). That is to first estimate S_v and S_l conditional upon the estimate of β from Step 3.1.2. Secondly, it is to estimate β conditional upon the above estimates of S_v and S_l . These two steps are iterated until the optimal estimates of β , S_v and S_l are found.*

Remark 3.2. *An alternative way to estimate the causal parameters β and to obtain efficiency gain is to make use of the knowledge about S_v and S_l . This involves computing the Cochrane–Orcutt plus RE–GLS transformations*

$$\ddot{\mathcal{X}} = \widehat{\Omega}_\varepsilon^{-1/2} \dot{\mathcal{X}} \text{ and } \ddot{\mathcal{Y}} = \widehat{\Omega}_\varepsilon^{-1/2} \dot{\mathcal{Y}}, \quad (3.11)$$

where $\widehat{\Omega}_\varepsilon^{-1/2} = \widehat{\Omega}_v^{-1/2} \otimes Q_0 + \widehat{\Omega}_l^{-1/2} \otimes Q_l$, and performing the final equation-by-equation TL–Tobit estimation, which was explained in Section 3.1.2, by using the resulting transformations. Since we will make use of this estimation strategy in our empirical analysis in Section 4, we provide a discussion of the validity of this procedure in detail. However, such a discussion is delegated to the Online Appendix due to space limitation.

3.4. Prediction Under the TL–RE–SED Tobit Model

Previous studies in econometrics (e.g., Ay et al., 2017; Baltagi et al., 2012; Baltagi & Li, 2006; Chakir & Le Gallo, 2013) show that

$$\hat{y}_{k,i,T+\tau}^* = X_{k,i,T+\tau} \hat{\beta}_k + \hat{\omega}_{k,i}^T \hat{\Omega}_{kk,u}^{-1} \hat{u}_k \quad (3.12)$$

is the best linear unbiased predictor of the i th individual at the future period $T+\tau$. The derivation in Section 2.4 suggests that we compute

$$\hat{\omega}_{k,i}^T \hat{\Omega}_{kk,u}^{-1} = \frac{\hat{\sigma}_{kk,\mu}^2}{\hat{\sigma}_{kk,l}^2} (e_T^T \otimes I_{k,i}^T) \text{ and } \hat{\sigma}_{kk,\mu}^2 = (\hat{\sigma}_{kk,l}^2 - \hat{\sigma}_{kk,v}^2) / T. \quad (3.13)$$

Moreover, $\hat{\beta}_k$, $\hat{\sigma}_{kk,1}^2$, and $\hat{\sigma}_{kk,v}^2$ are parameter estimates obtained in Step 3.3. Using $\tilde{u}_k = (\tilde{u}_{k,11}, \tilde{u}_{k,21}, \dots, \tilde{u}_{k,N1}, \tilde{u}_{k,12}, \dots, \tilde{u}_{k,NT})^T$ from Step 3.1.3 to represent \hat{u}_k suggests that $\hat{y}_{k,i,T+\tau}^*$ is the predictor of the latent variable $y_{k,i,T+\tau}^*$, which brings about inclusion of the “*” superscript. As a result, the best linear unbiased predictor of the land-use share for the i th plot of land at the future period $T+\tau$ is computed as

$$\hat{y}_{k,i,T+\tau} = \begin{cases} 0 & \text{if } \hat{y}_{k,i,T+\tau}^* \leq 0 \\ \hat{y}_{k,i,T+\tau}^* & \text{if } 0 < \hat{y}_{k,i,T+\tau}^* < 1, \\ 1 & \text{if } \hat{y}_{k,i,T+\tau}^* \geq 1 \end{cases} .$$

Finally, it should be noted that $\hat{y}_{k,i,T+\tau}^*$ modifies the usual predictor simply by adding a fraction of the corresponding residuals to the i th unit of land. A similar result was obtained in [Baltagi and Li \(2006\)](#), [Baltagi et al. \(2012\)](#), and [Ay et al. \(2017\)](#). Here, the addition is equivalent to that of a random-effects model without the spatial autocorrelation, which deviates from the results formulated in [Baltagi and Li \(2004, 2006\)](#). This is because our SAR random effects model differs from that of [Anselin \(1988\)](#) in that the disturbance term itself follows a SAR process whereas the remainder term follows an error component structure. This point will be useful when performing the hypothesis test for comparing our model's predictive accuracy in Section 4.4.

4. EMPIRICAL ANALYSIS OF SELECTED LAND-USE SHARES IN ENGLAND

The objective of this section is to illustrate the applicability of our framework by using it to investigate farmers' responses to the behavioral drivers of land-use decision in England. We are particularly interested in investigating whether environmental schemes and grants have assisted in freeing up land used for arable, rough grazing, temporary and permanent grasslands and converting it to bio-energy crops to help to achieve deep emission reductions and prepare for climate change. We are also interested in finding out whether our predictor, which includes the above-derived fraction of the residuals to the i th unit of land, can improve accuracy for the prediction of future land-use share.

4.1. Data Descriptions and Sources

To achieve these goals, our analysis focuses on land-use shares of (i) arable, i.e., cereals (wheat, barley, and oats) and root crops (excluding oilseed rape), (ii) temporary grassland (grassland typically part of an arable crop rotation), (iii) permanent grassland (grassland maintained perpetually without reseeding), (iv) rough grazing (uncultivated land used for grazing livestock), and (v) oilseed rape.

While the first four categories are the main land-use types for the English agricultural sector, the fifth is a representation of bio-energy crops, which should be financially incentivized in order to help to reduce emissions and prepare for climate change (CCC, 2018). Moreover, we consider a set of land-use determinants and drivers, which can be classified into three categories, namely (i) economics, (ii) climatic and physical environment, and (iii) environmental policy.

Regarding the data used, they are from a unique database that consists of data compiled from various sources at the Land, Environment, Economics and Policy (LEEP) Institute. Data on agricultural land-use are derived from the June Agricultural Census on a 2- km^2 (400 ha) grid available online from Edinburgh University Data Library. These data cover England and Wales for 17 irregular spaced years between 1969 and 2006 and yield roughly 38,000 grid-square records each year. [Table 1](#) presents a full list of exogenous variables considered in our model. Due to space limitation, more details of the data are presented in the Online Appendix.

Remark 4.1. *A lack of information on the spatial variation of market input and output prices hinders an explicit modeling of their effects on land-use shares. Hence, in our empirical analysis these will be accounted for by a set of yearly and regional dummy variables (see also Fezzi et al., 2015; Sterling et al., 2013, who used a similar approach).*

Remark 4.2. *We suggested in Section 2.1.1 that aggregation of the individual data is done only on a scale that is small enough to capture the spatial variation in the environmental and climatic drivers of farmers' behavior, and proportionate with the scale of the decision-making unit. In order to satisfy the second characteristic, we shall assume that the spatial unit considered, i.e., 2- km^2 grid, represents a decision-making unit. The average farm size in the North East of England, for example, was about 1.5- km^2 (150 hectares) in 2018 (Department for Environment, Food and Rural Affairs, 2020).*

The formulation in Section 2.2 suggests that TL-RE-SED-Tobit Model only accepts balanced panel data. To satisfy such a condition, a subset of the data in the space dimension is selected by randomly extracting one grid square and then sampling every fourth grid cell along both the latitude and longitude axes. In the time dimension, since the original data cover unevenly spaced years, only observations from 1976, 1979, 1981, 1988, 2000, and 2004 are selected to stay as close to a regular time series as possible. In the other words, $T = 6$ years. For England, this leads to $NT = 10,034$ or $N = 1,729$ observations. This spatial sampling method has been used extensively in the literature (see e.g., Carrion-Flores & Irwin, 2004; Fezzi & Bateman, 2011; Nelson & Hellerstein, 1997) and should help to improve estimation performance since undesirable noises are also removed.

Finally, to compute the land-use share, we first calculate total amount of land within a 2- km^2 grid used for cultivating arable, temporary grassland, permanent grassland, rough grazing, oilseed rape, then compute land-use share of a given crop as a percentage of such a total. [Table 2](#) presents descriptive statistics for the areas of land used in hectares. The table also indicates cases in which p -values for Welch's unequal variances t -test for mean-comparison are less than 0.01, 0.05 and 0.1, respectively. These results suggest that only the area used for temporary

Table 1. Land-Use Determinants.

Abbreviations	Definitions
<i>Group 1:</i>	
<i>alt0</i>	$d_{eb200} \times elev$, where $d_{eb200} = 1$ if $elev < 200$ and 0 otherwise
<i>alt200</i>	$d_{ea200} \times elev$, where $d_{ea200} = 1$ if $elev > 200$ and 0 otherwise
<i>alt200d</i>	$alt200d = 1$ if $elev > 200$ and 0 otherwise
<i>slope6</i>	Share of each grid square with a slope higher than 6°
<i>rain</i>	Accumulated rainfall for the growing season
<i>temp</i>	Average temperature for the growing season
<i>ratemp</i>	$rain \times temp$ (i.e., an interaction term)
<i>dist300</i>	Distance to the closest major market
<i>speat</i>	Proportion of soil characteristic “Peat”
<i>sgravel</i>	Proportion of soil characteristic “Gravel”
<i>sstoney</i>	Proportion of soil characteristic “Stone”
<i>sfragipan</i>	Proportion of soil characteristic “Fragipan Soil”
<i>scoarse</i>	Proportion of soil texture “Coarse”
<i>sfine</i>	Proportion of soil texture “Fine”
<i>smedium</i>	Proportion of soil texture “Medium”
<i>sud</i>	$sud = 1$, if the grid square is located in the Southern England
<i>nor</i>	$nor = 1$, if the grid square is located in the Northern England
<i>mid</i>	$mid = 1$, if the grid square is located in the Midlands
<i>y_ℓ</i>	Yearly dummies, where $\ell = 1976, 1979, 1981, 1988, 2000, 2004$
<i>npark</i>	Share of each grid square designated as a National Park
<i>esa</i>	Share of each grid square designated as an Environmentally Sensitive Area
<i>greenbelt</i>	Share of each grid square designated as a Greenbelt
<i>setaside</i>	Share of each grid square designated as a Set-aside
<i>Group 2:</i>	
<i>rain_ℓ</i>	$rain_{\ell} = (rain - \ell)d_{r\ell}$ for $\ell = 300, 350, 400, 450, 500, 600$
<i>temp_ℓ</i>	$temp_{\ell} = (temp - \ell)d_{t\ell}$ for $\ell = 9, 10, 11, 12, 13, 14$

Notes: Since *sud*, *nor* and *mid* are summed to one, *mid* is omitted in the estimation because of multicollinearity. *smedium* is also omitted for a similar reason.

Table 2. Descriptive Statistics for Land-Uses (in ha).

	1976	1979	1981	1988	2000	2004
Temp. grassland	35.929	30.133 ^a	29.645	25.363 ^a	23.428 ^b	20.753 ^a
Perm. grassland	96.705	96.623	94.938	91.338	81.641 ^a	88.755 ^a
Rough grazing	25.772	25.274	24.995	24.341	23.622	24.948
Arable	113.255	117.333	121.042	119.114	107.460 ^a	99.035 ^a
Total arg. land ^d	272.910	271.415	274.330	269.853 ^c	245.715 ^a	248.451

Notes: ^a, ^b, and ^c signify cases where *p*-values for Welch's unequal variances *t*-test (e.g., $H_0: \mu_{k,1979} - \mu_{k,1976} = 0$ or $H_0: \mu_{k,1988} - \mu_{k,1981} = 0$) are less than 0.01, 0.05, and 0.1, respectively.

^dTotal agricultural land is computed as the summation of temporary, permanent, rough grassland and oilseed rape.

grassland has statistically significantly declined between 1976 and 2004. The level of land used for permanent grassland (arable) remained unchanged between 1976 and 1988, then fluctuated slightly (decreased steadily) between 1988 and 2004.

4.2. Empirical Specifications

This section discusses a number of empirical specifications, which are important to the analysis that follows.

Conditional mean: Regarding the empirical specifications of the conditional mean, the most basic specification is to impose linear effects on all the determinants of the land-use shares. In the other words, how the expected value of the unobserved and censored land-use share $y_{k,it}^*$ varies with the environmental, climatic and policy variables is described by

$$E[y_{k,it}^* | x_{k,it}] = x_{k,it}\beta_k, \quad (4.1)$$

where $x_{k,it}$ is an 1×29 row-vector whose elements are a constant one and the variables listed under Group 1 in Table 1. Since the specification in (4.1) can be overly restrictive, we also consider an alternative which (i) allows for some nonlinear flexibility within the parametric specification, and (ii) does so without imposing too much computational pressure. This is to capture the potential nonlinear effects of climatic factors by modeling the measures of rainfall and temperature as piecewise linear functions. In particular, how the expected value of the unobserved and censored land-use share $y_{k,it}^*$ varies with the environmental, climatic and policy variables is described by

$$E[y_{k,it}^* | x_{k,it}] = x_{k,it}\beta_k + \vartheta_k(rain_{it}) + \zeta_k(temp_{it}),$$

where $\vartheta_k(rain_{it}) = \beta_{k,r300}rain_{300,it} + \dots + \beta_{k,r600}rain_{600,it}$ and $\zeta_k(temp_{it}) = \beta_{k,t9}temp_{9,it} + \dots + \beta_{k,t14}temp_{14,it}$.

Spatial Weighting Matrices: Various studies have reported that predictive accuracy and empirical results in general are sensitive to the choice of spatial weighting matrix W_k (e.g., Anselin & Bera, 1998; Bhattacharjee & Jensen-Butler, 2006). To investigate such sensitivity, we consider weighting matrices based on two types of schemes, namely “inter-point-distance” and “graph-based-neighbors.” Particularly, we construct the κ -Nearest-Neighbors weighting matrices, $W_k^{\kappa NN}$, where either $\kappa = 2$ or $\kappa = 5$, and the Sphere-of-Influence-Neighbors weighting matrix, W_k^{SOI} . All these spatial weighting matrices are row-normalized.

Reference Land-Use Category: Note that the adding-up restriction on the land-use shares only holds for the latent shares in (2.1), but it is unsatisfiable for the observed shares. In the demand study literature, such a problem is avoided by treating one of the categories as a reference and omitting it from the system (i.e., Chakir & Le Gallo, 2013; Marcos-Martinez et al., 2017; Yen et al., 2003). In the study that follows, we drop the category “oilseed rape” and jointly estimate a system of four TL–RE–SED–Tobit models for arable, temporary grassland, permanent grassland and rough grazing for England.

4.3. Estimation Results and Important Findings

We have prepared detailed results for each of the steps discussed in Section 3. They are presented in the Online Appendix due to space limitation. In the current section, we first provide a brief description of these results, then thoroughly explain important findings from each of the estimation steps.

In the Online Appendix, Tables 3–10 present estimation results for the four land-use shares under consideration. In these tables, we present these results for four modeling strategies: (i) Without RE–SED, (ii) RE–SED under W_k^{2NN} , (iii) RE–SED under W_k^{5NN} , and (iv) RE–SED under W_k^{SOI} . In addition, for each strategy, we present the results in three columns, namely parameter estimates ($\hat{\beta}_k$), associated standard errors (SEs), and p -values (p -vals). At the bottom of the tables, we also present numbers of coefficient estimates that are statistically significant at 0.01, 0.05 and 0.1 significance levels. Furthermore, Table 11 in the Online Appendix presents results of the GMM estimation for the autoregressive parameters ρ_k , where $k = 1, \dots, \mathcal{K} = 4$. For each of the four land-use shares under consideration, we compute six estimates, i.e., three based on the linear specification under W_k^{5NN} , W_k^{2NN} and W_k^{SOI} , and the remainders based on the partial linear specification. In addition, Tables 12–14 in the Online Appendix presents the resulting estimates for Ω_v and Ω_μ denoted by $\hat{\Omega}_v$ and $\hat{\Omega}_\mu = (1/T)(\hat{\Omega}_1 - \hat{\Omega}_v)$. Based on these, we also compute correlations matrices ρ_v and ρ_μ , which gauge the cross-equation correlations in the error terms v_k and the random effects μ_k , respectively. These estimates are for the RE–SED models under W_k^{2NN} , W_k^{5NN} and W_k^{SOI} matrices, respectively, and both the linear and partial linear specifications.

4.3.1. Pooled QML Estimation of the TL–Tobit Panel Data Models

The results, which concern the pooled QML estimation of the TL–Tobit panel data models, are presented under W/O RE–SED (Without RE–SED) in Tables 3–10 in the Online Appendix. We shall discuss these numbers more thoroughly below.

4.3.2. GMM Estimation of the Spatial Parameters

In Table 11 in the Online Appendix, it is clear that at $\kappa = 5$, the corresponding estimates of ρ_k for $W_k^{\kappa NN}$ are close to those of W_k^{SOI} for all cases. The most likely reason underpinning such a phenomenon is the similarity in the degree of sparseness of these weighting matrices. Furthermore, a higher degree of sparseness in the weighting matrix is usually associated with higher estimates of ρ_k . The estimates for shares of arable and permanent grassland are statistically significant at 0.05 significance level for all cases. For the share of temporary grassland, the estimates are significant at 0.1 significance level for W_k^{2NN} and at 0.05 for both W_k^{5NN} and W_k^{SOI} . For the share of rough grazing, the estimates are significant at 0.1 level, except that of W_k^{2NN} . Moreover, these estimates enable computation of $\hat{A}_{kk} = I_T \otimes \hat{H}_k$, $\hat{H}_k = (I_N - \hat{\rho}_k W_k)^{-1}$, and the Cochrane–Orcutt transformations $\dot{\mathcal{X}} = \hat{A}^{-1} \tilde{\mathcal{X}}$ and $\dot{\mathcal{Y}} = \hat{A}^{-1} \tilde{\mathcal{Y}}$ as explained in Step 3.2.

4.3.3. Iterative QML Method

This step performs the iterative QML method discussed in Remark 3.1. Our main focus is on estimating $S_v = [\sigma_{11,v}^2, \dots, \sigma_{KK,v}^2, \sigma_{12,v}^2, \dots, \sigma_{K-1,K,v}^2]^T$ and $S_i = [\sigma_{11,i}^2, \dots, \sigma_{KK,i}^2, \sigma_{12,i}^2, \dots, \sigma_{K-1,K,i}^2]^T$, where $K = 4$. We use these estimates to construct the Cochrane–Orcutt plus RE–GLS transformations defined in (3.11), then perform the final Tobit QML estimation of the causal parameters (as discussed in Remark 3.2) in order to obtain the associated SEs. Estimation results for the four land-use shares in question are also presented in Tables 3–10 in the Online Appendix under the headings RE–SED under W_k^{2NN} , RE–SED under W_k^{5NN} and RE–SED under W_k^{SOI} . Below, let us summarize a number of key findings.

Form these tables, it is clear that taking into consideration cross-equation correlations and RE–SED reduces the number of coefficient estimates that are considered statistically significant in all cases. Let us take the share of arable as an example. The number of estimates that are significant at 0.01, 0.05 and 0.1 significance levels are 17, 20 and 21 (15, 16 and 17) under the linear (partial linear) specification and without RE–SED. These reduces to 13, 15 and 16 (11, 13 and 13) when the RE–SED is modeled under W_k^{2NN} . The increase in the degree of sparseness in the weighting matrices leads to further reduction of the figures to 10, 11 and 12 (8, 10 and 10) for modeling under W_k^{5NN} . Similar results are also obtained for modeling the RE–SED under W_k^{SOI} . The most likely reason underpinning such a phenomenon is the similarity in the degree of sparseness of these weighting matrices.

Inevitably, these lead to differences in the conclusions drawn about the causal effects of climatic and physical environment, and environmental policy on the land-use shares. We now discuss the implications of the above findings on the individual land-use shares.

Share of Arable: The coefficient estimates associated with different measures of altitude are statistically significant under the model without RE–SED, but are insignificant under RE–SED irrespective of the weighting matrices used. The coefficient estimates associated with *slope* (negative), *rain* (positive) and *temp* (positive) are statistically significant across the modeling strategies considered. However, only that associated with *sfragipan* (negative) remains statistically significant after taking into consideration RE–SED. Being in the north of England has a negative effect on the share of arable compared to Midlands. Year dummies are all statistically significant irrespective of the modeling strategies. Finally, the coefficient estimates associated with environmental policies becomes insignificant when RE–SED is modeled irrespective of the weighting matrices used.

Share of Permanent Grassland: Unlike those of arable, the coefficients estimates associated with measures of altitude are statistically significant under all the models considered for share of permanent grassland. Also unlike those for arable, the coefficient estimates associated with *slope*, *rain*, and *temp*, which are statistically significant under the model without RE–SED, become insignificant when RE–SED is taken into consideration. Although, soil characteristics remains matter for the share of permanent grassland, locations of the land (i.e., *nor* or *sud*) become insignificant after taking into consideration RE–SED. A similar conclusion can also be drawn for all the year dummies (except *y4* and *y5* which

remains statistically significant). Finally, the coefficient estimates associated with *esa* becomes statistically significant after including RE-SED.

Share of Temporary Grassland: We now shift our attention to the share of temporary grassland. The coefficient estimates associated with measures of altitude and *slope* are not statistically significant in all models. On the contrary, those associated with *rain* and *temp* are significant irrespective of the modeling strategies used. Soil characteristics seem to matter when modeling without RE-SED, but the coefficient estimates become insignificant under weighting matrices with higher degree of sparseness. A similar conclusion can also be drawn for *nor* and *sud*, and *y1* and *y2*. Finally, all the coefficient estimates associated with the environmental policies (except that of *setaside* (positive)) are not statistically significant.

Share of Rough Grazing: Regarding share of rough grazing, the coefficient estimates associated with measures of altitude are statistically insignificant, while that associated with *slope* are statistically significant irrespective of the modeling strategies considered. A similar conclusion can also be drawn for *rain* and *temp*. Nonetheless, the coefficient estimates associated with soil characteristics become insignificant after taking into consideration RE-SED. The location of the land, i.e., *nor* and *sou*, contribute positively to the share of rough grazing compared to the Midlands. Furthermore, all year dummies are statistically significant across all the model used. Finally, all coefficient estimates associated with *nspark*, *esa*, and *greenbelt* are significant when modeled without RE-SED, but only that of *greenbelt* remains significant after taking RE-SED into consideration.

We complete this section by discussing the resulting estimates for Ω_v and Ω_μ denoted by $\widehat{\Omega}_v$ and $\widehat{\Omega}_\mu = (1/T)(\widehat{\Omega}_1 - \widehat{\Omega}_v)$. Firstly, we find that the cross-equation correlations μ_k are much stronger than those of v_k . In the other words, the cross-equation correlations in the TL-Tobit system of the land-use shares are dominated by those of the random effects. Secondly, switching between the weighting matrices does not alter the signs of the estimated cross-equation correlations. However, their magnitudes change more significantly when switching from W_k^{2NN} to W_k^{5NN} than from W_k^{5NN} to W_k^{SOI} . These changes are largely dominated by those in the cross-equation correlations of μ_k . Finally, we find that the above findings hold for both the linear and partial linear specifications.

4.4. Improvement in Prediction Accuracy

In the current section, we investigate whether inclusion the above-derived fraction of the residuals to the i th unit of land is able to improve accuracy for the prediction of future land-use share. To this end, note that predictive evaluation must be performed by treating

$$\hat{y}_{k,i,T+\tau} = \begin{cases} 0 & \hat{y}_{k,i,T+\tau}^* \leq 0 \\ \hat{y}_{k,i,T+\tau}^* & \text{if } 0 < \hat{y}_{k,i,T+\tau}^* < 1, \\ 1 & \text{if } \hat{y}_{k,i,T+\tau}^* \geq 1 \end{cases}$$

as the test data set since $\hat{y}_{k,i,T+\tau}^*$ is the best linear unbiased predictor of the latent variable $y_{k,i,T+\tau}^*$. Moreover, our examination focuses on comparing root mean squared errors (RMSEs) from a number of alternative predictors. These are computed based on: (A.1) Linear TL Tobit model without the random effects and SED $\hat{y}_{k,i,T+\tau}^* = x_{k,i,T+\tau} \hat{\beta}_k$. (A.2) Partially linear TL Tobit model without the random effects and SED $\hat{y}_{k,i,T+\tau}^* = x_{k,i,T+\tau} \hat{\beta}_k + \hat{\vartheta}_k(rain_{it}) + \hat{\zeta}_k(temp_{it})$. (B.1) Linear TL Tobit model with the random effects and SED, but without the fraction of the residuals corresponding to the i th unit of land $\hat{y}_{k,i,T+\tau}^* = x_{k,i,T+\tau} \hat{\beta}_k$. (B.2) Partially linear TL Tobit model with the random effects and SED, but without the fraction of the residuals corresponding to the i th unit of land

$$\hat{y}_{k,i,T+\tau}^* = x_{k,i,T+\tau} \hat{\beta}_k + \hat{\vartheta}_k(rain_{it}) + \hat{\zeta}_k(temp_{it}).$$

(C.1) Linear TL Tobit model with the random effects and SED, and with the fraction of the residuals corresponding to the i th unit of land

$$\hat{y}_{k,i,T+\tau}^* = x_{k,i,T+\tau} \hat{\beta}_k + \frac{\hat{\sigma}_{\mu,k}^2}{\hat{\sigma}_{1,kl}^2} (e'_T \otimes l'_{k,i}) \hat{u}_k. \quad (\text{C.2}) \text{ Partially linear TL Tobit model with the random effects and SED, and with the fraction of the residuals corresponding to the } i \text{ th unit of land } \hat{y}_{k,i,T+\tau}^* = x_{k,i,T+\tau} \hat{\beta}_k + \hat{\vartheta}_k(rain_{it}) + \hat{\zeta}_k(temp_{it}) + \frac{\hat{\sigma}_{\mu,k}^2}{\hat{\sigma}_{1,kl}^2} (e'_T \otimes l'_{k,i}) \hat{u}_k.$$

In this regard, we are interested in two set of comparisons: (a) It is the comparison between the predictors listed under categories A and B. These are important because they can help to confirm the asymptotic compatibility between $\tilde{\beta}_k$ and $\hat{\beta}_k$. (b) Comparison between the predictors listed in categories B and C. These are significant since they can affirm that improvement in predictive accuracy can be achieved by including the above-derived fraction of the corresponding residuals to the i th unit of land without adjusting the causal specification. In this regard, it should also be noted that such an inclusion is not possible without incorporating the random effects and SED into the model.

Furthermore, we reinforce these results by conducting hypothesis testing for the equivalence of predictors listed under categories B and C. To this end, it is useful to recall the difference between the predictors in these categories, namely the added fraction of the corresponding residuals to the i th land. Unlike other error component models (e.g., those formulated in [Baltagi & Li, 2004, 2006](#)), the addition here is equivalent to that of a random-effect model without the spatial autocorrelation. This suggests that the absence of random effect should lead to simplification of the predictors in category C to those in B, and therefore that a testing procedure such as that of [Breusch and Pagan \(1980\)](#), which tests for the random-effects model, could be used for checking the equivalence of these predictors. [Breusch and Pagan \(1980\)](#) devised a Lagrange multiplier test for the random-effects model, in which the test statistic is

$$LM_{BP} = \frac{NT}{2(T-1)} \left[\left(\frac{\sum_{i=1}^N \left[\sum_{t=1}^T \varepsilon_{it} \right]^2}{\sum_{i=1}^N \sum_{t=1}^T \varepsilon_{it}^2} \right) - 1 \right]^2. \quad (4.2)$$

The limiting distribution of LM_{BP} is chi-squared with one degree of freedom under the null hypothesis $H_0: \sigma_{k,\mu}^2 = 0$. The practical implementation of the test relies on computation of $\hat{\varepsilon}_{k,it}$ using $\hat{\rho}_k$ and \hat{u}_k , and based either on (2.8) or $\varepsilon_k = u_k - \rho_k(I_T \otimes W_k)u_k$.

In the empirical analysis, the observed land-use shares in 2010 are treated as the validation data set. In this regard, let $\hat{y}_{k,i,2010}^*$ denote the best linear unbiased predictor of crop k 's land-use share on the i th plot of land in 2010. Hence, $\hat{y}_{k,i,2010}^*$ quantifies the expected level of crop k 's land-use share on the i th plot of land in 2010 for a given scenario of climatic, economic and policy drivers based on our estimated models and specifications. Clearly, if changes in environmental policies (e.g., an increase in farming in London's greenbelt) or climate (e.g., a higher level of rainfall and/or temperature in some regions) are expected in 2010, then $\hat{y}_{k,i,2010}^*$ is adjusted accordingly. This suggests an important benefit of land-use prediction in practice that is the ability to accurately forecast the effects of policy and/or climate changes on agricultural production and land-use in the UK.

In the Online Appendix, Tables 15–18 present the RMSEs for the out-of-sample predictions of the land-use shares in 2010. Some important findings are as follows: (i) It is clear that the RMSEs reported in rows [a] within each of the tables do not vary significantly from one another. These suggest that $\tilde{\beta}_k$ and $\hat{\beta}_k$ are closely similar. Such findings are as anticipated and theoretically deducible from the estimation consistency. (ii) The RMSEs reported in rows [b] in each table (even under different weighting matrices) are always smaller than those in rows [a]. These differences are particularly significant for permanent grassland and arable. Such findings stress the need to incorporate the random effects and SED into the model in order to improve the predictive accuracy. (iii) It seems that at $\kappa = 5$, the RMSEs reported for $W_k^{\kappa NN}$ are relatively close to those of W_k^{SOI} . Nonetheless, the evidence is not conclusive on which specifications of the weighting matrix is able to bring about better forecasts. Moreover, in Tables 15–17 in the Online Appendix, rows [c] present the corresponding LM_{BP} test statistics and p -values under the different weighting matrices. In all cases, the LM_{BP} test statistics far exceed 3.84, which is the 95% critical value for the chi-squared distribution with one degree of freedom. These lead to rejection of the null hypothesis and a suggestion that superiority in the predictive accuracy reported in the previous paragraph was not caused by measurement error. The predictors under category C are statistically different from those listed under category B and are able to provide more accurate prediction.

5. CONCLUSIONS

We contributed toward building a better understanding of farmers' responses to behavioral drivers of land-use decision by establishing a new analytical procedure that can handle complex data structures and overcome various drawbacks suffered by existing methods. Firstly, our procedure made use of spatially high-resolution data so that idiosyncratic effects of physical environment drivers could be explicitly modeled. Secondly, we addressed the famous censored data problem to ensure

theoretical consistency of the parameter estimates. Thirdly, we incorporated SED and heterogeneity in order to gain efficiency, more accurate formulation of the variances for the parameter estimates and hence more effective statistical inferences. Finally, we reduced the computational burden and improved estimation accuracy by introducing a GMM/QML hybrid estimation procedure. We applied our method to spatially high resolution data in England and found that the number of coefficient estimates that are statistically significant reduces significantly when the SED and heterogeneity are taken into consideration. Inevitably, this leads to conclusions about causal effects of climatic and physical environment, and environmental policy on land-use shares that differed significantly from those made based on methods that are currently used in the literature. Moreover, we showed that our method enables derivation of a more effective predictor of the land-use shares, which is utterly useful from the policy-making point of view.

REFERENCES

- Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica: Journal of the Econometric Society*, 41, 997–1016.
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic.
- Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics: Regression models. In *Handbook of applied economic statistics* (pp. 257–259). CRC Press.
- Ay, J. S., Chakir, R., & Le Gallo, J. (2017). Aggregated versus individual land-use models: Modeling spatial autocorrelation to increase predictive accuracy. *Environmental Modeling & Assessment*, 22(2), 129–145.
- Baltagi, B. H. (1980). On seemingly unrelated regressions with error components. *Econometrica: Journal of the Econometric Society*, 48, 1547–1551.
- Baltagi, B. H. (2008). *Econometric analysis of panel data*. John Wiley & Sons.
- Baltagi, B. H., Bresson, G., & Pirotte, A. (2012). Forecasting with spatial panel data. *Computational Statistics & Data Analysis*, 56(11), 3381–3397.
- Baltagi, B. H., & Li, D. (2004). Prediction in the panel data model with spatial correlation. In *Advances in spatial econometrics* (pp. 283–295). Springer.
- Baltagi, B. H., & Li, D. (2006). Prediction in the panel data model with spatial correlation: The case of liquor. *Spatial Economic Analysis*, 1(2), 175–185.
- Baltagi, B. H., & Pirotte, A. (2011). Seemingly unrelated regressions with spatial error components. *Empirical Economics*, 40(1), 5–49.
- Bhattacharjee, A., & Jensen-Butler, C. (2006). *Estimation of spatial weights matrix, with an application to diffusion in housing demand* [Centre for Research into Industry, Enterprise and the Firm Discussion Paper, 519].
- Breusch, T. S., & Pagan, A. R. (1980). Lagrange multiplier test and its applications to model specification. *The Review of Economic Studies*, 47(1), 239–253.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconomics: Methods and applications*. Cambridge University Press.
- Carrion-Flores, C., & Irwin, E. G. (2004). Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics*, 86(4), 889–904.
- Chakir, R., & Le Gallo, J. (2013). Predicting land use allocation in France: A spatial panel data analysis. *Ecological Economics*, 92, 114–125.
- Chambers, R. G., & Just, R. E. (1989). Estimating multioutput technologies. *American Journal of Agricultural Economics*, 71(4), 980–995.
- Cliff, A. D., & Ord, J. K. (1973). Spatial autocorrelation (No. 04; QA278. 2, C5.).
- Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: Models & applications*. Taylor & Francis.

- Committee on Climate Change (CCC). (2018). *Land use: Reducing emissions and preparing for climate change.* www.theccc.org.uk/publication/land-use-reducing-emissions-and-preparing-for-climate-change
- Committee on Climate Change (CCC). (2020). Land use: Policies for a Net Zero UK. www.theccc.org.uk/publication/land-use-policies-for-a-net-zero-uk/
- Deng, Q., & Xue, J. (2014). Multivariate Tobit system estimation of education expenditure in urban China. *The Singapore Economic Review*, 59, 1450005.
- Department for Environment, Food and Rural Affairs. (2020). *Defra statistics: Agricultural facts england regional profiles February 2020.* www.gov.uk/government/statistics/agricultural-facts-england-regional-profiles
- Dong, D., Gould, B. W., & Kaiser, H. M. (2004). Food demand in Mexico: An application of the Amemiya-Tobin approach to the estimation of a censored food system. *American Journal of Agricultural Economics*, 86(4), 1094–1107.
- Fezzi, C., & Bateman, I. J. (2011). Structural agricultural land use modelling for spatial agro-environmental policy analysis. *American Journal of Agricultural Economics*, 93(4), 1168–1188.
- Fezzi, C., Harwood, A., Lovett, A., & Bateman, I. J. (2015). The environmental impact of climate change adaptation on land use and water quality. *Nature Climate Change*, 5, 255–260.
- Greene, W. H. (2008). The econometric approach to efficiency analysis. In *The measurement of productive efficiency and productivity growth* (Vol. 1, Iss. 1, pp. 92–250).
- Kapoor, M., Kelejian, H. H., & Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1), 97–130.
- Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2), 509–533.
- Lacroix, A., & Thomas, A. (2011). Estimating the environmental impact of land and production decisions with multivariate selection rules and panel data. *American Journal of Agricultural Economics*, 93(3), 784–802.
- Lee, L. F. (1993). Multivariate Tobit models in econometrics. In G. S. Maddala, C. R. Rao, & H. D. Vinod, Eds. *Handbook of Statistics* (Vol. 11., chap. 6, pp. 145–173). North-Holland.
- Lee, L. F., & Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of econometrics*, 154(2), 165–185.
- Li, M., JunJie, W., & Deng, X. (2013). Identifying drivers of land use change in China: A spatial multinomial logit model analysis. *Land Economics*, 89(4), 632–654.
- Liu, S. F., & Yang, Z. (2015). Asymptotic distribution and finite sample bias correction of QML estimators for spatial error dependence model. *Econometrics*, 3(2), 376–411.
- Marcos-Martinez, R., Bryan, B. A., Connor, J. D., & King, D. (2017). Agricultural land-use dynamics: Assessing the relative importance of socioeconomic and biophysical drivers for more targeted policy. *Land Use Policy*, 63, 53–66.
- Mattison, E. H., & Norris, K. (2005). Bridging the gaps between agricultural policy, land-use and biodiversity. *Trends in Ecology & Evolution*, 20(11), 610–616.
- Meyerhoefer, C. D., Ranney, C. K., & Sahn, D. E. (2005). Consistent estimation of censored demand systems using panel data. *American Journal of Agricultural Economics*, 87(3), 660–672.
- Ministry of Housing, Communities and Local Government. (2017). *Land use in England 2017.* www.gov.uk/government/collections/land-use-in-england-experimental-statistics
- Moscone, F., Knapp, M., & Tosetti E. (2007). Mental health expenditure in England: A spatial panel approach. *Journal of Health Economics*, 26(4), 842–864.
- Nelson, G. C., & Hellerstein, D. (1997). Do roads cause deforestation? Using satellite images in econometric analysis of land use. *American Journal of Agricultural Economics*, 79(1), 80–88.
- Reidsma, P., Tekelenburg, T., Van den Berg M., & Alkemade R. (2006). Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union. *Agriculture, Ecosystems & Environment*, 114(1), 86–102.
- Sterling, S. M., Ducharme, A., & Polcher, J. (2013). The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change*, 3(4), 385–390.
- Tsay, R. S. (2005). *Analysis of financial time series* (Vol. 543). John Wiley & Sons.
- Wang, X., & Kockelman, K. M. (2007). Specification and estimation of a spatially and temporally autocorrelated seemingly unrelated regression model: Application to crash rates in China. *Transportation*, 34(3), 281–300.

- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.
- Yang, Z. (2013). *Quasi-Maximum likelihood estimation for spatial panel data regressions*. Research Collection School of Economics. <https://ink.library.smu.edu.sg>
- Yen, S. T., Lin, B. H., & Smallwood, D. M. (2003). Quasi-and simulated-likelihood approaches to censored demand systems: Food consumption by food stamp recipients in the United States. *American Journal of Agricultural Economics*, 85(2), 458–478.

This page intentionally left blank

CHAPTER 12

LOCAL CLIMATE SENSITIVITY: WHAT CAN TIME SERIES OF DISTRIBUTIONS REVEAL ABOUT SPATIAL HETEROGENEITY OF CLIMATE CHANGE?

J. Isaac Miller

Department of Economics, University of Missouri, Columbia, Missouri, USA

ABSTRACT

Transient climate sensitivity relates total climate forcings from anthropogenic and other sources to surface temperature. Global transient climate sensitivity is well studied, as are the related concepts of equilibrium climate sensitivity (ECS) and transient climate response (TCR), but spatially disaggregated local climate sensitivity (LCS) is less so. An energy balance model (EBM) and an easily implemented semiparametric statistical approach are proposed to estimate LCS using the historical record and to assess its contribution to global transient climate sensitivity. Results suggest that areas dominated by ocean tend to import energy, they are relatively more sensitive to forcings, but they warm more slowly than areas dominated by land. Economic implications are discussed.

Keywords: Transient climate sensitivity; local climate sensitivity; energy balance model; temperature anomalies; functional time series; econometric analysis of climate change

JEL Classification: C14; C23; Q54

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 319–350

Copyright © 2023 by J. Isaac Miller

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:10.1108/S0731-90532023000045B014

1. INTRODUCTION

Attribution of the evident upward movement in global temperatures over recent decades is a topic of central importance in climate science and growing importance in economics. Characterizing and predicting the anthropogenic footprint on our climate is a difficult but important task, and one that has been tackled both by complex physical climate models and relatively simple statistical models that aim to make the most out of historical series by capturing their most salient properties. The latter are typically based on low-dimensional EBMs. Statistical climate models that are complex enough to capture salient features of the data and make meaningful predictions yet simple enough to allow sufficient replications to assess uncertainty are useful to economists in estimating damages and comparing policies.

In particular, zero-dimensional EBMs provide physical bases for many statistical analyses that relate total radiative forcings (TRF) to global mean temperature anomalies (GMTA). The physical link is described by [Gregory and Forster \(2008\)](#) and [Schwartz \(2012\)](#), *inter alia*. A key parameter in these models, often labeled $1/\lambda$ and measured in $^{\circ}\text{C}/(\text{W}/\text{m}^2)$, is referred to as *transient climate sensitivity* or simply *climate sensitivity* ([Boer & Yu, 2003](#); [Held et al., 2010](#)).

A number of authors have estimated climate sensitivity using statistical methods applied to zero-dimensional EBMs. Recent methods generally allow for one of two types of long-run co-movement: cointegrating or co-breaking series. Proponents of the former argue that both TRF and GMTA have a common stochastic trend, and they include [Stern and Kaufmann \(2000\)](#), [Kaufmann and Stern \(2002\)](#), [Kaufmann et al. \(2006a, 2006b, 2010, 2013\)](#), and [Pretis \(2020\)](#). Those of the latter argue that TRF and GMTA have deterministic trends that have experienced one or more contemporaneous breaks, and they include [Estrada et al. \(2013a, 2013b\)](#) and [Estrada and Perron \(2014\)](#). A lack of consensus among these authors suggests the use of statistical methods that maintain some robustness to either type of trending behavior in the data.

Zero-dimensional EBMs predict the temperature of a theoretically isothermal (spatially homogeneous) planet. While GMTA provides a convenient measure of central tendency, it ignores the temperature anomaly distribution, and thus zero-dimensional EBMs cannot account for spatial heterogeneity. In contrast, one-dimensional EBMs ([Budyko, 1969](#); [Sellers, 1969](#)) account for heterogeneity in temperature and net (horizontal) heat transport (NHT) across latitudes, while two-dimensional EBMs ([Sellers, 1976](#), [North et al., 1983](#)) additionally account for heterogeneity within latitudes.

In this chapter, a non-gridded one-dimensional EBM is proposed for the purpose of estimating spatially disaggregated climate sensitivity or *LCS* using the historical record.¹ This model takes into account the distribution of spatially disaggregated temperature anomalies – or their fast components in the spirit of [Held et al. \(2010\)](#). Considering this distribution simplifies the statistical analysis relative to a gridded model without sacrificing as much information about global heterogeneity as is lost to aggregation for a zero-dimensional EBM or even for a one-dimensional EBM based on latitude.

A novel but easily implemented semiparametric estimation procedure using functional time series is proposed to estimate spatially heterogeneous NHT as a nonlinear function of temperature anomalies and its contribution to climate sensitivity. Not surprisingly, homogeneity is rejected. An NHT function is estimated that is strikingly nonconstant and nonlinear in temperature anomalies, but we show that the nonlinearity is consistent with the linear Budyko-Sellers model. Over the globe, oceans tend to be net importers of energy, while continents tend to be net exporters.

Spatial heterogeneity of NHT implied by the nonlinearity is an important feature of more complicated climate models (Trenberth et al., 2001, e.g.). The geographical distribution of the estimated local climate sensitivities is roughly comparable to that of the analogous parameter mapped by Boer and Yu (2003) using a completely different approach and data. The heterogeneity of LCS is tied to polar amplification, and recent work by Francis and Vavrus (2012), Liu et al. (2012), Screen and Simmonds (2013), *inter alia* emphasize a link between polar amplification and severe weather at mid-latitudes. The IPCC's Fifth Assessment Report (IPCC, 2014) echoes this link with a prediction of increasingly likely heat waves and “extreme precipitation events” over mid-latitudes. Increasingly frequent and severe weather events, such as the active Atlantic hurricane season of 2020, the historic European heatwave of 2022, and the massive flooding in Pakistan in 2022, certainly have had negative economic impacts. Moreover, as Brock and Xepapadeas (2017) point out, ignoring local differences in NHT – as economic models based on zero-dimensional EBMs do – may systematically bias optimal mitigation policy.

Both the proposed methodology and the empirical application of that methodology to climate change firmly link the present research to that of Professor Joon Y. Park, in addition to that of his wife and frequent coauthor, Professor Yoosoon Chang, some of his former students, and even a recent “grandstudent” of theirs. The link between a single time series and a time series of distributions is directly related to Professor Park’s work on functional time series and less directly to that on functional coefficients in published papers by Park and Hahn (1999), Park et al. (2010), Park and Qian (2012), Chang et al. (2014, 2016a, 2016b, 2016c, 2020), Nam (2018), and Miller and Nam (2020) in addition to a number of as-yet unpublished papers. The latter three squarely lie within a growing literature on the analysis of climate and climate change using econometric methods, loosely termed *climate econometrics*, around which special issues of *Journal of Econometrics*, *Energy Economics*, and other journals have recently been organized.

The rest of the chapter is organized as follows. Section 2 describes the physical and statistical models considered in this research, and then Section 3 proposes a simple estimation procedure that is summarized in three steps. Section 4 presents the data employed with results of the estimation procedure and discussions of these results. Section 5 concludes. An appendix contains technical details on the approximation to the EBM that is empirically analyzed in the chapter.

2. PHYSICAL AND STATISTICAL MODELS

The first law of thermodynamics equates absorbed energy Q , incoming energy net of ice and land albedo (reflection), with radiated energy E of the Earth in a steady state (Peixoto & Oort, 1992; Taylor, 2005, e.g.). Both Q and E are functions of the Earth's effective temperature, but are often expressed as functions $Q(\bar{\tau})$ and $E(\bar{\tau})$ of surface temperature, denoted here by $\bar{\tau}$ (Schwartz, 2012, e.g.).

Following Schwartz (2012), planetary net heat flux into the planet in a zero-dimensional EBM is given by $N(\bar{\tau}) = h + Q(\bar{\tau}) - E(\bar{\tau})$ with the introduction of an external spatially uniform forcing h (TRF) expressed in W/m^2 . $Q(\bar{\tau}) - E(\bar{\tau})$ is typically assumed to be linear in $\bar{\tau}$ due to Budyko's (1969) linear representation of E and with a constant ice line for Q , so we let $Q(\bar{\tau}) - E(\bar{\tau}) = QS - A - \lambda\bar{\tau}$ following North and Cahalan (1981) with constants Q , S , and A but replacing $-B$ in their notation with $-\lambda$ to match the notation of Schwartz (2012) for the derivative of $Q(\bar{\tau}) - E(\bar{\tau})$ with respect to $\bar{\tau}$.

The planetary net heat flux becomes

$$N(\bar{\tau}) = h + QS - A - \lambda\bar{\tau} \quad (1)$$

under the linear specification. Noting that global temperature is an implicit function of TRF h , total differentiation of equation (1) yields

$$dN(\bar{\tau}) = dh - \lambda d\bar{\tau}, \quad (2)$$

which is the well-known energy balance equation (EBM) of Gregory and Forster (2008, equation 1), Schwartz (2012, equation 5), *inter alia*.²

Solving for a change in the global temperature $d\bar{\tau}$ resulting from a change in TRF dh yields

$$\frac{d\bar{\tau}}{dh} = \frac{1}{\lambda} \left[1 - \frac{d}{dh} N(\bar{\tau}) \right], \quad (3)$$

which is the *climate sensitivity* (Boer & Yu, 2003; Held et al., 2010), and it equals $1/\lambda$ in the steady state in which $dN(\bar{\tau}) = 0$. When the system is out of equilibrium, the derivative $d\bar{\tau}/dh$ is referred to as the transient climate sensitivity (Schwartz, 2012). Allowing for measurement errors and idiosyncrasies, climate sensitivity may be estimated by regressing GMTA onto TRF over the historical record (Estrada et al., 2013b), similarly to equation (12) below.

Often discussed in the literature are the related concepts of ECS and TCR (Bindoff et al., 2013, e.g.). Both are defined as temperature responses to a doubling in atmospheric concentration of CO_2 from pre-industrial levels and both may be calculated from the transient climate sensitivity, but ECS takes into account ocean heat uptake, while TCR does not. In particular, TCR is $h_{2\times} d\bar{\tau} / dh$ with $h_{2\times} = 3.71 \text{ W/m}^2$, and ECS relates to climate sensitivity by a parameter governing the ocean's heat uptake (Schwartz, 2012). It is appropriate to think of climate sensitivity, as defined here, as a model parameter and TCR and ECS as model outputs based on that parameter.

The EBM in [equation \(2\)](#) is described as zero-dimensional in the sense that temperature is homogeneous over the globe. While such an assumption is obviously unrealistic in and of itself, it allows simple analyses to estimate aggregate global characteristics of the climate, such as climate sensitivity. [Budyko \(1969\)](#) and [Sellers \(1969\)](#) extend the zero-dimensional planetary EBM to a one-dimensional EBM that allows for horizontal (advective) heat transport. At a given latitude θ (in radians), a forced EBM is given by

$$X(\tau_\theta - \bar{\tau}, \theta) = h + QS(\theta) - A - \lambda\tau_\theta, \quad (4)$$

where $X(\tau_\theta - \bar{\tau}, \theta)$ is NHT (outgoing for positive values), a function of latitude and the deviation of surface temperature τ_θ (itself a function of latitude) from global mean temperature $\bar{\tau}$, and where $S(\theta)$ depends on location (see North and Cahalan, 1981, e.g.) and integrates globally to S defined above.

A variety of generalizations of the one-dimensional EBM in [equation \(4\)](#) have been proposed. Two-dimensional EBMs allow heterogeneity in temperature based on other geographic characteristics, such as land versus ocean coverage (e.g., [North et al., 1983](#); [Sellers, 1976](#)). Moist EBMs such as that of [Langen and Alexeev \(2007\)](#) and [Merlis and Henry \(2018\)](#) allow for water vapor in the atmosphere, while [Siler et al. \(2018\)](#) take moist EBMs a step further by allowing for the effect of the Hadley Circulation in the tropics.

These approaches rely on an EBM defined over a continuum of temperatures that are themselves defined over a continuum of latitudes. Simple “two-box” models such as that of [Langen and Alexeev \(2007\)](#) can be derived from underlying continuous processes and aggregate concepts such as the global mean are derived accordingly. However, statistical estimation of LCS requires spatially disaggregated data, so discretization to a much finer resolution than that of a two-box model or a global aggregate is needed. We define functions of spatially disaggregated data, such as the global mean temperature $\bar{\tau}$, as aggregates of observations τ_L drawn systematically from locations $L(\theta, \varphi)$ given by grid boxes of equal latitude θ and equal longitude φ .

To ameliorate measurement errors acknowledged in the climate literature, disaggregated temperature data typically are expressed as temperature anomalies $r_L = \tau_L - \tau_L^B$, where τ_L^B is the mean temperature over box L during a base period. The expression

$$X(r_L, L) = h + QS(L) - A - \lambda(r_L + \tau_L^B) \quad (5)$$

gives a two-dimensional EBM that is a function of discrete boxes L , but that is derived in the appendix as discretization of a two-dimensional EBM that is continuous in latitude and longitude.

We further define the NHT function $X(r)$ to be a locationless measure of NHT at temperature anomaly r , which is calculated as the average NHT $X(r_L, L)$ over all boxes $\{L : r_L = r\}$ that have the same temperature anomaly r at a given time. The EBM can be written as

$$X(r) = h + QS(r) - A - \lambda(r + \tau^B(r)), \quad (6)$$

where τ^B is defined to be the average of the base temperatures over the set of locations $\{L : r_L = r\}$ and is thus a nonlinear function of r . $S(r)$ is defined analogously to $X(r)$. See the [Appendix](#) for further details. Note that the nonlinearity in r allows for spatial heterogeneity of NHT.

It is useful to step back and consider the significance of r and $X(r)$ more carefully. Aggregating NHT across locations using $X(r)$ groups them by anomaly. In order for sharp distinctions between NHT at different locations to remain, those locations must share common characteristics. Such commonality is certainly plausible in the North Atlantic, in the Arctic, and around Greenland, where some of the largest anomalies occur, or in the Tropics, where some of the smallest occur. The aggregation reduces the model to a single dimension r , like that of early one-dimensional EBMs. However, NHT is not a function of latitude as it is in those models, allowing for more realistic complexity. For example, the Southern Ocean has not warmed as fast as the Arctic Ocean, as a one-dimensional EBM implies it should.

As a frame of reference, [Leduc et al. \(2016\)](#) aggregate temperatures by region in order to assess temperature change as a function of cumulative carbon emissions. They find the relationship to be strikingly linear in most regions (see their [Fig. 2](#)), supportive of the linearity of regional NHT functions along the lines of the Budyko-Sellers EBM discussed above. Yet, their slopes are quite different across regions, supporting the nonlinearity of $X(r)$ to capture spatial heterogeneous climate sensitivity. Indeed, nonlinearity of $X(r)$ induced by aggregation may explain why the evidence for linearity in some of the regions considered by [Leduc et al. \(2016\)](#) is stronger than in others. Regions with relatively homogeneous NHT – near the Equator, say – have nearly linear NHT functions, and the results of those authors are consistent with this reasoning.

The work of [Castruccio et al. \(2014\)](#) provides another frame of reference, both due to the long-run linearity of temperature in forcings imposed by those authors (as derived by [Miller & Brock, 2021](#)) in emulating model output and the comparison in their [Fig. 6](#) with pattern scaling. Pattern scaling assumes a linear relationship between local anomalies and GMTA. If the relationship between GMTA and TRF is also linear, pattern scaling imposes linearity of local NHT. The logic of nonlinearity resulting from aggregating heterogeneous climate sensitivities discussed above is therefore not inconsistent with linearity assumption of pattern scaling.

A methodological benefit of the aggregation across anomalies is that $X(r)$ may be viewed as a functional in a Hilbert space. Specifically, we define the Hilbert space H similarly to [Chang et al. \(2016c, 2020\)](#) but without temporal demeaning, as

$$H = \left\{ w \middle| \int w^2(r) dr < \infty \right\}, \quad (7)$$

with inner product $\langle v, w \rangle = \int v(r)w(r) dr$ for $v, w \in H$, where the integrals are evaluated over the range of temperature anomalies $[r^-, r^+]$.

The distribution of temperature anomalies, given by $f(r)$, is a functional in that space that has been employed already by [Chang et al. \(2020\)](#) and [Miller and](#)

[Nam \(2020\)](#). The former analyzed and tested for the type of nonstationarity in the temporal evolution of the temperature anomaly distribution, while the latter used the distribution of ocean temperature anomalies to measure a multidecadal and multibasin temperature cycle to predict the next slowdown in global warming.

Integrating the EBM in [equation \(6\)](#) over all temperature anomalies observed at a given time yields

$$\langle X, f \rangle = h + QS - A - \lambda \bar{\tau}, \quad (8)$$

where $S = \langle S, f \rangle$ and $\bar{\tau} = \langle \tau, f \rangle$ are the same as in [equation \(1\)](#) up to an approximation error.³ The right hand sides of [equations \(1\)](#) and [\(8\)](#) are the same, but the left-hand sides appear to be different. In fact, the left-hand sides of both equations are zero. The left-hand side of [equation \(1\)](#) shows planetary incoming energy less aggregate outgoing planetary energy, which must be zero in equilibrium. In contrast, the left-hand side of [equation \(8\)](#) shows aggregate net energy transported horizontally from each region of the globe, which must also aggregate to zero. Even when the planet is in a state of equilibrium, so that $N(\bar{\tau}) = 0$, $X(r)$ will be positive over some regions (energy “exporters”) and negative over others (energy “importers”).

How does the EBM in [equation \(8\)](#) change when the climate is forced? Going back to [equations \(2\)](#) and [\(3\)](#) for the globe, a change in forcing drives a change in GMTA, which is reflected by $1/\lambda$ when planetary net heat flux is zero. In order for that to happen, f must be an implicit function of h . We maintain the assumption discussed above that $Q(\bar{\tau}) - E(\bar{\tau})$ is linear in temperature $\bar{\tau}$ globally, and we extend it to apply locally. In particular, we assume that local changes in the temperature anomaly distribution affect the local temperature but not $S(\theta)$.

Analogously to [equations \(2\)](#) and [\(3\)](#), we differentiate [equation \(8\)](#) to get

$$d\langle X, f \rangle = dh - \lambda d\bar{\tau}, \quad (9)$$

and

$$\frac{d\bar{\tau}}{dh} = \frac{1}{\lambda} \left[1 - \frac{d}{dh} \langle X, f \rangle \right] = \frac{1}{\lambda} \left[1 - \left\langle X, \frac{df}{dh} \right\rangle \right], \quad (10)$$

where the last equality follows by assuming that the NHT function $X(r)$ is invariant with respect to TRF.⁴

The conclusion that $d\bar{\tau}/dh = 1/\lambda$ in the steady state holds for both [equations \(3\)](#) and [\(10\)](#). Although planetary net flux must be zero in a steady-state equilibrium, local NHT might never be zero, because the temperature anomaly distribution may change. Casual observation of the data suggest that this distribution does in fact change over time, and [Chang et al. \(2020\)](#) provide evidence supporting an even stronger result that the distributions are stochastically trending over time. Clearly $df/dh \neq 0$.

In order to express heterogeneous NHT more explicitly, [equation \(10\)](#) may be rewritten as

$$\frac{d\bar{\tau}}{dh} = \langle 1, C \rangle / (r^+ - r^-)$$

where

$$C(r) = \frac{1}{\lambda} \left(1 - X(r) \frac{df(r)}{dh} (r^+ - r^-) \right) \quad (11)$$

is defined to be the *LCS*, expressed in $^{\circ}\text{C}/(\text{W}/\text{m}^2)$. NHT across the globe integrates to zero, so that climate sensitivity, given by the normalized integral of local climate sensitivities across all temperature anomalies, is $d\bar{\tau} / dh = 1/\lambda$ just as with the zero-dimensional EBM used to obtain [equation \(3\)](#).

The negative sign after $1/\lambda$ in $C(r)$ implies that as TRF increases, areas that pass along the additional energy from this increase by exporting it ($X(r) > 0$) are less sensitive to changes in TRF ($C(r) < 1/\lambda$). Those that do not pass along the additional energy and thus import energy on net ($X(r) < 0$) are more sensitive to changes in TRF ($C(r) > 1/\lambda$).

3. METHODOLOGY

3.1. Estimation of LCS

We propose estimating the LCS by separately estimating each component of $C(r)$ in [equation \(11\)](#): $df(r)/dh$, $1/\lambda$, and $X(r)$. $f(r)$ is estimated over each year $t = 1, \dots, T$, and $f(r)$, $df(r)/dh$, and $X(r)$ are estimated at equidistant points $r_i \in [r^-, r^+]$ for $i = 0, 1, \dots, n$, such that $r_0 = r^-, r_n = r^+, \epsilon = r_i - r_{i-1}$, and thus $(r^+ - r^-) = n\epsilon$ for some small number ϵ , which we set to be 0.005.

3.1.1. Step 1: $df(r)/dh$

Before estimating $df(r)/dh$, the density $f_t(r)$ is estimated for each year t using a standard kernel density estimator. Because global temperature anomaly data sets contain a large number of cross-sectional (spatial) observations for each time period, estimates of $f_t(r)$ are precise enough that we make no further distinction between the densities and their estimates, following [Chang et al. \(2020\)](#) and [Miller and Nam \(2020\)](#).

One way to estimate $df(r)/dh$ is by linear regression. Specifically, this task is accomplished by regressing $f_t(r_i)$ at each r_i on h_i and an intercept. Estimates of the coefficient on h_i from these $n - 1$ regressions estimate the derivative $df(r_i)/dh$ at each r_i .

3.1.2. Step 2: Climate Sensitivity, $1/\lambda$

Because the last term of [equation \(10\)](#) equals zero, climate sensitivity is given by the first term, $1/\lambda$. Estimating $1/\lambda$ is straightforward using linear regression. To this end, a conditional expectation function may be defined as

$$\langle \iota, f_t \rangle = \mathbf{E}[\langle \iota, f_t \rangle | h_t] + \langle \varepsilon_{S0}, f_t \rangle,$$

where $\langle \iota, f_t \rangle$ with identity function $\iota(r)$ is the GMTA and $\langle \varepsilon_{S0}, f_t \rangle = \langle \iota, f_t \rangle - \mathbf{E}[\langle \iota, f_t \rangle | h_t]$ may be interpreted as the aggregate of errors resulting from stochastic forcing ([North et al., 1981](#)), measurement error, or other sources of ephemeral disequilibrium.

Adding and subtracting $\alpha_{S0,0} + \alpha_{S0,1}h_t$ yields

$$\langle \iota, f_t \rangle = \alpha_{S0,0} + \alpha_{S0,1}h_t + (\mathbf{E}[\langle \iota, f_t \rangle | h_t] - (\alpha_{S0,0} + \alpha_{S0,1}h_t)) + \langle \varepsilon_{S0}, f_t \rangle,$$

where the third term reflects unmodeled nonlinearity in the spatial aggregate. In fact, we know that $d\bar{r}/dh = d\bar{\tau}/dh$ is constant in the physical model, so the third term reduces to zero. By defining $\varepsilon_{S0t} = \langle \varepsilon_{S0}, f_t \rangle$ the model may be written simply as

$$S0: \bar{r}_t = \alpha_{S0,0} + \alpha_{S0,1}h_t + \varepsilon_{S0t}, \quad (12)$$

for $t = 1, \dots, T$ years. We may interpret $\alpha_{S0,1}$ as the climate sensitivity, or more precisely as global mean climate sensitivity, given by $1/\lambda$, as long as the unmeasured forcing ε_{S0t} is idiosyncratic with respect to changes in the measured forcing h_t .

The simple linear regression in [equation \(12\)](#) is employed by [Estrada et al. \(2013b\)](#) and other authors to estimate climate sensitivity. Those authors explicitly link the regression model to a zero-dimensional EBM, but we have now shown how it also links to a higher-dimensional EBM.

3.1.3. Step 3: NHT Function, $X(r)$

The most complicated task and main methodological contribution of this research is to estimate the local NHT function $X(r)$. As a simple motivating example, suppose we know that $X(r)$ is quadratic such that $X(r)/\lambda = \gamma_0 + \gamma_1 r + \gamma_2 r^2$ with unknown γ_0 , γ_1 , and γ_2 . The regression given by Model S0 in [equation \(12\)](#) is motivated by [equation \(10\)](#) with the equilibrium condition that $\langle X, f \rangle = 0$. But if the equilibrium condition holds only in expectation then information about the NHT function may be gleaned from adjustment to that equilibrium, which motivates including the NHT function directly into the regression.

Substituting the equilibrium condition

$$\mathbf{E}\langle X/\lambda, f_t \rangle = \mathbf{E}\langle \gamma_0 + \gamma_1 \iota + \gamma_2 \iota^2, f_t \rangle = 0 \quad (13)$$

into Model S0 and recalling that $\langle \iota, f_t \rangle = \bar{r}_t$ and $\langle 1, f_t \rangle = 1$ for all t yields

$$\bar{r}_t = \alpha_{S0,0} + \alpha_{S0,1} h_t + (\gamma_0 + \gamma_1 \bar{r}_t + \gamma_2 \langle \iota^2, f_t \rangle) + \tilde{\varepsilon}_{S0t}, \quad (14)$$

where $\tilde{\varepsilon}_{S0t} = \varepsilon_{S0t} + (\mathbf{E}\langle X/\lambda, f_t \rangle - \langle X/\lambda, f_t \rangle)$. The parameter γ_0 is not identified due to the presence of $\alpha_{S0,0}$. The parameter γ_1 is not identified either, because \bar{r}_t appears on both sides of the equation. For known parameters γ_0 and γ_1 , [equation \(14\)](#) becomes a feasible regression and γ_2 is identified.

We condition out the nuisance terms γ_0 and $\gamma_1 \bar{r}_t$ to estimate γ_2 . Define δ_{20} and δ_{21} by

$$\begin{bmatrix} \delta_{20} \\ \delta_{21} \end{bmatrix} = \left(\mathbf{E} \begin{bmatrix} 1 \\ \bar{r}_t \end{bmatrix} \begin{bmatrix} 1 \\ \bar{r}_t \end{bmatrix}' \right)^{-1} \mathbf{E} \begin{bmatrix} 1 \\ \bar{r}_t \end{bmatrix} \langle \iota^2, f_t \rangle$$

so that $x_{2t} = \langle \iota^2, f_t \rangle - \delta_{21} \bar{r}_t - \delta_{20}$ orthogonalizes $\langle \iota^2, f_t \rangle$ and the quadratic in parentheses in [equation \(14\)](#) more generally with respect to GMTA and a constant. Specifically, x_{2t} is estimated by the series of fitted residuals from a regression of $\langle \iota^2, f_t \rangle$ onto \bar{r}_t and a constant with linear projection coefficients given by δ_{21} and δ_{20} .

The orthogonalization allows a feasible regression given by

$$S1: \bar{r}_t = \alpha_{S0,0} + \alpha_{S0,1} h_t + x_{2t} \gamma_2 + \tilde{\varepsilon}_{S0t} \quad (15)$$

in place of the infeasible regression in [equation \(14\)](#). Co-movements of the components of the equilibrium condition in [equation \(13\)](#) with the GMTA \bar{r}_t are removed, leaving only adjustments to the equilibrium given by non-zero γ_2 .

Having estimated δ_{20} and δ_{21} using the regression to obtain x_{2t} and having estimated γ_2 using the regression in [equation \(15\)](#), we can now identify γ_0 and γ_1 . Setting

$$x_{2t} \gamma_2 = \gamma_2 (\langle \iota^2, f_t \rangle - \delta_{21} \bar{r}_t - \delta_{20}) = (\gamma_0 + \gamma_1 \bar{r}_t + \gamma_2 \langle \iota^2, f_t \rangle)$$

equates the models in [equations \(14\)](#) and [\(15\)](#). We could identify $\gamma_0 = -\gamma_2 \delta_{20}$ and $\gamma_1 = -\gamma_2 \delta_{21}$ in this way.

Suppose we want to impose the stronger equilibrium condition that $\langle X, f \rangle = 0$ and not only in expectation. Setting $\gamma_0 = \gamma_1 = \gamma_2 = 0$ trivially accomplishes this task. With a non-zero γ_2 , however, the same condition is imposed by introducing a time-varying intercept identified by

$$\gamma_{0t} = -\gamma_1 \bar{r}_t - \gamma_2 \langle \iota^2, f_t \rangle = -\gamma_2 (\langle \iota^2, f_t \rangle - \delta_{21} \langle \iota, f_t \rangle)$$

so that $x_{2t} = 0$. This gives us an alternative to identify γ_0 while at the same time allowing for a time-varying NHT.

Results not shown show very little change in the intercept over time, so that

$$-\gamma_2\delta_{20} \approx -\gamma_2(\langle \iota^2, f_i \rangle - \delta_{21}\langle \iota, f_i \rangle) \approx -\gamma_2(\langle \iota^2, f \rangle - \delta_{21}\langle \iota, f \rangle)$$

with f defined as an average of (f_i) over time for each r_i in its support. The first approximation would be an equality under the stricter equilibrium condition that $x_{2t} = 0$ and the last approximation follows from the empirical finding that the relationship is stable over time, so we identify γ_0 as $-\gamma_2(\langle \iota^2, f \rangle - \delta_{21}\langle \iota, f \rangle)$.

In fact, a time-invariant relationship is consistent with the literature. [Estrada et al. \(2013b\)](#) allow for structural instability (breaks) in the trends of TRF and GMTA, yet they find that the series break together so that their relationship is stable. [Eroğlu et al. \(2022\)](#) explicitly test for time-varying cointegration against linear cointegration, and find that GMTA linearly cointegrates with well-mixed greenhouse gases, the most persistent component of TRF.

Finally, in the case of a known quadratic, $X(r)$ is identified by dividing $\gamma_0 + \gamma_1 r + \gamma_2 r^2$ by the climate sensitivity $1/\lambda$, an estimate of which comes from the estimate of $\alpha_{S0.1}$ in Model S0 in [equation \(12\)](#). This completes the estimation procedure for a known quadratic NHT function.

More generally, assume that the NHT function admits a series expansion that is approximated by

$$X/\lambda = \gamma_0 + \gamma_1 r + \gamma'_{2:m} g_{2:m}(r)$$

with known functions $g_{2:m}(r) = (g_2(r), \dots, g_m(r))'$ and $\gamma_{2:m} = (\gamma_2, \dots, \gamma_m)'$ for $m \geq 2$. This assumption is in the spirit of the assumption of [North \(1975\)](#) that the NHT function admits a Legendre polynomial expansion in latitude, except that we have aggregated over locations with common observed anomalies.

The logic of the infeasibility of [equation \(14\)](#) remains. Rather than a single regression to create a series of fitted residuals x_{2t} as above, orthogonalization is accomplished using $m-1$ regressions given by

$$\begin{aligned} \bar{g}_{2t} &= \delta_{20} + \delta_{21}\bar{r}_t + x_{2t} \\ \bar{g}_{3t} &= \delta_{30} + \delta_{31}\bar{r}_t + \delta_{32}\bar{g}_{2t} + x_{3t} \\ &\vdots \\ \bar{g}_{mt} &= \delta_{m0} + \delta_{m1}\bar{r}_t + \delta_{m2}\bar{g}_{2t} + \cdots + \delta_{m,m-1}\bar{g}_{m-1,t} + x_{mt}. \end{aligned} \tag{16}$$

with $\bar{g}_{it} = \langle g_i, f_i \rangle$ for $i = 2, \dots, m$,

$$\begin{bmatrix} \delta_{20} \\ \delta_{21} \end{bmatrix} = \left(\mathbf{E} \begin{bmatrix} 1 \\ \bar{r}_t \\ \bar{r}_t \end{bmatrix} \begin{bmatrix} 1 \\ \bar{r}_t \\ \bar{r}_t \end{bmatrix}' \right)^{-1} \mathbf{E} \begin{bmatrix} 1 \\ \bar{r}_t \\ \bar{r}_t \end{bmatrix} \bar{g}_{2t}$$

analogously to δ_{20} and δ_{21} defined above, and

$$\begin{bmatrix} \delta_{i0} \\ \delta_{i1} \\ \delta_{i2} \\ \vdots \\ \delta_{i,i-1} \end{bmatrix} = \left(\mathbf{E} \begin{bmatrix} 1 \\ \bar{r}_t \\ \bar{g}_{2t} \\ \vdots \\ \bar{g}_{i-1,t} \end{bmatrix} \begin{bmatrix} 1 \\ \bar{r}_t \\ \bar{g}_{2t} \\ \vdots \\ \bar{g}_{i-1,t} \end{bmatrix}' \right)^{-1} \mathbf{E} \begin{bmatrix} 1 \\ \bar{r}_t \\ \bar{g}_{2t} \\ \vdots \\ \bar{g}_{i-1,t} \end{bmatrix}$$

for $i = 3, \dots, m$.

The system in [equation \(16\)](#) may be rewritten more succinctly as

$$x_{2:m,t} = \Delta \langle g_{2:m}, f_t \rangle - \delta_1 \bar{r}_t - \delta_0 \quad (17)$$

by defining $x_{2:m,t} = (x_{2t}, \dots, x_{mt})'$, an $(m-1) \times (m-1)$ matrix given by

$$\Delta = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\delta_{32} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ -\delta_{m2} & \cdots & -\delta_{m,m-1} & 1 \end{bmatrix},$$

and $(m-1) \times 1$ vectors given by $\delta_k = (\delta_{2k}, \dots, \delta_{mk})'$ for $k = 0, 1$.

Model $S0$ is augmented as Model $S1$, generalized from that of [equation \(15\)](#) in the quadratic case to

$$S1: \bar{r}_t = \alpha_{S1,0} + \alpha_{S1,1} h_t + x_{2:m,t}' \beta + \varepsilon_{S1t}, \quad (18)$$

such that $\gamma_{2:m}' = \beta' \Delta$. Model $S1$ is a semiparametric regression where $x_{2:m,t}$ represents the fitted residuals from the regressions in (16) and implied by (17). Parameters $\alpha_{S1,0}$, $\alpha_{S1,1}$, and β may be estimated using least squares.

The final step to obtain X/λ is to identify the γ 's. Analogously to the simpler quadratic example, this is accomplished by setting $\gamma_1 = -\beta' \delta_1$ and $\gamma_{2:m}' = \beta' \Delta$, and the intercept is identified as $\gamma_0 = -\gamma_1 \langle \iota, f \rangle - \gamma_{2:m}' \langle g_{2:m}, f \rangle$, but δ_0 is not used in identification. Once again, $X(r)$ is obtained by dividing the resulting function $\gamma_0 + \gamma_1 r + \gamma_{2:m}' g_{2:m}(r)$ by the climate sensitivity $1/\lambda = \alpha_{S0,1}$ estimated by Model $S0$ in the previous step.

3.1.4. Synthesis: Local Climate Sensitivity, $C(r)$

A brief summary of the three steps for estimation follows.

1. Estimate $f_i(r)$ using a standard density estimation technique, then estimate $df(r)/dh$ by regressing $f_i(r_i)$ onto h_i and an intercept at points $r_i \in [r^-, r^+]$. Retain the estimates of the coefficients on h_i as $df(r_i)/dh$.

2. Estimate $1/\lambda$ by regressing \bar{r}_t onto h_t and an intercept, as in Model S0 in [equation \(12\)](#). Retain the estimate of the coefficient $\alpha_{S0,1}$ on h_t .
3. Estimate $X(r)$.
 - (a) Iteratively regress $\langle g_i, f_t \rangle$ onto an intercept, $\langle g_j, f_t \rangle$ for all $j < i$, and \bar{r}_t , as in [equation \(17\)](#). Retain the estimates of δ_1 , Δ , and $(x_{2:m,i})$.
 - (b) Estimate β using Model S1 in [equation \(18\)](#) by regression.
 - (c) Identify $\gamma_1 = -\beta'\delta_1$, $\gamma'_{2:m} = \beta'\Delta$, and $\gamma_0 = -\gamma_1 \langle \iota, f \rangle - \gamma'_{2:m} \langle g_{2:m}, f \rangle$ with f defined as an average of (f_t) over time.
 - (d) Set $X(r) = (\gamma_0 + \gamma_1 r + \gamma'_{2:m} g_{2:m}(r)) / \alpha_{S0,1}$.

These steps require techniques no more complicated than kernel density estimation and least squares. Once the three components have been estimated, $C(r)$ is assembled from [equation \(11\)](#). Specifically, $C(r_i) = \alpha_{S0,1} - \alpha_{S0,1} X(r_i) (df(r_i) / dh) \epsilon$ gives the LCS near each of n points r_i .

3.2. Time Scales and Dynamic Statistical Models

The discussion thus far has focused on static statistical models. In fact, there are equilibria occurring in two time scales in the static model. Planetary net heat flux is zero on a multidecadal time scale (long-run, in the statistical model), but NHT integrates to zero on a subannual time scale (instantaneous, in the statistical model).

[Held et al. \(2010\)](#) and the IPCC's Fifth Assessment Report (Chapter 10, [Bindoff et al., 2013](#)) differentiate two (additional) time scales in the context of a two-compartment zero-dimensional EBM. Specifically, they identify "fast" and "recalcitrant" components of temperatures. The fast component is proportional to forcings and, based on simulations of a general circulation model, they find that it reverts from 2.5 °C to 0 °C about one decade after an abrupt return of TRF to pre-industrial levels. A first-order autoregressive model is appropriate for a 10-year rate of decay.

To take into account the additional time scales, the static Model S0 is replaced by a dynamic model, Model D0, given by

$$D0 : \bar{r}_t^F = \rho_{D0} \bar{r}_{t-1}^F + \alpha_{D0,0} + \alpha_{D0,1} h_t + \varepsilon_{D0,t}, \quad (19)$$

where $\bar{r}_t^F = \bar{r}_t - d_t$ is the fast component of GMTA and d_t is the recalcitrant component. The additional term $\rho_{D0} \bar{r}_{t-1}^F = \rho_{D0} \langle \iota, f_{t-1}^F \rangle = \rho_{D0} (\langle \iota, f_{t-1} \rangle - d_{t-1}) = \rho_{D0} (\bar{r}_{t-1} - d_{t-1})$ captures prior forcings still relevant to the present fast component.

The coefficient $\alpha_{D0,1}$ gives the immediate response of the fast component to a unit change in TRF. If $d_t = 0$ for all t , there is no recalcitrant component and the long-run (multidecadal) climate sensitivity is estimated by $\alpha_{D0,1} / (1 - \rho_{D0}) \approx \alpha_{S0,1}$. More realistically, some of the upward movement in temperatures that would have been attributed to an increasing TRF will instead be accounted for by an increasing recalcitrant component, so that $\alpha_{D0,1} / (1 - \rho_{D0}) < \alpha_{S0,1}$. In other words, the recalcitrant component tempers the climate sensitivity apparent in the historical record.

A dynamic model given by

$$D1: \bar{r}_t^F = \rho_{D1} \bar{r}_{t-1}^F + \alpha_{D1,0} + \alpha_{D1,1} h_t + x'_{2:m,t} \beta + \varepsilon_{D1t}, \quad (20)$$

is analogous to Model *S1*. In Model *D1*, $x'_{2:m,t}$ is redefined as

$$x'_{2:m,t} = \Delta \langle g_{2:m}, f_t^F \rangle - \delta_1 \bar{r}_t^F - \delta_0 - \delta_{-1} \bar{r}_{t-1}^F \quad (21)$$

so that the residuals are also orthogonal to the autoregressive term in Model *D1*. The parameters β , Δ , etc., are redefined accordingly. The parameters γ_0 , γ_1 , and $\gamma_{2:m}$ are identified in exactly the same way as with the static model with using β , Δ , and δ_1 but using neither δ_0 nor δ_{-1} .

4. DATA AND EMPIRICAL RESULTS

4.1. Data Sources and Construction

HadCRUT5 temperature anomaly data⁵ from Morice et al. (2020) are employed to measure temperature anomalies. Monthly HadCRUT5 data observed over 5° latitude by 5° longitude grid boxes are pooled into years over the time period 1850–2018 ($T = 169$), providing up to $36 \times 72 \times 12 = 31,104$ data points per year for the globe. The base period for the HadCRUT5 anomaly data is 1961–1990.

Annual global data⁶ from Hansen et al. (2017) are used to proxy for spatially uniform TRF. Specifically, TRF is given by the sum of all but volcanic forcings following Estrada et al. (2013b). It is not hard to find deficiencies in the proxy: the effect of aerosols is notoriously difficult to pin down and most of the forcings are far from being spatially uniform.

An alternative would be to use only well-mixed greenhouse gases, which are nearly uniform by virtue of being well-mixed and are measured most precisely. Because there seems to be a consensus in the literature that both temperatures and forcings have a nonstationary (stochastic or deterministic) trend as discussed above, short-lived measurement error should not bias estimation. In this light, including series with short-lived measurement error may be preferable to omitting those series.

Six models are initially compared. The two static models and two dynamic models given by Model *S0*, *S1*, *D0*, and *D1* are included. Further, *S⁺0* and *D⁻0* are included as comparisons with *S0* and *D0*. They are analogous to *S0* and *D0*, but with an accounting of the recalcitrant component d_t added to *S0* and removed from *D0*.

Held et al. (2010) estimate the recalcitrant component of temperature by taking the average of simulated temperatures 10–30 years after an abrupt shutoff of TRF in their general circulation model at years 2000 and 2100 under the projection that CO₂ increases to 720 ppm by 2100. Doing so estimates the pre-industrial recalcitrant component (0 °C), that in 2000 (0.1 °C), and that in 2100 (0.4 °C) under their projection. In the present analysis, the recalcitrant component d_t is

calculated as piecewise linear, passing through the two anomalies above in 2000 and 2100 and through 0 °C in 1880, and then demeaned by the average of the piecewise linear function over the HadCRUT5 base period of 1961–1990, (0.08 °C) so that the recalcitrant component of the anomalies crosses zero between 1975 and 1976.

The recalcitrant component is common to all locations, so that the fast component accounts for all spatial heterogeneity in temperature anomalies. Calculating the entire distribution of the fast component from the entire distribution of anomalies is accomplished by subtracting the recalcitrant component from the support of the distribution. For example, the probably mass of the anomaly at 1 °C in 2000 has a recalcitrant component of $0.1 - 0.08 = 0.02$ °C, so is assigned to the fast component of $1 - 0.02 = 0.98$ °C. Doing so for all anomalies results in a fast component distribution in 2000 which is that of the anomalies in 2000 only shifted to the left by 0.02 °C.

4.2. Main Estimation Results

4.2.1. Step 1: $df(r)/dh$

Densities of the HadCRUT5 temperature anomalies are estimated on a compact support that excludes 1% of the outliers of the distribution of temperature anomalies across the entire time span following [Chang et al. \(2020\)](#) and [Miller and Nam \(2020\)](#). The support is then adjusted to also accommodate densities of the fast component as discussed above, so that the same support is used for both densities across all years. The support employed is $[-5.40, 5.58]$ with increments of $\epsilon = 0.005$, so that the sample size $n + 1$ is $(r^+ - r^-)/\epsilon + 1 = 2,197$ for each year. Density estimation is accomplished using a Bartlett kernel density estimator with Silverman bandwidth.

[Fig. 1](#) compares three series: (a) the published HadCRUT5 ensemble mean GMTA, (b) GMTA using the global distributions estimated from disaggregated HadCRUT5 data, and (c) the global mean fast components calculated using the estimated fast component distributions. The figure shows that the distribution estimates yield means that align well with the published means, and that the fast component accounts for nearly all of the temperature change over the observational record, as expected. The appendix contains a detailed comparison of constructions of the published and estimated means.

[Fig. 2](#) shows estimates of the derivative $df(r)/dh$ from Step 1 for both the temperature anomaly and fast component distributions, along with 95% confidence intervals created using a first-order sieve bootstrap with 999 replications. The sieve allows for temporal correlation at specific anomalies r_i . The residuals of the sieve are drawn contemporaneously, so that any cross-sectional correlation is preserved.

The shape that these derivatives display is sensible. Any change in a density must sum to zero along its domain, because both the original density and the density that results from the change must sum to unity. A change in TRF does not seem to affect the probability of observing the most extreme positive or negative outliers in an absolute sense: they are still outliers with low probabilities of occurring.

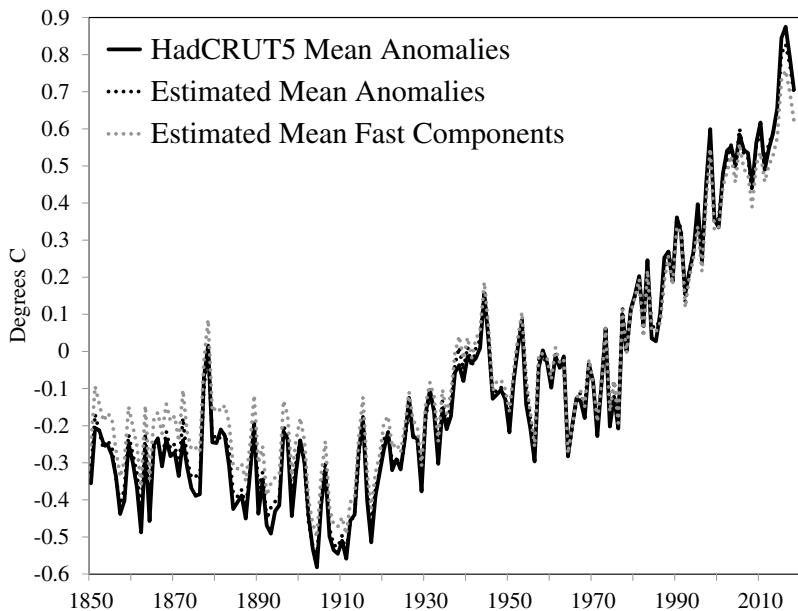


Fig. 1. HadCRUT5 and Estimated Means. Notes: *HadCRUT5 Mean Anomalies* denotes published HadCRUT5 ensemble mean GMTA. *Estimated Mean Anomalies* (\bar{r}_t) denote estimates of the HadCRUT5 GMTA using the global distribution of anomalies. *Estimated Mean Fast Components* (\bar{r}_t^F) are analogous to Estimated Mean Anomalies but are constructed using the global distribution of fast components.

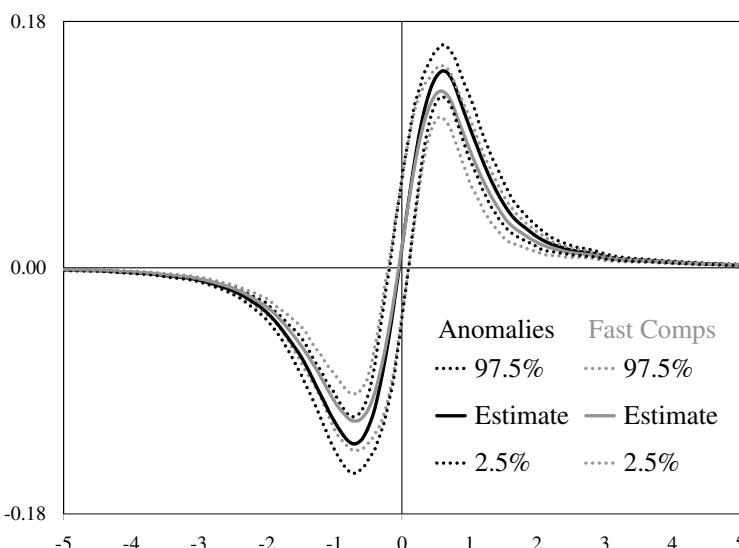


Fig. 2. Estimates of $df(r)/dh$. Notes: Estimated using temperature anomalies and their fast components, with 95% uncertainty intervals constructed using a first-order sieve bootstrap as described in the text.

Over the domain of $df(r)/dh$, the change in observing negative anomalies is mostly decreasing in TRF and the change in observing positive anomalies is entirely increasing in TRF, consistent with GMTA increasing with TRF – i.e., consistent with global warming. The set of derivatives for the fast components generally have smaller magnitudes than those of the anomalies because the fast components are less sensitive to TRF by design. This difference is not statistically significant, however.

4.2.2. Step 2: Climate Sensitivity, $1/\lambda$

Climate sensitivity estimates, or more precisely global mean climate sensitivity estimates, are shown in Table 1. The least squares estimate of α_1 from estimating Model S0 is given by $0.434 \text{ } ^\circ\text{C}/(\text{W/m}^2)$, which is comparable to estimates in the extant literature. For example, [Estrada et al. \(2013b\)](#) obtain $0.40 \text{ } ^\circ\text{C}/(\text{W/m}^2)$ using a filtered sample mean series from an earlier version of the HadCRUT data.

The TCR, given by $h_{2\times}d\bar{\tau}/dh$, is estimated to be $3.71 \times 0.434 = 1.610 \text{ } ^\circ\text{C}$ with an interval estimate of $(1.466, 1.750) \text{ } ^\circ\text{C}$ using Model S0, well within the IPCC's "likely" range of $1\text{--}2.5 \text{ } ^\circ\text{C}$ ([Bindoff et al., 2013](#)). Model $S0^+$, which is the static model including the recalcitrant component d_r , gives a global mean climate sensitivity of $0.376 \text{ } ^\circ\text{C}/(\text{W/m}^2)$, which is lower but still with a plausible TCR of $1.394 \text{ } ^\circ\text{C}$.

The dynamic models allow muted responses of $0.180 \text{ } ^\circ\text{C}/(\text{W/m}^2)$ and $0.154 \text{ } ^\circ\text{C}/(\text{W/m}^2)$ after only one year, but the long-run responses, $0.448 \text{ } ^\circ\text{C}/(\text{W/m}^2)$ and $0.389 \text{ } ^\circ\text{C}/(\text{W/m}^2)$, are quite close to those of the static models as they should be. The TCRs from the dynamic models are $1.662 \text{ } ^\circ\text{C}$ for Model D^-0 and $1.443 \text{ } ^\circ\text{C}$ for Model $D0$, quite similar to those from the corresponding static models and well within margins of error.

Notice that an increasing recalcitrant temperature component decreases the climate sensitivity and therefore the TCR, because the recalcitrant component is invariant with respect to forcings on a short time scale. The intermediate Models S^+0 and D^-0 will not be considered henceforth because there is little difference between the long-run results from Models S0 and D^-0 and those from Models D0 and S^+0 .

Table 1. Estimates of Climate Sensitivity $1/\lambda$ and TCR.

Model	α_1	CS in $^\circ\text{C}/(\text{W/m}^2)$	95% Conf. Int.	TCR in $^\circ\text{C}$	95% Conf. Int.
S0	0.434	0.434	(0.395, 0.472)	1.610	(1.466, 1.750)
S^+0	0.376	0.376	(0.336, 0.415)	1.394	(1.248, 1.539)
D^-0	0.180	0.448	(0.381, 0.528)	1.662	(1.413, 1.959)
D0	0.154	0.389	(0.322, 0.473)	1.443	(1.193, 1.754)

Notes: Least squares estimates from models S0, S^+0 , D^-0 , and D0. Climate sensitivity (labeled CS) is calculated as $\alpha_1 / (1 - \rho)$ with $\rho = 0$ for static models, with 95% uncertainty intervals constructed using a first-order sieve bootstrap as described in the text.

4.2.3. Step 3: NHT Function, $X(r)$

The flexible Fourier functional form, which approximates g by a series of polynomial and trigonometric functions, has been used to estimate semiparametric cointegrating regressions under different assumptions (Park et al., 2010; Park & Hahn, 1999). This form may be written as

$$g_j^S(s) = \begin{cases} s^j & \text{for } j = 2, \dots, p \\ \cos 2\pi ks & \text{for } j = p + 2k - 1 \text{ and } k = 1, \dots, q \\ \sin 2\pi ks & \text{for } j = p + 2k \text{ and } k = 1, \dots, q \end{cases}$$

for $s \in [0, 1]$. Using this notation, $m = p + 2q$.

The functions g_j^S must be defined over the unit interval, so let $g_j(r) = (r^+ - r^-)g_j^S((r - r^-)/(r^+ - r^-))$. Thus,

$$\langle g_j(r), f_t(r) \rangle = \langle (r^+ - r^-)g_j^S((r - r^-)/(r^+ - r^-)), f_t(r) \rangle$$

holds, so that the inner products in equations (17) and (21) can be evaluated even though g_j^S is defined on the unit interval. Table 2 shows estimates of δ_1 and Δ from Models S1 and D1 in equations (17) and (21). Because the point of these regressions is orthogonalization rather than estimation, standard errors are not reported.

Next, β in Model S1 in equation (18) or Model D1 in equation (20) is estimated using least squares. Selection of p and q may be accomplished using an information criterion, such as Bayesian (BIC) or Hannan-Quinn (HQ). Setting the maximum of (p, q) to be $(2, 2)$, or $m = 6$, BIC and HQ are minimized at $(2, 0)$ and $(1, 1)$ respectively for Model S1. The results presented here use the HQ choice $(1, 1)$, or $m = 3$ (no polynomial terms, a single set of periodic functions), which are qualitatively similar to those using $(2, 0)$. The specification is fixed to $(1, 1)$ in Model D1 for comparison with Model S1.

Estimates from Models S1 and D1 in equations (18) and (20) are shown in Table 3. Standard errors given in the table should be interpreted with caution, because they are not robust to nuisance parameters that may result from trending data. More robust and meaningful interval estimates are presented subsequently. Estimates of α_1 are numerically comparable to those obtained in the previous

Table 2. Estimates of $-\delta_1$ and Δ .

	Model S1			Model D1		
	$-\delta_1$	Δ		$-\delta_1$	Δ	
x_{2t}	0.295	1.000	0.000	0.203	1.000	0.000
x_{3t}	4.987	0.091	1.000	4.946	0.100	1.000

Notes: Least squares estimates based on the regressions implied by equations (17) and (21) with residuals denoted by x_{2t} and x_{3t} . Estimates of δ_0 and δ_{-1} are not displayed because they are not used for identification.

Table 3. Estimates of Models S1 and D1.

	Model S1		Model D1	
	est.	s.e.	est.	s.e.
ρ	—	—	0.546	0.064
α_0	-0.352	0.012	-0.132	0.021
α_1	0.440	0.013	0.180	0.026
β_2	0.104	0.031	0.065	0.029
β_3	-0.192	0.097	-0.120	0.083

Notes: Least squares estimates based on the regressions in (18) and (20).

step, but these are not intended to estimate climate sensitivity. Estimates of climate sensitivity are still given by those in Table 1.

Estimates of β , δ_1 , and Δ in Tables 2 and 3 are used to identify $\gamma_1 = -\beta'\delta_1$ and $\gamma'_{2:m} = \beta'\Delta$, and then $\gamma_0 = -\gamma_1 \langle \iota, f \rangle - \gamma'_{2:m} \langle g_{2:m}, f \rangle$. Finally, $\gamma_1 r + \gamma'_{2:m} g_{2:m}(r)$ is identified by multiplying $\gamma_1 s$ and $g_j^S(s)$ by $(r^+ - r^-)$ to recover r and $g_j(r)$ for $j = 2, \dots, m$.

Before proceeding and in light of the multiple steps leading up to this point, it is useful to consider uncertainty in estimation of the NHT and thus the LCS. Uncertainty in estimation of $df(r)/dh$ and $1/\lambda$ were both dealt with by way of a first-order sieve bootstrap with 999 replications, so it is natural to deal with uncertainty in estimating β and thus γ_0 , γ_1 , and $\gamma'_{2:m}$ in a similar way.

The bootstrap strategy redraws the fitted residuals from equation (18) or (20). However, the terms in $x_{2:m,t}$ are not re-orthogonalized with respect to the bootstrapped regressand. An estimate of $X(r)$ also requires an estimate of $\alpha_{S0,1} = 1/\lambda$, so the respective bootstrap samples must be drawn jointly. In other words, pseudo-samples are drawn using the same random seed for $X(r)$ as for $\alpha_{S0,1}$ in the previous step.

Fig. 3 shows the NHT function $X(r)$ in W/m^2 . In spite of the uncertainty, $X(r)$ is clearly nonconstant and nonlinear, inconsistent with a zero-dimensional EBM (homogenous planet). However, as discussed above, the nonlinearity in r is *not* inconsistent with the one-dimensional Budyko-Sellers EBM. Because there appears to be little statistical difference between the Models S1 and D1, but the former is estimated more precisely, we henceforth focus exclusively on the results from Model S1.

The magnitudes of the point estimates of the NHT function, on the order of $\pm 10 \text{ W/m}^2$, are considerably less than those estimated in the extant literature. For example, Trenberth et al. (2001) estimate positive zonal averages up to 50 W/m^2 and negative zonal averages down to about -125 W/m^2 . The smaller magnitudes of the present analysis are likely attributable to the aggregation of locations and months with the same temperature anomaly to create the temperature distribution for the statistical analysis.

Fig. 3 is easy to interpret as the NHT function, but not as easy to read, because outlying values of the graph dominate. With this in mind, Fig. 4 shows $X(r)f(r)$, which may be interpreted roughly as NHT at a given temperature anomaly weighted by the probability of observing that anomaly. Under this weighting, it

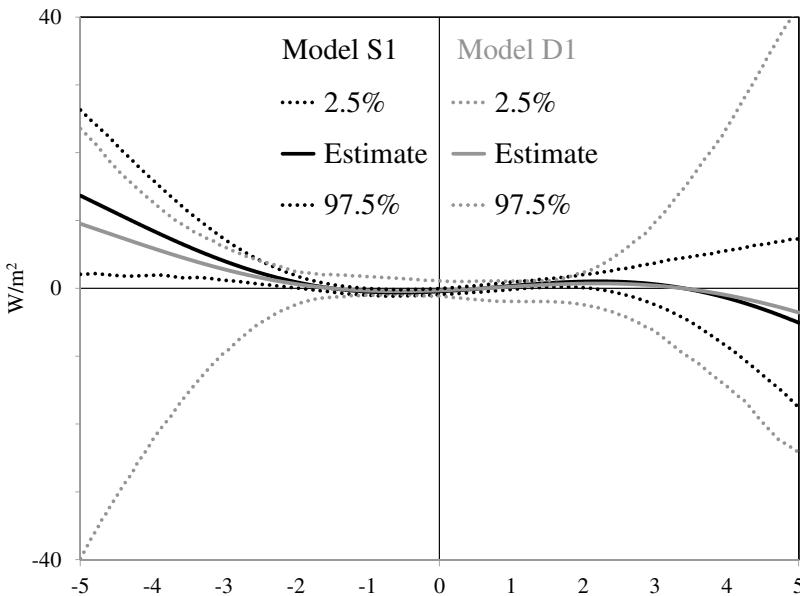


Fig. 3. Estimates of NHT $X(r)$. Notes: Estimated using Model S1 and Model D1, with 95% uncertainty intervals constructed using a first-order sieve bootstrap as described in the text.

is easier to see a statistically significant difference between the estimated function and a constant or otherwise linear function.

It is clear from Fig. 4 that coldest anomalies correlate with positive (outgoing) NHT – i.e., these areas tend to export energy. In contrast, moderately negative or positive anomalies, over roughly -1.5°C to 0.6°C , correlate with negative (incoming) NHT – these areas tend to import energy. Over roughly 0.6°C to 3.0°C , NHT is positive again but no longer statistically significant beyond 2.0°C .

Fig. 5 shows a map of HadCRUT5 temperature anomalies averaged over 120 months, 2009–2018. Fig. 6 maps the locationless NHT function $X(r)$ in Fig. 3 to locations in Fig. 5 based on the average anomaly component over this period. NHT over most of the globe are estimated to be less than 1.0 W/m^2 in absolute value. That this number is quite low in comparison with those in Fig. 3 above results from the averaging of the anomalies over time used to generate the figure. Very few areas have sustained anomalies of a particularly large magnitude.

Areas with positive NHT tend to be those with warmer or near-zero anomalies in Fig. 5. This tendency seems counter-intuitive in light of the point estimate of the NHT function in Fig. 3, which is mostly downward sloping. However, the finding results from the NHT function being negative between about -1.5°C and 0.6°C and positive between about 0.6°C and 3.0°C . There are very few observations more extreme than -1.5°C or 3.0°C . This finding may be interpreted to mean that while energy exporting areas are warming up, energy importing areas are not warming up as fast. Such an interpretation is reasonable because the capacity of

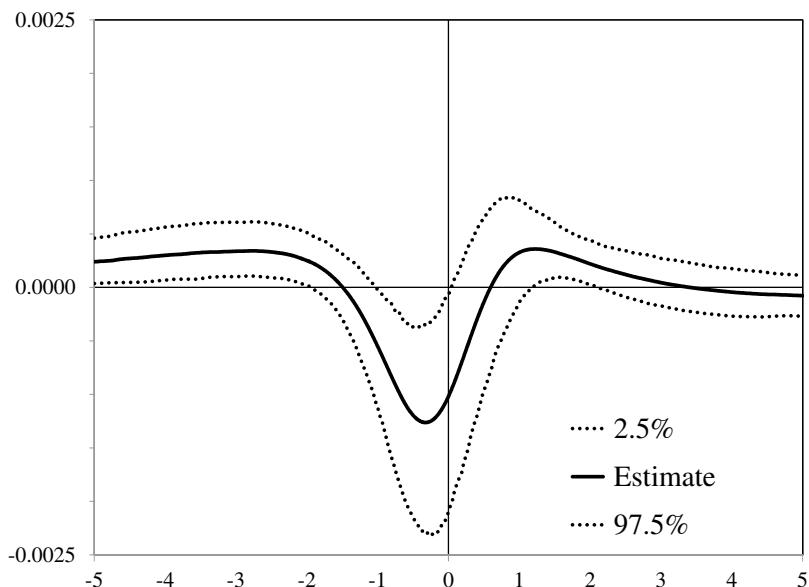


Fig. 4. Estimates of $X(r)f(r)$. Notes: Estimates using Model S1, with 95% uncertainty interval constructed using first-order sieve bootstrap as described in the text.

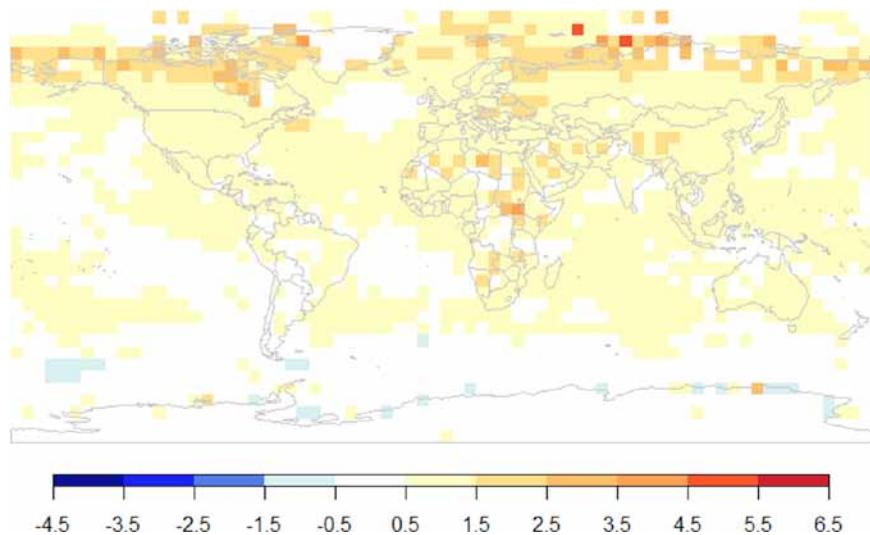


Fig. 5. Temperature Anomalies, in $^{\circ}\text{C}$. Notes: 10-Year average over 2009–2018.

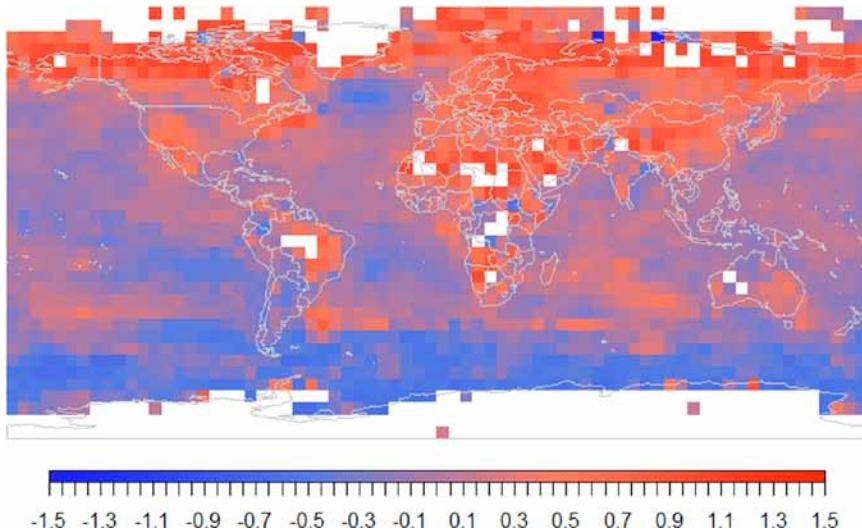


Fig. 6. Estimates of NHT $X(r)$ from Model S1 Mapped to Location, 2009–2018, in W/m^2 . Notes: White cells denote either missing anomaly data from Fig. 5 (most) or $X(r)$ estimated to be outside of $[-1.5, 1.5] \text{ W/m}^2$ (few).

areas to export energy is bounded by their geographical features, which do not change over time. Exports cannot keep up with forcings, so areas that export energy are warming up the most.

NHT tends to be positive over continental landmasses, so that landmasses generally export energy to the oceans. Trenberth et al. (2001) found the net effect to be the opposite over a longer and earlier time period, citing the strength of the ocean's moderation of temperatures over land during the Northern Hemisphere winter to that during the summer. On the other hand, evidence of a divide in the Southern Hemisphere between the cold Antarctic Circumpolar Current and the warm currents of the southern gyres of the South Pacific, South Atlantic, and Indian Oceans is roughly consistent with their findings. Fig. 6 does not show a similar divide in the Northern Hemisphere. Estimated NHT conforms more to the asymmetry of landmasses across the hemispheres than to symmetry in latitude imposed by traditional one-dimensional EBMs.

4.2.4. Synthesis: Local Climate Sensitivity, $C(r)$

Finally, the steps are synthesized to estimate LCS $C(r)$. Fig. 7 shows the result using Model S1. Results using Model D1 (not shown) mostly lie within the uncertainty interval for Model S1 but have a much wider uncertainty interval around them. The 95% confidence intervals are constructed by drawing from the same empirical error distributions resulting from first-order sieve bootstraps as discussed above, but it is important to fix the random seed, so that the bootstrapped distribution of $C(r)$ preserves correlations between the bootstrapped distributions of the individual components.

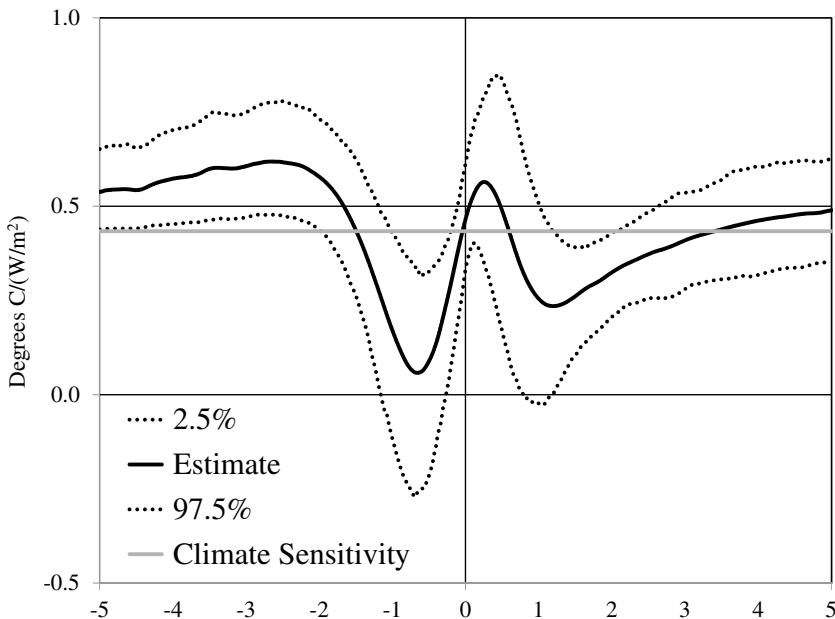


Fig. 7. Estimates of LCS $C(r)$, in $^{\circ}\text{C}/(\text{W}/\text{m}^2)$. Notes: Estimated using Model S1, with 95% uncertainty interval constructed using a first-order sieve bootstrap as described in the text. For reference, the global mean climate sensitivity estimated to be $1/\lambda = 0.434 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$ is plotted.

As a reference, recall that the point estimate of global mean climate sensitivity estimated in Step 2 is $0.434 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$ and the mean LCS given here is $0.440 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$. The difference $0.006 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$ is attributable to variation in $X(r)$ over time. As discussed above, a time-varying intercept may be identified, so that the last term in equation (10) is not assumed to be zero. Because this difference is well within the margin of error of estimating climate sensitivity and does not have very much impact on the shape of the LCS, the assumption of time-invariance is plausible.

LCS exceeds the global mean climate sensitivity at particularly cold anomalies, decreasing dramatically as the temperature anomaly warms with a trough at about $-0.667 \, ^{\circ}\text{C}$. The point estimate of the nadir exceeds $0 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$, but negative LCSs are within the range of uncertainty. LCS then increases dramatically and crosses the global mean climate sensitivity at an anomaly close to zero and peaks (again) just above zero. The horizontal distance between this peak and the trough is on the order of $1 \, ^{\circ}\text{C}$. The LCS has one more trough and then increases to the level of the global mean again. The total range estimated by the point estimates from Model S1 is $0.058 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$ to $0.564 \, ^{\circ}\text{C}/(\text{W}/\text{m}^2)$.

Similar to Fig. 6, Fig. 8 presents LCS geographically. Some striking patterns emerge. Areas that were light or dark red in Fig. 6 tend to be light gray in Fig. 8, meaning that areas with a positive energy balance – that export rather than import

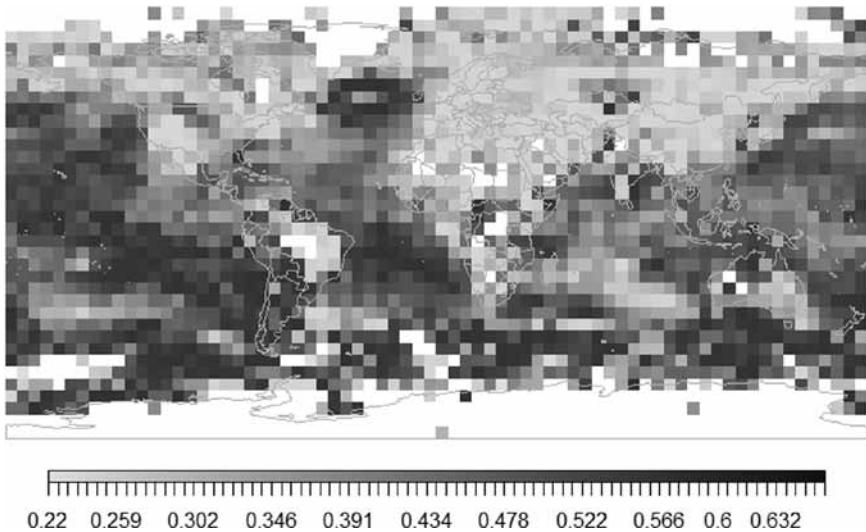


Fig. 8. Estimates of $C(r)$ from Model S1 Mapped to Location, 2009–2018, in $^{\circ}\text{C}/(\text{W}/\text{m}^2)$. Notes: White cells denote either missing anomaly data from Fig. 5 (most) or $C(r)$ estimated to be outside $[0.22, 0.66] ^{\circ}\text{C}/(\text{W}/\text{m}^2)$ (a few smaller; none larger).

energy – are less sensitive to TRF. The most sensitive areas, darker gray in Fig. 8, tend to be energy importers, blue in Fig. 6. This relationship is consistent with the minus sign in front of $X(r)$ in equation (11) derived from the physical model.

The difference between positive and negative NHT is especially noticeable in the ocean regions. Most are more sensitive than average, with a few large “islands” of insensitivity, particularly noticeable in the southern parts of the Indian, Pacific, and Atlantic, but also at mid-latitudes of the Western coast of Mexico and Eastern coast of Canada.

Going back to Fig. 5, those areas that are the most sensitive to TRF appear to have experienced no more than a $0.5 ^{\circ}\text{C}$ increase over the base period, which seems counter-intuitive. Note especially the Atlantic regions off the coasts of Greenland and Brazil. In contrast, the areas that have experienced increases of more than $0.5 ^{\circ}\text{C}$ may have no more than average sensitivity. This is possible because the most sensitive areas tend to be over oceans, which are net importers of energy. Manabe et al. (1991), Sejas et al. (2014), *inter alia* attribute the disparity in temperature responses over ocean and land to ocean heat storage and especially to evaporation over the ocean, both of which ameliorate warming.

The LCS estimated here from the historical record is somewhat comparable to the negative inverse of the measure of climate sensitivity mapped by Boer and Yu (2003, Fig. 3) using simulated values from an equilibrium over 500 years in the future. They obtain $1/\lambda$ roughly twice as large as the value $0.434 ^{\circ}\text{C}/(\text{W}/\text{m}^2)$ obtained here and those used or obtained by a number of other authors. It is difficult to compare the magnitudes of their measure of local climate sensitivity with those presented here, because the inverse creates an asymptote near zero, but the signs may be compared.

Most notably, Boer and Yu (2003) also find the oceans to be relatively sensitive but with a pocket of negative sensitivity in the Central Pacific and negative values particularly to the south and around Antarctica. Sensitivities around Antarctica, given by about $-1/7 = -0.14 \text{ }^{\circ}\text{C}/(\text{W/m}^2)$ in some localities in their analysis, are roughly comparable to the low end of the uncertainty interval of sensitivities in the present analysis that map to similar regions. Boer and Yu (2003) also find the area around the North Pole, Hudson and Baffin Bays, and Central Asia to be negatively sensitive. In the present analysis using historical data rather than simulated data, these areas have positive sensitivity though some are quite low.

5. CONCLUDING REMARKS

The statistical methodology proposed in this paper estimates local climate sensitivity, a disaggregation of climate sensitivity as defined by Boer and Yu (2003) *inter alia*, using well-known aggregated forcings and disaggregated temperature anomaly data sets over the historical record. The underlying physical model is a hybrid of existing one- and two-dimensional energy balance models, in the sense that the data are subject to some aggregation similarly to one-dimensional EBMs, but that variation within latitudes is allowed similarly to two-dimensional EBMs.

Both static and dynamic statistical models are proposed and estimated, with some specifications that also allow for the decomposition of temperatures into recalcitrant and fast components, along the lines of Held et al. (2010) and Bindoff et al. (2013). However, as those authors note, the recalcitrant component does not contribute substantively over the historical record, so our main results are based on models that do not distinguish between these components. The methodology is semiparametric but each step employs techniques no more complicated than least squares or kernel density estimation.

Estimates of local horizontal NHT are intermediate to those of the LCS estimates. The estimated NHT function is strongly nonconstant and nonlinear in temperature anomalies, which is not inconsistent with the linear one-dimensional Budyko-Sellers EBM, because temperature anomalies rather than levels are used.

Areas of the globe with estimated positive NHT tend to be those with warmer anomalies. These areas also tend to be over land, suggesting that landmasses are net exporters of energy to oceans. Some similarities emerge with NHT mapped in the extant literature (Trenberth et al., 2001), especially in the Southern Hemisphere.

In turn, local climate sensitivity is estimated with a global mean climate sensitivity of $0.434 \text{ }^{\circ}\text{C}/(\text{W/m}^2)$ which is consistent with the extant literature (Estrada et al., 2013b). The total range over the globe is estimated to be $0.058 \text{ }^{\circ}\text{C}/(\text{W/m}^2)$ to $0.564 \text{ }^{\circ}\text{C}/(\text{W/m}^2)$, but mostly above about $0.22 \text{ }^{\circ}\text{C}/(\text{W/m}^2)$. The smaller values are roughly comparable to those of Boer and Yu (2003), as are the regions over which the sensitivity is very low or possibly negative.

The heterogeneity of local climate sensitivity detected and estimated here has broad economic implications. Polar amplification allowed by such heterogeneity drives more severe weather at mid-latitudes (Francis & Vavrus, 2012; IPCC, 2014; Liu et al., 2012), implying an increase in economic damages in addition to the

rising mean predicted by zero-dimensional EBMs. Furthermore, [Brock and Xepapadeas \(2017\)](#) emphasize the differences in economic costs of climate change and bias in mitigation policies from ignoring heterogeneity in NHT.

Future research linking local climate sensitivity to local damage functions, such as those used by [Miller and Brock \(2021\)](#) based on those of [Gasparini et al. \(2015\)](#) for relative risk of mortality or those of [Hsiang et al. \(2017\)](#) for a number of climate-induced damages, may further scientific knowledge about the heterogeneity and economic disparity expected from climate change.

NOTES

1. [Chapman et al. \(2013\)](#) use *local* trends in conjunction with climate sensitivity to describe heterogeneous climate responses in °C. In this chapter, the terminology applies to local variations in the parameter $1/\lambda$, but of course these also imply differing local trends.

2. These two sets of authors write $N(\bar{\tau})$ and h in levels, because the steady-state unforced planetary net heat flux is zero. For small changes $dN(\bar{\tau}) \approx \Delta N(\bar{\tau}) = N(\bar{\tau}) - 0$ and $dh \approx \Delta h = h - 0$.

3. The approximation error comes from weighting all grid boxes equally to obtain [equation \(8\)](#) as discussed in the Appendix. The plots labeled *HadCRUT5 Mean Anomalies* and *Estimated Mean Anomalies* in [Fig. 1](#) show estimates of GMTA using equal weights to be very similar to those weighted by latitude.

4. Empirical evidence not shown allowing for time-variation in the estimation scheme discussed below suggests that $dX / dh \approx 0$ for all r , so this assumption seems reasonable.

5. Ensemble mean of HadCRUT5.0.0.0 downloaded from <https://www.metoffice.gov.uk/hadobs/hadcrut5/index.html> on December 29, 2020.

6. Downloaded from Makiko Sato's webpage at <http://www.columbia.edu/~mhs119> on December 29, 2020.

ACKNOWLEDGMENTS

I appreciate insightful comments from Myles Allen, William A. (Buz) Brock, Yoosoon Chang, David Hendry, Joon Park, and participants of the 2016 Conference on Econometric Models of Climate Change (Aarhus), and the 2016 International Conference of Computational and Financial Econometrics (Seville), 2017 North American Summer Meeting of the Econometric Society (Washington University in St. Louis), the 2017 Midwest Econometrics Group Meeting (Texas A&M), and a colloquium at the University of Missouri. I am grateful to Kyungsik Nam, whose dissertation research provided some new insights into estimating the semiparametric cointegrating models in the paper. All errors are mine.

REFERENCES

- Bindoff, N. L., Stott, P. A., AchutaRao, K. M., Allen, M. R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I. I., Overland, J., Perlitz, J., Sebbari, R., & Zhang, X. (2013). Detection and attribution of climate change: From global to regional. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I*

- to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 867–952). Cambridge University Press.
- Boer, G. J., & Yu, B. (2003). Climate sensitivity and response. *Climate Dynamics*, 20, 415–429.
- Brock, W. A., & Xepapadeas, A. (2017). Climate change policy under polar amplification. *European Economic Review*, 94, 263–282.
- Budyko, M. I. (1969). The effect of solar radiation variations on the climate of the Earth. *Tellus*, 21, 611–619.
- Castruccio, S., McInerney, D. J., Stein, M. L., Crouch, F. L., Jacob, R. L., & Moyer, E. J. (2014). Statistical emulation of climate model projections based on precomputed GCM runs. *Journal of Climate*, 27, 1829–1844.
- Chang, Y., Choi, Y., Kim, C. S., Miller, J. I., & Park, J. Y. (2016). Disentangling temporal patterns in elasticities: A functional coefficient panel analysis of electricity demand. *Energy Economics*, 60, 232–243.
- Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics*, 214, 274–294.
- Chang, Y., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2014). Time-varying long-run income and output elasticities of electricity demand. *Energy Economics*, 46, 334–347.
- Chang, Y., Kim, C. S., Miller, J. I., Park, J. Y., & Park, S. (2016). A new approach to modeling the effects of temperature fluctuations on monthly electricity demand. *Energy Economics*, 60, 206–216.
- Chang, Y., Kim, C. S., & Park, J. Y. (2016). Nonstationarity in time series of state densities. *Journal of Econometrics*, 192, 152–167.
- Chapman, S. C., Stainforth, D. A., & Watkins, N. W. (2013). On estimating local long-term climate trends. *Philosophical Transactions of the Royal Society A*, 371, 20120287.
- Eroğlu, B. A., Miller, J. I., & Yigit, T. (2022). Time-varying cointegration and the Kalman filter. *Econometric Reviews*, 41, 1–21.
- Estrada, F., & Perron, P. (2014). Detection and attribution of climate change through econometric methods. *Boletín de la Sociedad Matemática Mexicana*, 20, 107–136.
- Estrada, F., Perron, P., Gay-García, C., & Martínez-López, B. (2013a). A time-series analysis of the 20th century climate simulations produced for the IPCC's fourth assessment report. *PLoS ONE*, 8, e60017.
- Estrada, F., Perron, P., & Martínez-López, B. (2013b). Statistically derived contributions of diverse human influences to twentieth-century temperature changes. *Nature Geo-Science*, 6, 1050–1055.
- Francis, J. A., & Vavrus, S. J. (2012). Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophysical Research Letters*, 39, L06801.
- Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., Leone, M., De Sario, M., Bell, M. L., Guo, Y.-L. L., Wu, C., Kan, H., Yi, S.-M., Coelho, M. de S. Z. S., Saldiva, P. H. N., ... Armstrong, B. (2015). Mortality risk attributable to high and low ambient temperature: A multicountry observational study. *The Lancet*, 386, 369–375.
- Gregory, J. M., & Forster, P. M. (2008). Transient climate response estimated from radiative forcing and observed temperature change. *Journal of Geophysical Research*, 113, D23105.
- Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K., Beerling, D. J., Cao, J., Marcott, S., Masson-Delmotte, V., Prather, M. J., Rohling, E. J., Shakun, J., Smith, P., Lacis, A., Russell, G., & Ruedy, R. (2017). Young people's burden: Requirement of negative CO₂ emissions. *Earth System Dynamics*, 8, 577–616.
- Held, I. M., & Suarez, M. J. (1974). Simple albedo feedback models of the icecaps. *Tellus*, 26, 613–629.
- Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, 23, 2418–2427.
- Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., Rasmussen, D. J., Muir-Wood, R., Wilson, P., Oppenheimer, M., Larsen, K., & Houser, T. (2017). Estimating economic damage from climate change in the United States. *Science*, 356, 1362–1369.

- IPCC. (2014). *Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]*. IPCC.
- Kaufmann, R. K., Kauppi, H., Mann, M. L., & Stock, J. H. (2013). Does temperature contain a stochastic trend: Linking statistical results to physical mechanisms. *Climatic Change*, 118, 729–743.
- Kaufmann, R. K., Kauppi, H., & Stock, J. H. (2006a). Emissions, concentrations and temperature: A time series analysis. *Climatic Change*, 77, 249–278.
- Kaufmann, R. K., Kauppi, H., & Stock, J. H. (2006b). The relationship between radiative forcing and temperature: What do statistical analyses of the instrumental temperature record measure? *Climatic Change*, 77, 279–289.
- Kaufmann, R. K., Kauppi, H., & Stock, J. H. (2010). Does temperature contain a stochastic trend? Evaluating conflicting statistical results. *Climatic Change*, 101, 395–405.
- Kaufmann, R. K., & Stern, D. I. (2002). Cointegration analysis of hemispheric temperature relations. *Journal of Geophysical Research*, 107, 4012.
- Langen, P. L., & Alexeev, V. I. (2007). Polar amplification as a preferred response in an idealized aqua-planet GCM. *Climate Dynamics*, 29, 305–317.
- Leduc, M., Matthews, H. D., & de Elia, R. (2016). Regional estimates of the transient climate response to cumulative CO₂ emissions. *Nature Climate Change*, 6, 474–478.
- Liu, J., Curry, J. A., Wang, H., Song, M., & Horton, R. M. (2012). Impact of declining Arctic sea ice on winter snowfall. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 4074–4079.
- Manabe, S., Stouffer, R. J., Spelman, M. J., & Bryan, K. (1991). Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO₂, Part I: Annual mean response. *Journal of Climate*, 4, 785–818.
- Merlis, T. M., & Henry, M. (2018). Simple estimates of polar amplification in moist diffusive energy balance models. *Journal of Climate*, 31, 5811–5824.
- Miller, J. I., & Brock, W. A. (2021). Beyond RCP8.5: Marginal mitigation using quasi-representative concentration pathways. *Journal of Econometrics*. Forthcoming.
- Miller, J. I., & Nam, K. (2020). Dating hiatuses: A statistical model of the recent slowdown in global warming and the next one. *Earth System Dynamics*, 11, 1123–1132.
- Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., & Simpson, I. R. (2020). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 dataset. *Journal of Geophysical Research*. In press.
- Nam, K. (2018). *Essays on climate econometrics* [PhD dissertation]. University of Missouri.
- North, G. R. (1975). Theory of energy-balance climate models. *Journal of Atmospheric Sciences*, 32, 2033–2043.
- North, G. R., Cahalan, R. F., & Coakley, J. A., Jr. (1981). Energy balance climate models. *Reviews of Geophysics and Space Physics*, 19, 91–121.
- North, G. R., & Cahalan, R. F. (1981). Predictability in a solvable stochastic climate model. *Journal of the Atmospheric Sciences*, 38, 504–513.
- North, G. R., Mengel, J. G., & Short, D. A. (1983). Simple energy balance model resolving the seasons and the continents: Application to the astronomical theory of the ice ages. *Journal of Geophysical Research*, 88, 6576–6586.
- Park, J. Y., & Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory*, 15, 664–703.
- Park, J. Y., & Qian, J. (2012). Functional regression of continuous state distributions. *Journal of Econometrics*, 167, 397–412.
- Park, J. Y., Shin, K., & Whang, Y. J. (2010). A semiparametric cointegrating regression: Investigating the effects of age distributions on consumption and saving. *Journal of Econometrics*, 157, 165–178.
- Peixoto, J. P., & Oort, A. H. (1992). *Physics of climate*. American Institute of Physics.
- Pretis, F. (2020). Econometric models of climate systems: The equivalence of energy balance models and cointegrated vector autoregressions. *Journal of Econometrics*, 214, 256–273.
- Schwartz, S. E. (2012). Determination of Earth's transient and equilibrium climate sensitivities from observations over the twentieth century: Strong dependence on assumed forcing. *Surveys in Geophysics*, 33, 745–777.

- Screen, J. A., & Simmonds, I. (2013). Exploring links between Arctic amplification and mid-latitude weather. *Geophysical Research Letters*, 40, 959–964.
- Sejas, S. A., Albert, O. S., Cai, M., & Deng, Y. (2014). Feedback attribution of the land-sea warming contrast in a global warming simulation of the NCAR CCSM4. *Environmental Research Letters*, 9, 124005.
- Sellers, W. D. (1969). A global climatic model based on the energy balance of the earth-atmosphere system. *Journal of Applied Meteorology*, 8, 392–400.
- Sellers, W. D. (1976). A two-dimensional global climatic model. *Monthly Weather Review*, 104, 233–248.
- Siler, N., Roe, G. H., & Armour, K. C. (2018). Insights into the zonal-mean response of the hydrologic cycle to global warming from a diffusive energy balance model. *Journal of Climate*, 31, 7481–7493.
- Stern, D. I., & Kaufmann, R. K. (2000). Detecting a global warming signal in hemispheric temperature series: A structural time series analysis. *Climatic Change*, 47, 411–438.
- Taylor, F. W. (2005). *Elementary climate physics*. Oxford University Press.
- Trenberth, K. E., Caron, J. M., & Stepaniak, D. P. (2001). The atmospheric energy budget and implications for surface fluxes and ocean heat transports. *Climate Dynamics*, 17, 259–276.

APPENDIX

A.1. A DISCRETE REPRESENTATION OF THE EBM

Hemispheric mean temperature in a model that varies only over latitude is defined as

$$\bar{\tau} = \int_0^{\pi/2} \tau_\theta \cos \theta d\theta$$

(cf. [Held & Suarez, 1974](#)), where the cosine adjusts for the curvature of the Earth. For temperature that also varies over longitude, the global average is

$$\bar{\tau} = \frac{1}{4\pi} \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2} \tau_{\theta\varphi} \cos \theta d\theta d\varphi,$$

with the convention that τ_{00} is the temperature where the Prime Meridian intersects the Equator. Division by 2π converts longitude from radians and division by 2 averages the hemispheric means evaluated over positive and negative latitudes.

Generalizing the one-dimensional EBM in [equation \(4\)](#) to vary over longitude yields

$$X_\tau(\tau_{\theta\varphi} - \bar{\tau}, \theta, \varphi) = h + QS(\theta) - A - \lambda \tau_{\theta\varphi}, \quad (22)$$

where the subscript τ denotes that the function X is defined with respect to the temperature rather than its anomaly. S reflects mean annual radiation ([North et al., 1981](#)), so it varies over latitude but not over longitude.

However, global observational data sets report temperature anomalies relative to a base period in order to ameliorate known measurement errors and biases in the data. Although the temperature τ is effectively observable at a single location (θ, φ) where a temperature reading is taken, the base temperature relies on aggregation of many observations and thus is not generally observable at a single location, especially over oceans where observations rely on ships. Instead, the set of all locations on the surface of the globe is divided into n disjoint subsets (“boxes”) such that the union of all n of these subsets covers the surface of the globe. In particular, these sets are created by a grid with intervals of equal latitude and equal longitude. For mathematical convenience, no boxes overlap the Equator and locations on the borderline between multiple boxes are assigned systematically to only one box.

Let $L(\theta, \varphi)$, or simply box L , denote the subset that contains (θ, φ) . For a generic function $G(\tau_{\theta\varphi}, \theta, \varphi)$ define

$$G(\tau_L, L) = \frac{1}{2\pi} \iint_L w_L(\theta) G(\tau_{\theta\varphi}, \theta, \varphi) d\theta d\varphi,$$

where the double integral aggregates the function over the box with weights

$$w_L(\theta) = \cos\theta \left(\frac{1}{2\pi} \iint_L \cos\theta d\theta d\varphi \right)^{-1}$$

given by the cosine of latitude divided by the total surface area of the box.

In particular, the expression

$$\tau_L = \frac{1}{2\pi} \iint_L w_L(\theta) \tau_{\theta\varphi} d\theta d\varphi,$$

gives the mean temperature for box L . Moreover, the expression

$$X_\tau(\tau_L, L) = \frac{1}{2\pi} \iint_L w_L(\theta) X_\tau(\tau_{\theta\varphi} - \bar{\tau}, \theta, \varphi) d\theta d\varphi$$

gives the NHT function for box L . And, finally, the expression

$$S(L) = \frac{1}{2\pi} \iint_L w_L(\theta) S(\theta) d\theta d\varphi$$

gives the mean annual radiation for box L .

The base temperature is defined by averaging τ_L over a base period. We denote the base temperature by τ_L^B and treat it as fixed over time. The temperature anomaly for box L is thus defined by $r_L = \tau_L - \tau_L^B$ using the observed values of τ_L and τ_L^B for each box. Finally, defining a function X such that

$$X(r_L, L) = X_\tau((r_L - \bar{r}) - (\tau_L^B - \bar{\tau}^B), L) = X_\tau(\tau_L, L),$$

where $\bar{r} = \bar{\tau} - \bar{\tau}^B$ is the GMTA and $\bar{\tau}^B$ is the mean of temperatures across the globe during the base period, and integrating [equation \(22\)](#) over the box yields the two-dimensional EBM given by [equation \(5\)](#) in Section 2.

A.2. APPROXIMATIONS TO THE EBM

Suppose that the range over which observed surface anomalies vary is given by $[r^-, r^+]$, and define disjoint subsets $L(r) = \{L : r_L = r\}$ of the set of all locations such that the union of all of these locations over $[r^-, r^+]$ covers the surface of the globe. In particular, $L(r)$ is the set of all boxes in which the temperature anomaly takes a value of r . For a generic function $G(\tau_L, L)$ define

$$G(r) = \int_{L(r)} w_{L(r)}(\theta) G(\tau_L, L) dL,$$

where

$$w_{L(r)}(\theta) = \cos\theta \left(\frac{1}{2\pi} \iint_{L(r)} \cos\theta d\theta d\varphi \right)^{-1}$$

is defined similarly to w_L above, but weights for the surface area of all boxes for which an anomaly r is observed rather than that of a single box.

An approximation used by Chang et al. (2020) and Miller and Nam (2020) and employed in the present analysis is to treat each of the boxes as being equally sized – i.e., $G(\tau_L, L) \approx G(\tau_L)$ is assumed. Along these lines, define

$$\bar{G}(r) = \int_{L(r)} \bar{w}_{L(r)} G(\tau_L) dL$$

with

$$\bar{w}_{L(r)} = \left(\frac{1}{2\pi} \iint_{L(r)} d\theta d\varphi \right)^{-1}$$

or, equivalently, define

$$\bar{G}(r) = n_r^{-1} \sum_{L(r)} G(r_L)$$

with n_r defined to be the number of anomalies in the set $L(r)$.

We may define $X(r)$ from $X(r_L, L)$ using either $G(r)$ or $\bar{G}(r)$ and $S(r)$ analogously from $S(\theta)$ to obtain equation (6) from equation (5). In practice, we estimate equation (6) using the equal-sized box approximation to facilitate the interpretation of the distribution function $f(r)$ as a spatial distribution.

A.3. ESTIMATORS OF GMTA

There are multiple ways to approximate GMTA. The simplest,

$$\bar{r} \approx n^{-1} \sum_L r_L,$$

simply weights all anomalies equally as $\bar{G}(r)$ does. A more sophisticated method weights the anomalies by the surface area of the boxes as $G(r)$ does. Yet another method estimates GMTA by

$$\int_{r^-}^{r^+} r \hat{f}(r) dr \approx \int_{r^-}^{r^+} r \left(\frac{1}{2\pi} \iint_{L(r)} \bar{w}_{L(r)} f(r) d\theta d\varphi \right) dr$$

by defining $\hat{f}(r)$ to be an estimate of the density $f(r)$ of anomalies at each box.

The first and third methods both assign equal weights to each box, but they differ in that the first estimates the population mean with the sample mean, while the third estimate the population mean using an estimate of the population density. The second uses a sample mean like the first but weights boxes by their surface area. Fig. 1 shows that the latter two – the most dissimilar of the three – are in fact quite similar, suggesting the adequacy of approximation by equal weighting.

PART IV

MICROECONOMETRICS AND PANEL DATA

This page intentionally left blank

CHAPTER 13

MAXIMUM LIKELIHOOD ESTIMATION OF DYNAMIC PANEL DATA MODELS WITH INTERACTIVE EFFECTS: QUASI-DIFFERENCING OVER TIME OR ACROSS INDIVIDUALS?

Cheng Hsiao^{a,b} and Qiankun Zhou^c

^a*Department of Economics, University of Southern California, University Park, Los Angeles, California, United States*

^b*WISE, Xiamen University, Xiamen, China*

^c*Department of Economics, Louisiana State University, Baton Rouge, Louisiana, United States*

ABSTRACT

The authors consider the quasi maximum likelihood (MLE) estimation of dynamic panel models with interactive effects based on the Ahn et al. (2001, 2013) quasi-differencing methods to remove the interactive effects. The authors show that the quasi-difference MLE (QDMLE) over time is inconsistent when $N \rightarrow \infty$ whether T is fixed or goes to infinity. On the other hand, the QDMLE is consistent and asymptotically unbiased if the difference is taken over individuals when T is large whether N is fixed or large. Monte Carlo studies are conducted to compare the performance of the QDMLE using different quasi-difference methods.

Keywords: Dynamic panel models; interactive effects; maximum likelihood estimation; quasi-difference; high dimensional data; unbiasedness

JEL Classification: C01; C13; C23

1. INTRODUCTION

Economic models typically focus on modeling causal relationships of a few variables an investigator considers important. Panel data provide information on individual outcomes across individuals and over time. Factors affecting individual outcomes are numerous. To control the impacts of unobserved individual and time heterogeneity and to obtain valid statistical inference, various approaches have been suggested such as error component model (e.g., Balestra & Nerlov, 1966; Hoch, 1962; Kuh, 1963), random coefficient model (e.g., Hsiao, 1974, 1975; Swamy, 1970), mixed fixed and random coefficients model (e.g., Hsiao et al., 1993), functional coefficients model (e.g., Chang et al., 2016, 2021), interactive effects model (e.g., Bai, 2009; Pesaran, 2009), etc. In this chapter, we consider the estimation of panel linear dynamic interactive effects models. Contrary to the approach of modeling unobserved individual- and time- specific effects in additive form (e.g., Hsiao, 2014, Ch.3 and 4), there is no simple transformation to get rid of the unobserved individual- and time-specific effects when they are in multiplicative form. Ahn et al. (2001, 2013) (ALS) have suggested a quasi-difference approach to remove the interactive effects in the model. They show that when the conditional covariates are strictly exogenous, applying GMM to the quasi-differenced equations can yield consistent and asymptotically normally distributed estimator if T is fixed and $N \rightarrow \infty$. However, the moments of quasi-differenced models are nonlinear functions of the original model parameters. There could be multiple solutions satisfying the moment conditions. To rule out the inconsistent solutions, separate objective functions need to be introduced (e.g., Honore, 1993; Powell, 1986 for the Tobit-type models). On the other hand, the likelihood approach provides a natural objective of maximizing the quasi-likelihood function. In this chapter, we consider the likelihood approach of estimating linear dynamic models with interactive effects following the ALS quasi-differencing approach. Since quasi-differencing approach can be approached from either the time-differencing or the pairwise differencing perspective, we focus on comparing the asymptotic properties of quasi-maximum likelihood estimates (QMLE). We show that if lagged dependent variables appear as conditional covariates in a model, the quasi-differencing over time is not consistent if T is fixed and $N \rightarrow \infty$ or both N and T go to infinity, $(N, T) \rightarrow \infty$. On the other hand, if one takes quasi-difference across individuals, then it is possible to get consistent and asymptotically normally distributed estimator whether N is fixed or large when $T \rightarrow \infty$.

We set up the model in Section 2. Section 3 introduces quasi-difference over time or across individuals. Section 4 derives the asymptotic properties of QDMLE when differencing over time period. Section 5 discusses the asymptotic properties of QDMLE when differencing across individuals. Results of Monte Carlo simulation studies are provided in Section 6. Concluding remarks are in Section 7.

Proofs of asymptotic results and extension to model with multiple factors are provided in the Appendix. The extension of model with exogenous regressors is provided in the online supplement.

Throughout this chapter, we assume the letter C stands for a generic finite positive constant. The notation \rightarrow_p denotes convergence in probability, \rightarrow_d denotes convergence in distribution, and $(N, T) \rightarrow \infty$ denotes both N and T go to infinity at the same time.

2. MODEL AND ASSUMPTIONS

We consider a simple dynamic panel model

$$y_{it} = \gamma y_{it-1} + \mathbf{x}_{it}' \boldsymbol{\beta} + v_{it}, i = 1, \dots, N; t = 1, \dots, T, \quad (2.1)$$

$$v_{it} = \boldsymbol{\lambda}_t' \mathbf{f}_t + u_{it}, \quad (2.2)$$

where $|\gamma| < 1$, \mathbf{x}_{it} are strictly exogenous variables with regard to u_{it} such that $E(\mathbf{x}_{it} u_{js}) = 0$, $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{ir})'$ denotes the factor loadings while $\mathbf{f}_t = (f_{t1}, \dots, f_{tr})'$ denotes unobservable common factors, r is the number of common factors. The initial observation y_{i0} is observable with $E(y_{i0}^2) < \infty$.

We assume

Assumption A1. u_{it} is independent of $\boldsymbol{\lambda}_i$ and \mathbf{f}_t and is independently and identically distributed over i and t , with mean zero, variance σ_u^2 and finite fourth moments.

Assumption A2. $\boldsymbol{\lambda}_i$ is a vector of constants over i with $N^{-1} \sum_{i=1}^N \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i'$ converges to a positive definite matrix.

Assumption A3. \mathbf{f}_t is a vector of constants over t with $T^{-1} \sum_{t=1}^T \mathbf{f}_t \mathbf{f}_t'$ converges to a positive definite matrix.

The i.i.d assumption on u_{it} (Assumption A1) is made to simplify algebraic derivation. In principle, we can generalize it to allow weak time and cross-sectional dependence (e.g., Bai, 2009; Jiang et al., 2021). However, it does not change the asymptotic properties of QMLE, but will substantially complicate the algebra. Assumptions A2 and A3 are standard assumptions for factor models (e.g., Anderson & Rubin, 1956; Bai, 2003, 2009).

3. QUASI-DIFFERENCING OF PANEL INTERACTIVE MODEL

To illustrate the basic idea of the ALS quasi-differencing approach, we assume $r = 1$, then

$$v_{it} = \lambda_i' f_t + u_{it}, \quad (3.1)$$

For the single common factor structure (3.1), it is impossible to identify f_t and λ_i separately without the normalization. To avoid this problem, we follow ALS (2013) to assume λ_1 or f_1 are nonzero constants¹ and let

$$\theta_t = \frac{f_t}{f_1}, t = 2, \dots, T, \quad (3.2)$$

then

$$y_{it} - \theta_t y_{i1} = \gamma(y_{i,t-1} - \theta_t y_{i0}) + (\mathbf{x}'_{it} - \theta_t \mathbf{x}'_{i1})\beta + (u_{it} - \theta_t u_{i1}), \quad (3.3)$$

for $t = 2, \dots, T$.

The ALS idea of quasi-differencing can also be applied cross-sectionally to get rid of $\lambda'_i f_i$. Suppose $\lambda_1 \neq 0$, let

$$\phi_i = \frac{\lambda_i}{\lambda_1}, i = 2, \dots, N, \quad (3.4)$$

then taking pairwise difference yields

$$y_{it} - \phi_i y_{1t} = \gamma(y_{i,t-1} - \phi_i y_{1,t-1}) + (\mathbf{x}'_{it} - \phi_i \mathbf{x}'_{1t})\beta + (u_{it} - \phi_i u_{1t}), \quad (3.5)$$

for $i = 2, \dots, N, t = 1, \dots, T$,

Remark 3.1. Although the aim of both the linear difference for the one-way fixed effects model and quasi-difference for the interactive effects model is to remove the incidental parameters, there is a fundamental difference between the two. In the one-way fixed effects model, say individual-specific effects, α_i , only appears in across-sectional dimension. Taking linear linear difference over time removes the incidental parameters, α_i , from the transformed model. If the conditional covariates involve lag dependent variables, the covariance transformation (e.g., Hsiao, 2014) or the first difference (Anderson & Hsiao, 1981, 1982) or the forward demean (Arellano & Bover, 1995) or the forward and backward transformation (Han et al., 2014) etc., to remove the incidental parameters in cross-sectional dimension also creates different forms of correlations between the transformed regressors and errors of the equation. Hence, the asymptotic bias depends on the way N or T goes to infinity. On the other hand, in the case of interactive effects, the incidental parameters appear in both cross-sectional and time series dimensions. One may be able to remove the interactive effects through quasi-differencing. However, the incidental parameters issues remain. For instance, as can be seen in (3.3) and (3.5), either θ_t or ϕ_i increases with T or N .

We note that from (3.3), the following moment conditions holds,

$$E(\mathbf{x}_i(u_{it} - \theta_t u_{i1})) = 0, t = 2, \dots, T, \quad (3.6)$$

where $\mathbf{x}_i = (\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT})'$. Ahn et al. (2013) show that applying the GMM to (3.6) yields asymptotically normally distributed estimates of γ and β when T is fixed and $N \rightarrow \infty$. However, the moment conditions (3.6) is nonlinear in $(\gamma, \beta, \theta_t)$. There could be multiple solutions satisfying (3.6). We need to introduce additional conditions to

rule out the irrelevant solutions. On the other hand, the QMLE of (3.3) gives the natural conditions for choosing $(\gamma, \beta, \theta_i)$ to maximize the likelihood function.

Remark 3.2. The asymptotic properties of the quasi-difference estimators are not affected by the presence of strictly exogenous variables \mathbf{x}_{it} . Therefore, for notational ease, we assume $\beta = 0$. Moreover, to have a clear idea of the source of difference between quasi-differencing over time or across individuals, we shall consider the case $r = 1$. The generalization to $r > 1$ is relegated to the Appendix.

4. QUASI-DIFFERENCING OVER TIME

For model (2.1) with a single common factor structure (3.1), conditional on y_{i0} being fixed constants, we have the following system of equations over time periods for (3.3) for each i ,

$$\mathbf{y}_i^*(\boldsymbol{\theta}) = \mathbf{y}_{i,-1}^*(\boldsymbol{\theta})\gamma + \mathbf{u}_i^*(\boldsymbol{\theta}), \quad (4.1)$$

or

$$\dot{\mathbf{y}}_i - \boldsymbol{\theta} y_{i1} = \gamma(\dot{\mathbf{y}}_{i,-1} - \boldsymbol{\theta} y_{i0}) + \dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1}, \quad (4.2)$$

where

$$\begin{aligned} \mathbf{y}_i^*(\boldsymbol{\theta}) &= \begin{pmatrix} y_{i2} - \theta_2 y_{i1} \\ \vdots \\ y_{iT} - \theta_T y_{i1} \end{pmatrix} = \dot{\mathbf{y}}_i - \boldsymbol{\theta} y_{i1}, \quad \dot{\mathbf{y}}_i = \begin{pmatrix} y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \\ \mathbf{y}_{i,-1}^*(\boldsymbol{\theta}) &= \begin{pmatrix} y_{i1} - \theta_2 y_{i0} \\ \vdots \\ y_{iT-1} - \theta_T y_{i0} \end{pmatrix} = \dot{\mathbf{y}}_{i,-1} - \boldsymbol{\theta} y_{i0}, \quad \dot{\mathbf{y}}_{i,-1} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT-1} \end{pmatrix}, \\ \mathbf{u}_i^*(\boldsymbol{\theta}) &= \begin{pmatrix} u_{i2} - \theta_2 u_{i1} \\ \vdots \\ u_{iT} - \theta_T u_{i1} \end{pmatrix} = \dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1}, \quad \dot{\mathbf{u}}_i = \begin{pmatrix} u_{i2} \\ \vdots \\ u_{iT} \end{pmatrix}, \end{aligned}$$

and $\boldsymbol{\theta} = (\theta_2, \dots, \theta_T)'$ is a $(T-1) \times 1$ vector with unrestricted parameters.

Let

$$\begin{aligned} \Omega_{(T-1) \times (T-1)} &= E(\mathbf{u}_i^* \mathbf{u}_i^{*'}) = E[(\dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1})(\dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1})'] \\ &= \sigma_u^2(\mathbf{I}_{T-1} + \boldsymbol{\theta} \boldsymbol{\theta}'), \end{aligned} \quad (4.3)$$

with

$$\Omega^{-1} = \frac{1}{\sigma_u^2} \left(\mathbf{I}_{T-1} - \frac{\boldsymbol{\theta} \boldsymbol{\theta}'}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \right). \quad (4.4)$$

The quasi-log-likelihood function of (4.1) is given by

$$\log L(\gamma, \theta, \sigma_u^2) = -\frac{N}{2} \log |\Omega| - \frac{1}{2} \sum_{i=1}^N (\mathbf{y}_i^*(\theta) - \mathbf{y}_{i,-1}^*(\theta)\gamma)' \Omega^{-1} (\mathbf{y}_i^*(\theta) - \mathbf{y}_{i,-1}^*(\theta)\gamma). \quad (4.5)$$

The first-order conditions for maximizing (4.5) with respect to $(\gamma, \theta, \sigma_u^2)$ are

$$\frac{\partial \log L}{\partial \gamma} = \sum_{i=1}^N \mathbf{y}_{i,-1}^*(\hat{\theta})' \hat{\Omega}^{-1} (\mathbf{y}_i^*(\hat{\theta}) - \mathbf{y}_{i,-1}^*(\hat{\theta})\hat{\gamma}) = 0, \quad (4.6)$$

and

$$\begin{aligned} \frac{\partial \log L}{\partial \theta} &= -N\sigma_u^2 \hat{\Omega}^{-1} \hat{\theta} + N\sigma_u^2 \hat{\Omega}^{-1} \mathbf{S}(\hat{\theta}) \hat{\Omega}^{-1} \hat{\theta} \\ &\quad + \hat{\Omega}^{-1} \sum_{i=1}^N [(y_{i1} - y_{i0}\hat{\gamma})(\dot{\mathbf{y}}_i - \dot{\mathbf{y}}_{i,-1}\hat{\gamma}) - \hat{\theta}(y_{i1} - y_{i0}\hat{\gamma})^2] \\ &= 0, \end{aligned} \quad (4.7)$$

where $\hat{\Omega} = \frac{1}{\hat{\sigma}_u^2} \mathbf{S}(\hat{\theta})$ with

$$\mathbf{S}(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N (\mathbf{y}_i^*(\hat{\theta}) - \mathbf{y}_{i,-1}^*(\hat{\theta})\hat{\gamma}) (\mathbf{y}_i^*(\hat{\theta}) - \mathbf{y}_{i,-1}^*(\hat{\theta})\hat{\gamma})', \quad (4.8)$$

and

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)} \sum_{i=1}^N (\mathbf{y}_i^* - \mathbf{y}_{i,-1}^*\hat{\gamma})' (\mathbf{I}_{T-1} + \hat{\theta}\hat{\theta}')^{-1} (\mathbf{y}_i^* - \mathbf{y}_{i,-1}^*\hat{\gamma}). \quad (4.9)$$

Conditional on θ , from (4.6), we obtain

$$\hat{\gamma}_{MLE}^{TS} = \left(\sum_{i=1}^N \mathbf{y}_{i,-1}^*(\theta)' \Omega^{-1} \mathbf{y}_{i,-1}^*(\theta) \right)^{-1} \sum_{i=1}^N (\mathbf{y}_{i,-1}^*(\theta)' \Omega^{-1} \mathbf{y}_i^*(\theta)), \quad (4.10)$$

where TS refers to using normalization (3.2) to apply the quasi-difference over time period.

We note that conditional on θ ,

$$\hat{\gamma}_{MLE}^{TS} - \gamma = \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}^*(\theta)' \Omega^{-1} \mathbf{y}_{i,-1}^*(\theta) \right)^{-1} \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}^*(\theta)' \Omega^{-1} \mathbf{u}_i^*(\theta), \quad (4.11)$$

and the denominator converges to

$$\frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}^*(\theta)' \Omega^{-1} \mathbf{y}_{i,-1}^*(\theta) \rightarrow_p \frac{\bar{\sigma}_\lambda^2 \bar{\sigma}_{\gamma,f}^2}{\sigma_u^2} + \frac{1}{1-\gamma^2}, \quad (4.12)$$

where $\bar{\sigma}_\lambda^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \lambda_i^2$ and $\bar{\sigma}_{\gamma,f}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \omega' (\mathbf{I}_{T-1} + \theta\theta')^{-1} \omega$ with

$$\omega = (\omega_1, \dots, \omega_{T-1})' \text{ and } \omega_t = \sum_{s=1}^t \gamma^{t-s} f_s.$$

The numerator of (4.11) converges to

$$\begin{aligned}
tr[\Omega(\boldsymbol{\theta})^{-1} \sigma_u^2 \boldsymbol{\theta}(1, \gamma, \dots, \gamma^{T-2})] &= tr\left[\left(\mathbf{I}_{T-1} - \frac{\boldsymbol{\theta}\boldsymbol{\theta}'}{1+\boldsymbol{\theta}'\boldsymbol{\theta}}\right)\boldsymbol{\theta}(1, \gamma, \dots, \gamma^{T-2})\right] \\
&= \frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} tr[\boldsymbol{\theta}(1, \gamma, \dots, \gamma^{T-2})] \\
&= \frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} tr(1, \gamma, \dots, \gamma^{T-2})\boldsymbol{\theta}' \\
&= \frac{1}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} \sum_{t=2}^T \theta_t \gamma^{t-2} = O\left(\frac{1}{T}\right).
\end{aligned} \tag{4.13}$$

Equations (4.12) and (4.13) imply that the numerator of (4.11) is $O_p\left(\frac{1}{T}\right)$, it does not go to zero when T is fixed no matter how large N is. In other words, $\hat{\gamma}_{MLE}^{TS}$ is inconsistent if T is fixed and $N \rightarrow \infty$.

Although conditional on $\boldsymbol{\theta}$, the QDMLE of γ is consistent when $T \rightarrow \infty$, if $\frac{N}{T} \rightarrow a \neq 0$ as $\frac{N}{T} \rightarrow a \neq 0$

$$E\left(\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}^*(\boldsymbol{\theta})' \Omega^{-1} \mathbf{u}_i^*(\boldsymbol{\theta})\right) = -\frac{\boldsymbol{\theta}' \Psi \boldsymbol{\theta}}{1+\boldsymbol{\theta}'\boldsymbol{\theta}} \sqrt{a} + o(1), \tag{4.14}$$

where Ψ is defined in (A.8).

Unfortunately, $\hat{\boldsymbol{\theta}}$ is inconsistent even $(N, T) \rightarrow \infty$. We note that conditional on γ , (4.7) leads to

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= \left[\left(\frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)^2 + \sigma_u^2 \right) \mathbf{I}_{T-1} - \frac{1}{1+\hat{\boldsymbol{\theta}}'\hat{\boldsymbol{\theta}}} \mathbf{S} \right]^{-1} \times \frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)(\dot{\mathbf{y}}_i - \dot{\mathbf{y}}_{i,-1}\gamma) \\
&= \left(\frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)^2 + \sigma_u^2 \right)^{-1} \frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)[\boldsymbol{\theta}(y_{i1} - \gamma y_{i0}) + (\dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1})] + o_p(1) \\
&= \boldsymbol{\theta} + \left(\frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)^2 + \sigma_u^2 \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)(\dot{\mathbf{u}}_i - \boldsymbol{\theta} u_{i1}) - \boldsymbol{\theta} \sigma_u^2 \right) + o_p(1) \\
&= \boldsymbol{\theta} - \frac{2\sigma_u^2}{\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (y_{i1} - y_{i0}\gamma)^2 + \sigma_u^2} \boldsymbol{\theta} + o_p(1) \\
&\not\asymp_p \boldsymbol{\theta}.
\end{aligned} \tag{4.15}$$

Since the score function evaluated at the true $(\gamma, \boldsymbol{\theta})$,

$$\frac{1}{NT} \begin{pmatrix} \frac{\partial \log L}{\partial \gamma} \\ \frac{\partial \log L}{\partial \boldsymbol{\theta}} \end{pmatrix} \not\asymp 0 \quad \text{as } (N, T) \rightarrow \infty, \tag{4.16}$$

the QDMLE (4.10) is inconsistent whether T is fixed or $T \rightarrow \infty$.

It then follows that

Theorem 4.1. *Under Assumptions A1–A3, conditional on y_{i0} being fixed constants, the QDMLE (4.10) over time for model (2.1) with single common factor structure (3.1) is inconsistent when $N \rightarrow \infty$ whether T is fixed or $T \rightarrow \infty$. If consistent estimator for θ can be found, the QDMLE (4.10) is consistent if $T \rightarrow \infty$. However, if $\frac{N}{T} \rightarrow a \neq 0 < \infty$, then*

$$\sqrt{NT} \left(\hat{\gamma}_{MLE}^{TS} - \left(\gamma - \frac{1}{T} \frac{b}{k_1} \right) \right) \xrightarrow{d} N\left(0, \frac{k_2}{k_1^2}\right), \quad (4.17)$$

where b denotes the bias term defined in (A.10), k_1 and k_2 are given by

$$k_1 = \frac{\bar{\sigma}_\lambda^2 \bar{\sigma}_{\gamma,f}^2}{\sigma_u^2} + \frac{1}{1-\gamma^2}, \quad k_2 = \frac{\bar{\sigma}_\lambda^2 \tilde{\sigma}_{\gamma,f}^2}{\sigma_u^2} + \frac{1}{1-\gamma^2}, \quad (4.18)$$

where $\bar{\sigma}_\lambda^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \lambda_i^2$, $\bar{\sigma}_{\gamma,f}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \omega' (\mathbf{I}_{T-1} + \theta\theta')^{-1} \omega$ and $\tilde{\sigma}_{\gamma,f}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \omega' \omega$ with $\omega = (\omega_1, \dots, \omega_{T-1})'$ and $\omega_t = \sum_{s=1}^t \gamma^{t-s} f_s$.

Remark 4.2. *There are many different ways to implement the ALS quasi-difference approach. For instance, we can use the long difference as (3.3). We can also use the first difference by letting $\theta_t^* = \frac{f_t}{f_{t-1}}$, then*

$$y_{it} - \theta_t^* y_{i,t-1} = \gamma(y_{i,t-1} - \theta_{t-1}^* y_{i,t-2}) + (\mathbf{x}'_{it} - \theta_t^* \mathbf{x}'_{i,t-1})\beta + (u_{it} - \theta_t^* u_{i,t-1}), \quad t=2, \dots, T. \quad (4.19)$$

Or the backward and forward differencing proposed by [Han et al. \(2014\)](#). In principle, one can derive the QDMLE (4.10) with corresponding transformed likelihood function. However, there is an advantage of using the long difference, the variance–covariance matrix of the transformed error terms has the form of standard random effects one-way error component model. The variance-covariance matrix of the errors has the form (4.3) and known pattern of its inverse (4.4). On the other hand, if we use (4.19), the variance-covariance matrix of the system $\tilde{\mathbf{u}}_i^* = (u_{i2} - \theta_2^* u_{i1}, \dots, u_{iT} - \theta_T^* u_{i,T-1})'$ takes the form

$$\sigma_u^2 \begin{pmatrix} (1+\theta_2^{*2}) & -\theta_2^* & 0 & 0 \\ -\theta_2^* & (1+\theta_3^{*2}) & \ddots & 0 \\ 0 & \ddots & \ddots & -\theta_T^* \\ 0 & 0 & -\theta_T^* & (1+\theta_T^{*2}) \end{pmatrix}. \quad (4.20)$$

The element of its inverse matrix consists of all $(\theta_2^*, \dots, \theta_T^*)$, and its computation is much more complicated. Similarly for the backward and forward transformation proposed by [Han et al. \(2014\)](#).

However, we expect the asymptotic properties of the different forms of transformations remain the same. The QDMLE (4.10) remains inconsistent whether T is fixed or goes to infinity as $N \rightarrow \infty$. For instance, in the case of (4.19), $\theta_t^* = \frac{f_t}{f_{t-1}}$. If θ_t cannot be consistently estimated either T is fixed or large no matter how large N is, then there is no reason to expect θ_t^* will be consistently estimated.

5. QUASI-DIFFERENCING ACROSS INDIVIDUALS

Rewrite (3.5) in vector form as

$$\mathbf{y}_t^*(\phi) = \mathbf{y}_{t-1}^*(\phi)\gamma + \mathbf{u}_t^*(\phi), t=1,\dots,T, \quad (5.1)$$

or

$$\dot{\mathbf{y}}_t - \phi y_{1t} = \gamma(\dot{\mathbf{y}}_{t-1} - \phi y_{1,t-1}) + \dot{\mathbf{u}}_t - \phi u_{1t}, \quad (5.2)$$

where

$$\begin{aligned} \mathbf{y}_{t-1}^*(\phi) &= \begin{pmatrix} y_{2t} - \phi_2 y_{1t} \\ \vdots \\ y_{Nt} - \phi_N y_{1t} \end{pmatrix} = \dot{\mathbf{y}}_t - \phi y_{1t}, \quad \dot{\mathbf{y}}_t = \begin{pmatrix} y_{2t} \\ \vdots \\ y_{Nt} \end{pmatrix}, \\ \mathbf{y}_{t-1}^*(\phi) &= \begin{pmatrix} y_{2,t-1} - \phi_2 y_{1,t-1} \\ \vdots \\ y_{N,t-1} - \phi_N y_{1,t-1} \end{pmatrix} = \dot{\mathbf{y}}_{t-1} - \phi y_{1,t-1}, \quad \mathbf{y}_{t-1} = \begin{pmatrix} y_{2,t-1} \\ \vdots \\ y_{N,t-1} \end{pmatrix}, \\ \mathbf{u}_t^*(\phi) &= \begin{pmatrix} u_{2t} - \phi_2 u_{1t} \\ \vdots \\ u_{Nt} - \phi_N u_{1t} \end{pmatrix} = \dot{\mathbf{u}}_t - \phi u_{1t}, \quad \dot{\mathbf{u}}_t = \begin{pmatrix} u_{2t} \\ \vdots \\ u_{Nt} \end{pmatrix}, \end{aligned}$$

and $\phi = (\phi_2, \dots, \phi_N)'$ is a $(N-1) \times 1$ vector with unrestricted parameters.

We note that for model (5.1), for any $s \neq t$,

$$\begin{aligned} E(\mathbf{u}_t^*(\phi)\mathbf{u}_s^*(\phi)') &= E[(\dot{\mathbf{u}}_t - \phi u_{1t})(\dot{\mathbf{u}}_s' - \phi' u_{1s})] \\ &= E(\dot{\mathbf{u}}_t \dot{\mathbf{u}}_s') - E(\dot{\mathbf{u}}_t u_{1s})\phi' - \phi E(u_{1t} \dot{\mathbf{u}}_s') + \phi \phi' E(u_{1t} u_{1s}) \\ &= 0, \end{aligned} \quad (5.3)$$

under serially uncorrelated assumption (A1). The variance-covariance matrix of (5.1) takes the form,

$$\begin{aligned} \tilde{\Omega}_{(N-1) \times (N-1)} &= E(\mathbf{u}_t^*(\phi)\mathbf{u}_t^*(\phi)') = E[(\dot{\mathbf{u}}_t - \phi u_{1t})(\dot{\mathbf{u}}_t' - \phi' u_{1t}')] \\ &= \sigma_u^2 (\mathbf{I}_{N-1} + \phi \phi'), \end{aligned} \quad (5.4)$$

with

$$\tilde{\Omega}^{-1} = \frac{1}{\sigma_u^2} \left(\mathbf{I}_{N-1} - \frac{\phi\phi'}{1+\phi'\phi} \right). \quad (5.5)$$

The quasi-log-likelihood function of (5.1) is given by

$$\log L = -\frac{T}{2} \log |\tilde{\Omega}| - \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t^*(\phi) - \mathbf{y}_{t-1}^*(\phi)\gamma)' \tilde{\Omega}^{-1} (\mathbf{y}_t^*(\phi) - \mathbf{y}_{t-1}^*(\phi)\gamma). \quad (5.6)$$

Conditional on ϕ ,

$$\hat{\gamma}_{MLE}^{CS} = \left(\sum_{t=1}^T \mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_{t-1}^*(\phi) \right)^{-1} \left(\sum_{t=1}^T \mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_t^*(\phi) \right), \quad (5.7)$$

where CS refers to using normalization (3.4) to apply the quasi-difference over cross-sectional dimension.

To see that the above QDMLE (5.7) across individuals is asymptotically unbiased, we note that

$$\sqrt{NT}(\hat{\gamma}_{MLE}^{CS} - \gamma) = \left[\frac{1}{NT} \sum_{t=1}^T \mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_{t-1}^*(\phi) \right]^{-1} \frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{u}_t^*(\phi), \quad (5.8)$$

where $\mathbf{y}_t^* = \mathbf{y}_{t-1}^* \gamma + \mathbf{u}_t^* = \sum_{j=0}^{\infty} \gamma^j \mathbf{u}_{t-j}^*$ since $|\gamma| < 1$. Thus, as $(N, T) \rightarrow \infty$, the denominator of (5.8) converges

$$\frac{1}{NT} \sum_{t=1}^T E[\mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_{t-1}^*(\phi)] = \frac{1}{NT} \sum_{t=1}^T \text{tr}\{\tilde{\Omega}^{-1} E[\mathbf{y}_{t-1}^*(\phi) \mathbf{y}_{t-1}^*(\phi)']\} \rightarrow \frac{1}{1-\gamma^2}, \quad (5.9)$$

because

$$E[\mathbf{y}_t^*(\phi) \mathbf{y}_t^*(\phi)'] = \frac{1}{1-\gamma^2} E[\mathbf{u}_t^*(\phi) \mathbf{u}_t^*(\phi)'] = \frac{1}{1-\gamma^2} \tilde{\Omega}. \quad (5.10)$$

For the numerator of (5.8), we notice that

$$\begin{aligned} E\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{u}_t^*(\phi)\right) &= \frac{1}{\sqrt{NT}} \sum_{t=1}^T E[\mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{u}_t^*(\phi)] \\ &= \frac{1}{\sqrt{NT}} \sum_{t=1}^T \text{tr}\{\tilde{\Omega}^{-1} E[\mathbf{u}_t^*(\phi) \mathbf{y}_{t-1}^*(\phi)']\} \\ &= 0, \end{aligned} \quad (5.11)$$

which suggests that the QDMLE (5.7) is asymptotically unbiased conditional on ϕ , and the asymptotical unbiasedness of QDMLE is independent of the way how (N, T) go to infinity.

Moreover, we can show that

$$\begin{aligned}
E\left(\frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{u}_t^*(\phi)\right)^2 &= \frac{1}{NT} \sum_{s,t} E\left[\mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{u}_t^*(\phi) \mathbf{u}_s^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_{s-1}^*(\phi)\right] \\
&= \frac{1}{NT} \sum_{t=1}^T E\left[\mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{u}_t^*(\phi) \mathbf{u}_t^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_{t-1}^*(\phi)\right] \\
&= \frac{1}{NT} \sum_{t=1}^T E\left[\mathbf{y}_{t-1}^*(\phi)' \tilde{\Omega}^{-1} \mathbf{y}_{t-1}^*(\phi)\right] \\
&\rightarrow \frac{1}{1-\gamma^2},
\end{aligned} \tag{5.12}$$

as $(N, T) \rightarrow \infty$. Furthermore, we can show the fourth moments of the numerator of (5.8) goes to zero by following [Bao and Ullah \(2010\)](#).

Combining the results of (5.9)–(5.12), we can show that

Theorem 5.1. *Under assumption A1-A3, the pairwise QDMLE (5.7) for model (2.1) with single common factor structure (3.1) is consistent and asymptotically unbiased as $N \rightarrow \infty$ whether T is fixed or $T \rightarrow \infty$, and*

$$\sqrt{NT} (\hat{\gamma}_{MLE}^{CS} - \gamma) \rightarrow_d N(0, 1 - \gamma^2), \tag{5.13}$$

as $(N, T) \rightarrow \infty$.

The pairwise QDMLE depends on the simultaneous solution of the first-order conditions (5.6) which can be computational cumbersome. One way to obtain a feasible solution is through the following iteration.

Step 1: Conditional on some initial estimates of ϕ , apply the least squares estimation to (3.5) to obtain initial estimates of γ and β .

Step 2: Conditional on γ and β , minimizing $\sum_{t=1}^T (\hat{u}_{it}^* - \phi_i \hat{u}_{it}^*)^2$ to obtain the least squares estimator of ϕ_i for $i = 2, \dots, N$ where \hat{u}_{it}^* denotes the residuals based on the initial estimates of γ and β .

Step 3: Iterate between Step 1 and 2 until the solutions of (γ, β, ϕ) converge.²

Step 4: Substitute the convergent solutions of (γ, β, ϕ) as initial estimates to implement the interactive procedure to obtain the pairwise QDMLE.

6. MONTE CARLO SIMULATION

In this section, we investigate the finite sample properties of the QDMLE for panels with cross-sectional dependence. The data generating process (DGP) is given by

$$y_{it} = \gamma y_{i,t-1} + \lambda_i f_t + u_{it}, \tag{6.1}$$

We generate $f_t \sim IIDN(3,1)$ for $t = 1, \dots, T$ and $\lambda_i \sim IIDN(2,0.3)$ for $t = 1, \dots, N$. We assume the error term $u_{it} \sim IIDN(0,1)$ for $i = 1, 2, \dots, N$, $t = 1, \dots, T$, and they are independent of λ_i and f_t for all i and t .

The true value of γ is set at $\gamma = 0.1, 0.5, 0.9$. We let $N = 50, 100, 500$ and $T = 50, 100, 500, 1,000$. The number of replication is 1,000 times, and the maximum number of iteration of feasible QDMLE is 100. We report mean estimates, bias, RMSE and size comparison for these two estimators using the nominal 5% significance level, QDMLE (4.10) using normalization (3.2) and MLE-CS refers to QDMLE (5.7) using normalization (3.4).

The simulation results are summarized in Tables 1–3. For the infeasible QDMLE estimators, which use the true prespecified factors or factor loadings (or ϕ_i) in the estimation. The pairwise QDMLE has negligible bias and the actual size is close to the nominal size whatever sample configurations of N and T are, even for cases where the lag coefficient is on the boundary (e.g., $\gamma = 0.1$ or $\gamma = 0.9$). However, the feasible pairwise QDMLE that depends on estimated ϕ_i , $\hat{\phi}_i$, although has negligible bias, could have significant size distortion when T is small due to the instability of $\hat{\phi}_i$. Nevertheless, the size distortion declines with the increase of T . On the other hand, the QDMLE over time has significant bias and significant size distortion even in the infeasible case. In all, the simulation results appear to support the findings that QDMLE over time is asymptotically biased while QDMLE pairwise is asymptotically unbiased.

7. CONCLUSION

Panel data blend inter-individual difference and intra-individual dynamics provide means to identify impacts of omitted time-invariant and individual-invariant variables. If the unobserved individual- and time-specific effects are in additive form, they can be easily removed through a covariance transformation (e.g., Hsiao, 2014, Ch3 and Ch4). On the other hand, if they are in multiplicative form, linear transformations cannot remove them. Ahn et al. (2001, 2013) have suggested a quasi-differencing method to remove the interactive effects. They show the resulting GMM method is consistent and asymptotically unbiased if the regressor are strictly exogenous. In this chapter, we show that the quasi-differencing MLE over time period for a dynamic panel model is inconsistent when either T is fixed and N is large or both N and T are large. On the other hand, if we take quasi-differencing across individuals, the resulting quasi-maximum likelihood estimator is consistent and asymptotically unbiased if $T \rightarrow \infty$ whether N is fixed or $N \rightarrow \infty$.

NOTES

1. In the extreme case when $\lambda_i = 0$ or $f_i = 0$, we can choose any other factor loading as long as $\lambda_i \neq 0$ and any other factor as long as $f_i \neq 0$. Then by switching index, the normalizations (3.2) and (3.4) are still valid, and the derivation will follow.
2. Steps 1–3 can be viewed as an iterative procedure to obtain the nonlinear least squares estimation of (3.5).

Table I. Simulation Results for γ of DGP (6.1) When $\gamma = 0.1$

N	T	50						100						500						1000					
		Feasible MLE			Infeasible MLE			Feasible MLE			Infeasible MLE			Feasible MLE			Infeasible MLE			Feasible MLE			Infeasible MLE		
		TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS	TS	CS
50	Mean	0.1529	0.0773	0.5224	0.0986	0.1475	0.0890	0.5224	0.0994	0.1554	0.0974	0.5247	0.0998	0.1544	0.0988	0.5249	0.0999	0.5249	0.0999	0.5249	0.0999	0.5249	0.0999	0.5249	0.0999
	Bias	0.0529	-0.0227	0.4224	-0.0014	0.0475	-0.0110	0.4224	-0.0006	0.0554	-0.0026	0.4237	-0.0002	0.0544	-0.0012	0.4249	-0.0001	0.4249	-0.0012	0.4249	-0.0001	0.4249	-0.0001	0.4249	-0.0001
	RMSE	0.0570	0.0304	0.4224	0.0199	0.0500	0.0185	0.4224	0.0141	0.0558	0.0071	0.4247	0.0064	0.0547	0.0047	0.4249	0.0046	0.4249	0.0046	0.4249	0.0046	0.4249	0.0046	0.4249	0.0046
	size	72%	19%	100%	4.7%	86%	11%	100%	5.5%	100%	5.2%	100%	5%	100%	5.8%	100%	5%	100%	5.8%	100%	5%	100%	5%	100%	5%
100	Mean	0.1237	0.0778	0.5194	0.0994	0.1391	0.0896	0.5263	0.0997	0.1488	0.0978	0.5234	0.0999	0.1512	0.0900	0.5247	0.0999	0.5247	0.0999	0.5247	0.0999	0.5247	0.0999	0.5247	0.0999
	Bias	0.0237	-0.0222	0.4194	-0.0006	0.0391	-0.0104	0.4263	-0.003	0.0488	-0.0022	0.4234	-0.0001	0.0512	-0.0100	0.4247	-0.0001	0.4247	-0.0001	0.4247	-0.0001	0.4247	-0.0001	0.4247	-0.0001
	RMSE	0.0286	0.0268	0.4194	0.0147	0.0404	0.0141	0.4263	0.0100	0.0490	0.0051	0.4234	0.0045	0.0513	0.0035	0.4247	0.0032	0.4247	0.0032	0.4247	0.0032	0.4247	0.0032	0.4247	0.0032
	size	28%	30%	100%	5.5%	96%	19%	100%	3.8%	100%	6%	100%	5.7%	100%	5.2%	100%	5.4%	100%	5.4%	100%	5.4%	100%	5.4%	100%	5.4%
500	Mean	0.1170	0.0795	0.5127	0.1002	0.1244	0.0900	0.5193	0.0999	0.1424	0.0980	0.5232	0.1000	0.1405	0.0900	0.5229	0.0999	0.5229	0.0999	0.5229	0.0999	0.5229	0.0999	0.5229	0.0999
	Bias	0.0170	-0.0205	0.4127	0.0002	0.0244	-0.0100	0.4193	-0.0001	0.0424	-0.0020	0.4232	0.0000	0.0405	-0.0010	0.4229	-0.0001	0.4229	-0.0001	0.4229	-0.0001	0.4229	-0.0001	0.4229	-0.0001
	RMSE	0.0182	0.0213	0.428	0.0065	0.0249	0.0110	0.4193	0.0044	0.0425	0.0028	0.4232	0.0019	0.0405	0.0017	0.4229	0.0014	0.4229	0.0014	0.4229	0.0014	0.4229	0.0014	0.4229	0.0014
	size	74%	91%	100%	5.4%	99%	58%	100%	5.8%	100%	18%	100%	5.2%	100%	11%	100%	5.1%	100%	5.1%	100%	5.1%	100%	5.1%	100%	5.1%

Notes: 1. “feasible MLE” refers to the feasible QDMLE using the iterative procedure, “infeasible MLE” refers to the QDMLE using the true prespecified factors or factor loadings to construct the variance-covariance matrix.

2. TS refers to QDMLE (4.10) using normalization (3.2), CS refers to QDMLE (5.7) using normalization (3.4).

3. The size is calculated at $H_0 : \gamma_0 = 0.1$.

Table 2. Simulation Results for γ of DGP (6.1) When $\gamma = 0.5$

N	T	50				100				500				1000			
		Feasible MLE		Infeasible MLE		Feasible MLE		Infeasible MLE		Feasible MLE		Infeasible MLE		Feasible MLE		Infeasible MLE	
		TS	CS	TS	CS												
50	Mean	0.6962	0.4693	0.4986	0.4988	0.6905	0.4852	0.4990	0.4994	0.6813	0.4970	0.4999	0.5000	0.6799	0.4985	0.4999	0.4999
	Bias	0.1962	-0.0307	-0.0014	-0.0012	0.1905	-0.0148	-0.0010	-0.0006	0.1813	-0.0030	-0.0001	0.0000	0.179	-0.0015	-0.0001	-0.0001
	RMSE	0.1964	0.0356	0.0042	0.0171	0.1096	0.0197	0.0031	0.0123	0.1813	0.0064	0.0013	0.0055	0.179	0.0042	0.0009	0.0040
100	size	100%	39%	6.8%	5.1%	100%	22%	6.9%	4.9%	100%	7.2%	5%	5%	100%	7.4%	5.4%	5.3%
	Mean	0.6875	0.4698	0.4982	0.4993	0.6773	0.4848	0.4992	0.4999	0.6747	0.4973	0.4998	0.4999	0.6807	0.4986	0.4999	0.4999
	Bias	0.1875	-0.0302	-0.0018	-0.0007	0.1773	-0.0142	-0.0008	-0.0001	0.1747	-0.0027	-0.0002	-0.0001	0.1807	-0.0014	-0.0001	-0.0001
500	RMSE	0.1877	0.0333	0.0037	0.0128	0.1774	0.0167	0.0022	0.0087	0.1747	0.0048	0.0010	0.0039	0.1807	0.0032	0.0006	0.0028
	size	100%	56%	7.6%	5.7%	100%	36%	6.9%	4.7%	100%	10%	6.1%	5.4%	100%	6.8%	4.4%	5.2%
	Mean	0.6577	0.4718	0.4983	0.5003	0.6383	0.4862	0.4992	0.5000	0.6634	0.4973	0.4999	0.5000	0.6632	0.4986	0.4999	0.5000
	Bias	0.1577	-0.0282	-0.0017	0.0003	0.1383	-0.0138	-0.0008	0.0000	0.1634	-0.0027	-0.0001	0.0000	0.1632	-0.0014	-0.0001	0.0000
	RMSE	0.1578	0.0287	0.0022	0.0057	0.1383	0.0144	0.0013	0.0038	0.1634	0.0032	0.0004	0.0017	0.1632	0.0018	0.0003	0.0012
	size	100%	100%	23%	5.1%	100%	92%	13%	5.8%	100%	36%	5.7%	4.6%	100%	18%	6.4%	4.6%

Notes: The size is calculated at $H_0 : \gamma_0 = 0.5$. Refer to notes of Table 1.

Table 3. Simulation Results for γ of DGP (6.1) When $\gamma = 0.9$

N	T	50				100				500				1000			
		Feasible MLE	Infeasible MLE														
50	Mean	0.9917	0.8556	0.9909	0.8994	0.9922	0.8796	0.9966	0.8994	0.9877	0.8965	0.9897	0.9001	0.9878	0.8982	0.9918	0.9000
	Bias	0.0917	-0.0444	0.0909	-0.0006	0.0922	-0.0204	0.0966	-0.0006	0.0877	-0.0035	0.0897	0.0001	0.0878	-0.0018	0.0918	0.0000
	RMSE	0.0917	0.0458	0.0909	0.0085	0.0922	0.0218	0.0966	0.0062	0.0877	0.0045	0.0897	0.0027	0.0878	0.0027	0.0918	0.0020
100	size	100%	96%	100%	4.8%	100%	73%	100%	5.2%	100%	22%	100%	5.4%	100%	13%	100%	5%
	Mean	0.9934	0.8563	1.0046	0.8996	0.9889	0.8804	0.9979	0.9001	0.9877	0.8966	0.9913	0.9000	0.9885	0.8983	0.9928	0.9000
	Bias	0.0934	-0.0437	0.1046	-0.0004	0.0889	-0.0196	0.0979	0.0001	0.0877	-0.0034	0.0913	0.0000	0.0885	-0.0017	0.0928	0.0000
500	RMSE	0.0934	0.0446	0.1046	0.0063	0.0889	0.0203	0.0979	0.0044	0.0877	0.0039	0.0913	0.0019	0.0885	0.0022	0.0928	0.0014
	size	100%	99%	100%	5.5%	100%	96%	100%	5%	100%	41%	100%	4.9%	100%	21%	100%	4.7%
	Mean	0.9854	0.8584	0.9792	0.9002	0.9854	0.8810	0.9909	0.9001	0.9871	0.8965	0.9917	0.9000	0.9875	0.8983	0.9928	0.9000
	Bias	0.0854	-0.0416	0.0742	0.0002	0.0854	-0.0190	0.0909	0.0001	0.0871	-0.0035	0.0917	0.0000	0.0875	-0.0017	0.0928	0.0000
	RMSE	0.0854	0.0418	0.0742	0.0028	0.0854	0.0192	0.0909	0.0019	0.0871	0.0036	0.0917	0.0009	0.0875	0.0017	0.0928	0.0006
	size	100%	100%	100%	5.4%	100%	100%	100%	4%	100%	97%	100%	4.6%	100%	76%	100%	5.1%

Notes: The size is calculated at $H_0 : \gamma_0 = 0.9$. Refer to notes of Table 1.

ACKNOWLEDGMENTS

We would like to thank the editor, a referee and Ruiqi Liu and Yonghui Zhang for helpful comments. All remaining errors are solely ours. Partial research support of China NSF grant #771631004 and #72033008 to the first author is gratefully acknowledged.

REFERENCES

- Ahn, S. C., Lee, Y. H., & Schmidt, P. (2001). GMM estimation of linear panel data models with time-varying individual effects. *Journal of Econometrics*, 101, 219–255.
- Ahn, S. C., Lee, Y. H., & Schmidt, P. (2013). Panel data models with multiple time-varying individual effects. *Journal of Econometrics*, 174, 1–14.
- Anderson, T. W. (1971). *The statistical analysis of time series*. John Wiley & Sons, Inc.
- Anderson, T. W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, 76, 598–606.
- Anderson, T. W., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18, 47–82.
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 5, pp. 111–150). Berkeley, CA, USA.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68, 29–51.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71, 135–171.
- Bai, J. (2009). Panel data models with interactive effects. *Econometrica*, 77, 1229–1279.
- Balestra, P., & Nerlove, M. (1966). Pooling cross-section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, 34, 585–612.
- Bao, Y., & Ullah, A. (2010). Expectation of quadratic forms in normal and nonnormal variables with applications. *Journal of Statistical Planning and Inference*, 140, 1193–1205.
- Chang, Y., Choi, Y., Kim, C., Miller, J. I., & Park, J. (2016). Disentangling temporal patterns in elasticities: A functional coefficient panel analysis of electricity demand. *Energy Economics*, 60, 232–243.
- Chang, Y., Choi, Y., Kim, C., Miller, Z., & Park, J. (2021). Forecasting regional long-run energy demand: A functional coefficient panel approach. *Energy Economics*, 96, 105–117.
- Fang, Y., Loparo, K., & Feng, X. (1994). Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control*, 39, 2489–2490.
- Han, C., Phillips, P., & Sul, D. (2014). X-differencing and dynamic panel model estimation. *Econometric Theory*, 30, 201–251.
- Hoch, I. (1962). Estimation of production function parameters combining time-series and cross-section data. *Econometrica*, 30, 34–53.
- Honore, B. E. (1993). Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics*, 59, 35–61.
- Hsiao, C. (1974). Statistical inference for a model with both random cross-sectional and time effects. *International Economic Review*, 15, 12–30.
- Hsiao, C. (1975). Some estimation methods for a random coefficients model. *Econometrica*, 43, 305–325.
- Hsiao, C. (2014). *Analysis of panel data* (3rd ed.). Cambridge University Press.
- Hsiao, C., Appelbe, T. W., & Dineen, C. R. (1993). A general framework for panel data analysis – with an application to Canadian customer dialed long distance service. *Journal of Econometrics*, 59, 63–86.
- Hsiao, C., & Zhou, Q. (2018). Incidental parameters, initial conditions and sample size in statistical inference for dynamic panel data models. *Journal of Econometrics*, 207, 114–128.
- Jiang, B., Yang, Y., Gao, J., & Hsiao, C. (2021). Recursive estimation in large panel data models: Theory and practice. *Journal of Econometrics*, 224, 439–465.

- Kuh, E. (1963). The validity of cross sectionally estimated behavior equations in time series applications. *Econometrica*, 27, 197–214.
- Magnus, J. R., & Neudecker, H. (1999) *Matrix differential calculus with applications in statistics and econometrics* (2nd ed.). Wiley.
- Pesaran, M. H. (2009). Estimation and inference in large heterogeneous panels with cross-section dependence. *Econometrica*, 74, 967–1012.
- Powell, J. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica*, 54, 1435–1460.
- Swamy, P. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38, 311–323.

APPENDIX. MATHEMATICAL DERIVATIONS

A. DERIVATION OF THEOREM 4.1

Conditional on θ , for (4.11), we have

$$\sqrt{NT}(\hat{\gamma}_{MLE}^{TS} - \gamma) = \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \Omega^{-1} \mathbf{y}_{i,-1}^* \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}' \Omega^{-1} \mathbf{u}_i^*. \quad (\text{A.1})$$

where

$$\begin{aligned} & \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \Omega^{-1} \mathbf{y}_{i,-1}^* \\ &= \frac{1}{\sigma_u^2} \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \left(\mathbf{I}_{T-1} - \theta (1 + \theta' \theta)^{-1} \theta' \right) \mathbf{y}_{i,-1} - \frac{1}{\sigma_u^2} \frac{1}{1 + \theta' \theta} \frac{2}{NT} \sum_{i=1}^N \theta' \mathbf{y}_{i,-1} y_{i0} \\ &+ \frac{1}{\sigma_u^2} \frac{\theta' \theta}{1 + \theta' \theta} \frac{1}{NT} \sum_{i=1}^N y_{i0}^2 \\ &= \frac{1}{\sigma_u^2} \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \left(\mathbf{I}_{T-1} - \theta (1 + \theta' \theta)^{-1} \theta' \right) \mathbf{y}_{i,-1} + O_p\left(\frac{1}{T}\right) \\ &= \frac{1}{\sigma_u^2} \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \mathbf{y}_{i,-1} - \frac{1}{\sigma_u^2} \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \theta (1 + \theta' \theta)^{-1} \theta' \mathbf{y}_{i,-1} + O_p\left(\frac{1}{T}\right), \end{aligned} \quad (\text{A.2})$$

where the penultimate identity holds since $\theta' \theta = O(T)$.

By continuous substitution,

$$y_{it} = \gamma^t y_{i0} + \lambda_i \sum_{s=1}^t \gamma^{t-s} f_s + \sum_{s=1}^t \gamma^{t-s} u_{is}, \quad (\text{A.3})$$

then we have

$$\begin{aligned} & \lim_{(N,T) \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N E(\mathbf{y}_{i,-1}' \mathbf{y}_{i,-1}) \\ &= \lim_{(N,T) \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-1} \gamma^{2t} E(y_{i0}^2) + \lim_{(N,T) \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-1} \lambda_i^2 \left(\sum_{s=1}^t \gamma^{t-s} f_s \right)^2 \\ &+ \lim_{(N,T) \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T-1} \left(\sum_{s=1}^t \gamma^{t-s} u_{is} \right)^2 \\ &= \bar{\sigma}_\lambda^2 \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^{T-1} \left(\sum_{s=1}^t \gamma^{t-s} f_s \right)^2 + \frac{\sigma_u^2}{1 - \gamma^2}, \end{aligned} \quad (\text{A.4})$$

where $\bar{\sigma}_\lambda^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \lambda_i^2$.

Also, for the second term of (A.2), we have

$$\frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \theta (1 + \theta' \theta)^{-1} \theta' \mathbf{y}_{i,-1} = \frac{T}{1 + \theta' \theta} \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{T} \mathbf{y}_{i,-1}' \theta \right)^2, \quad (\text{A.5})$$

whereas $T \rightarrow \infty$, $T(1 + \theta' \theta)^{-1}$ will converge to a constant and

$$\frac{1}{T} \mathbf{y}_{i,-1}' \boldsymbol{\theta} = \lambda_i \frac{1}{T} \sum_{t=1}^{T-1} \theta_{t+1} \sum_{s=1}^t \gamma^{t-s} f_s + o_p(1) = O_p(1),$$

it follows that

$$\begin{aligned} \frac{1}{NT} \sum_{i=1}^N \mathbf{y}_{i,-1}' \boldsymbol{\theta} (1 + \boldsymbol{\theta}' \boldsymbol{\theta})^{-1} \boldsymbol{\theta}' \mathbf{y}_{i,-1} &= \frac{T}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^N \left(\lambda_i \frac{1}{T} \sum_{t=1}^{T-1} \theta_{t+1} \sum_{s=1}^t \gamma^{t-s} f_s \right)^2 + o_p(1) \\ &= \bar{\sigma}_\lambda^2 \frac{1}{T(1 + \boldsymbol{\theta}' \boldsymbol{\theta})} \left(\sum_{t=1}^{T-1} \theta_{t+1} \sum_{s=1}^t \gamma^{t-s} f_s \right)^2 + o_p(1) = O_p(1). \end{aligned} \quad (\text{A.6})$$

Combining (A.4) and (A.6), and letting $\omega_t = \sum_{s=1}^t \gamma^{t-s} f_s$ (which is $O(1)$) and $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{T-1})'$ yields (4.12).

By similar manipulation and taking account of it is i.i.d, we can show (4.13).

Similarly, for $\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}^{*\prime} \Omega^{-1} \mathbf{u}_i^*$, we have

$$\begin{aligned} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}^{*\prime} \Omega^{-1} \mathbf{u}_i^* &= \frac{1}{\sigma_u^2} \frac{1}{\sqrt{NT}} \sum_{i=1}^N (\mathbf{y}_{i,-1}' - \boldsymbol{\theta}' y_{i0}) (\mathbf{I}_{T-1} - \boldsymbol{\theta} (1 + \boldsymbol{\theta}' \boldsymbol{\theta})^{-1} \boldsymbol{\theta}') (\mathbf{u}_i - \boldsymbol{\theta} u_{i1}) \\ &= \frac{1}{\sigma_u^2} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}' \mathbf{u}_i - \frac{1}{\sigma_u^2} \frac{1}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}' \boldsymbol{\theta} \boldsymbol{\theta}' \mathbf{u}_i \\ &\quad - \frac{1}{\sigma_u^2} \frac{1}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}' u_{i1} \boldsymbol{\theta} - \frac{1}{\sigma_u^2} \frac{1}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N \boldsymbol{\theta}' y_{i0} \mathbf{u}_i \\ &\quad + \frac{1}{\sigma_u^2} \frac{\boldsymbol{\theta}' \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N y_{i0} u_{i1} \\ &= A_1 + A_2 + A_3 + A_4 + A_5, \end{aligned}$$

where A_1 will contribute to the limiting distribution.

For A_2 , it will contribute to the asymptotical bias since

$$E(A_2) = -\frac{\boldsymbol{\theta}' \Psi \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \sqrt{\frac{N}{T}}, \quad (\text{A.7})$$

by following the derivation in the main paper and conditional on $\boldsymbol{\theta}$,

$$\begin{aligned} E[\dot{\mathbf{u}}_i \dot{\mathbf{y}}_{i,-1}'] &= E \left(\begin{array}{cccc} u_{i2} y_{i1} & u_{i2} y_{i2} & \cdots & u_{i2} y_{i,T-1} \\ u_{i3} y_{i1} & u_{i3} y_{i2} & \cdots & u_{i3} y_{i,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{iT} y_{i1} & u_{iT} y_{i2} & \cdots & u_{iT} y_{i,T-1} \end{array} \right) \\ &= \sigma_u^2 \left(\begin{array}{cccc} 0 & 1 & \cdots & \gamma^{T-3} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & 0 \end{array} \right) = \sigma_u^2 \Psi. \end{aligned} \quad (\text{A.8})$$

Also, it can be shown that $\text{Var}(A_2) = O\left(\frac{1}{T}\right) = o(1)$. Thus, we obtain

$$A_2 \rightarrow_p -\frac{\boldsymbol{\theta}' \Psi \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \sqrt{\frac{N}{T}},$$

as $(N, T) \rightarrow \infty$.

For A_3 , it can be shown to be $o_p(1)$ as $(N, T) \rightarrow \infty$ since

$$\begin{aligned} E(A_3) &= -\frac{1}{\sigma_u^2} \frac{1}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N E(\mathbf{y}_{i,-1}' u_{i1}) \boldsymbol{\theta} \\ &= -\frac{1}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N (1, \gamma, \dots, \gamma^{T-2}) \boldsymbol{\theta} \\ &= O\left(\frac{N^{1/2}}{T^{3/2}}\right) = o(1), \end{aligned}$$

as long as $\frac{N}{T^3} \rightarrow 0$ (which is true if $\frac{N}{T} \rightarrow a \neq 0$ as $(N, T) \rightarrow \infty$) and because

$1 + \boldsymbol{\theta}' \boldsymbol{\theta} = O(T)$ and $E(A_3^2) = O\left(\frac{1}{T}\right) + O\left(\frac{N}{T^2}\right) = o(1)$, which implies $A_3 = o_p(1)$.

For A_4 , we have

$$E(A_4) = -\frac{1}{\sigma_u^2} \frac{1}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N E(y_{i0} \boldsymbol{\theta}' \mathbf{u}_i) = 0,$$

and $E(A_4^2) = O\left(\frac{1}{T^2}\right)$, which implies $A_4 = o_p(1)$.

For A_5 , we have

$$E(A_5) = \frac{1}{\sigma_u^2} \frac{\boldsymbol{\theta}' \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \frac{1}{\sqrt{NT}} \sum_{i=1}^N E(y_{i0} u_{i1}) = 0,$$

and

$$\begin{aligned} E(A_5^2) &= \frac{1}{\sigma_u^4} \left(\frac{\boldsymbol{\theta}' \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \right)^2 \frac{1}{NT} \sum_{i,j} E(y_{i0} u_{i1} y_{j0} u_{j1}) \\ &= \frac{1}{\sigma_u^4} \left(\frac{\boldsymbol{\theta}' \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}} \right)^2 \frac{1}{NT} \sum_i E(y_{i0}^2 u_{i1}^2) = O\left(\frac{1}{T}\right), \end{aligned}$$

which gives $A_5 = o_p(1)$.

Combining these results yields (4.14). Following the derivation of [Anderson \(1971, Ch 5\)](#), we can show that

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}' \mathbf{u}_i \rightarrow_d N\left(0, \frac{\sigma_y^2}{\sigma_u^2}\right),$$

where

$$\sigma_y^2 = \frac{1}{T} \sum_{t=1}^T E(y_{it}^2) \rightarrow \bar{\sigma}_\lambda^2 \tilde{\sigma}_{\gamma,f}^2 + \frac{\sigma_u^2}{1 - \gamma^2},$$

$$\tilde{\sigma}_{\gamma,f}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \boldsymbol{\omega}' \boldsymbol{\omega} \text{ and } \boldsymbol{\omega} = (\omega_1, \dots, \omega_{T-1})' \text{ with } \omega_t = \sum_{s=1}^t \gamma^{t-s} f_s.$$

Equivalently, we have

$$\frac{1}{\sqrt{NT}} \sum_{i=1}^N \mathbf{y}_{i,-1}^* (\boldsymbol{\theta})' \Omega(\boldsymbol{\theta})^{-1} \mathbf{u}_i^* (\boldsymbol{\theta}) \rightarrow_d N \left(-b \sqrt{\frac{N}{T}}, \frac{\bar{\sigma}_\lambda^2 \tilde{\sigma}_{\gamma,f}^2}{\sigma_u^2} + \frac{1}{1-\gamma^2} \right), \quad (\text{A.9})$$

where

$$b = \lim_{T \rightarrow \infty} \frac{\boldsymbol{\theta}' \Psi \boldsymbol{\theta}}{1 + \boldsymbol{\theta}' \boldsymbol{\theta}}, \quad (\text{A.10})$$

denoting the bias term with $\boldsymbol{\theta} = (\theta_2, \dots, \theta_T)'$ is a $(T-1) \times 1$ vector with unrestricted parameters and Ψ is provided in (A.8), $\bar{\sigma}_\lambda^2$ is defined as before and

$$\tilde{\sigma}_{\gamma,f}^2 = \lim_{T \rightarrow \infty} \frac{1}{T} \boldsymbol{\omega}' \boldsymbol{\omega} \text{ with } \boldsymbol{\omega} = (\omega_1, \dots, \omega_{T-1})' \text{ and } \omega_t = \sum_{s=1}^t \gamma^{t-s} f_s.$$

Combining (A.6) and (A.9) yields Theorem 4.1 as required.

B. QDMLE FOR MODEL WITH MULTIPLE FACTORS

When $r > 1$, the regression error v_{it} takes the form of

$$v_{it} = \boldsymbol{\lambda}_i' \mathbf{f}_i + u_{it}, \quad (\text{A.11})$$

where $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{ir})'$ denotes the factor loadings while $\mathbf{f}_i = (f_{i1}, \dots, f_{ir})'$ denotes unobservable common factors. Let $\mathbf{v}_i = (v_{i1}, \dots, v_{iT})'$, then (A.11) can be rewritten in vector form as

$$\mathbf{v}_i = \mathbf{F} \boldsymbol{\lambda}_i + \mathbf{u}_i, i=1,2,\dots,N, \quad (\text{A.12})$$

where $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)'$ denotes the $T \times r$ matrix of r time-specific common factors, \mathbf{f}_i , over time and $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$. Similarly, let $\mathbf{v}_t = (v_{1t}, \dots, v_{Nt})'$, then (A.11) can be rewritten in vector form as

$$\mathbf{v}_t = \Lambda \mathbf{f}_t + \mathbf{u}_t, \quad (\text{A.13})$$

where $\Lambda = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_N)'$ denotes the $N \times r$ matrix of factor loading and $\mathbf{u}_t = (u_{1t}, \dots, u_{Nt})'$.

For model (2.1) with multiple common factor structure (A.11), it can be stacked in vector form as

$$\mathbf{y}_i = \gamma \mathbf{y}_{i,-1} + \mathbf{F} \boldsymbol{\lambda}_i + \mathbf{u}_i, \quad (\text{A.14})$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$, $\mathbf{y}_{i,-1} = (y_{i0}, \dots, y_{i,T-1})'$ and $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$.

Note that \mathbf{F} and $\boldsymbol{\lambda}_i$ in (A.12) or (A.14) are not separately identified since for any full rank matrix \mathbf{C} , we have

$$\mathbf{F}\boldsymbol{\lambda}_i = \mathbf{F}\mathbf{C}\mathbf{C}^{-1}\boldsymbol{\lambda}_i = \mathbf{F}^*\boldsymbol{\lambda}_i^*, \quad (\text{A.15})$$

which is the typical rotation problem for models with interactive effects (Bai, 2009). To avoid this problem, following ALS (2013), we use the normalization condition $\mathbf{F} = (-\mathbf{I}_r, \Xi')$, where Ξ is a $(T-r) \times r$ matrix of unrestricted parameters. Define the $T \times (T-r)$ matrix

$$\mathbf{H}(\vartheta) = (\Xi, \mathbf{I}_{T-r})' = [\mathbf{h}_1(\vartheta_1), \mathbf{h}_2(\vartheta_2), \dots, \mathbf{h}_{T-r}(\vartheta_{T-r})], \quad (\text{A.16})$$

where $\vartheta = \text{vec}(\Xi)$, ϑ_j is the j th column of Ξ' , and $\mathbf{h}_j(\vartheta_j)$ is the j th column of $\mathbf{H}(\vartheta)$ for $j = 1, \dots, T-r$. Given the definition of $\mathbf{H}(\vartheta)$, we have

$$\mathbf{H}(\vartheta)' \mathbf{F} = (\Xi, \mathbf{I}_{T-r})(-\mathbf{I}_r, \Xi')' = \mathbf{0}_{(T-r) \times r}. \quad (\text{A.17})$$

As a result, we can remove the interactive effects in (A.14) by multiplying $\mathbf{H}(\vartheta)'$ in both sides of (A.14), which in turn gives

$$\mathbf{H}(\vartheta)' \mathbf{y}_i = \mathbf{H}(\vartheta)' \mathbf{y}_{i,-1} \gamma + \mathbf{H}(\vartheta)' \mathbf{u}_i, \quad i=1, \dots, N. \quad (\text{A.18})$$

For the transformed errors, we have

$$\mathbf{H}(\vartheta)' \mathbf{u}_i = \begin{pmatrix} \mathbf{h}_1(\vartheta_1)' \mathbf{u}_i \\ \mathbf{h}_2(\vartheta_2)' \mathbf{u}_i \\ \vdots \\ \mathbf{h}_{T-r}(\vartheta_{T-r})' \mathbf{u}_i \end{pmatrix} = \begin{pmatrix} u_{i,r+1} + \vartheta_1' \mathbf{u}_{i,1,r} \\ u_{i,r+2} + \vartheta_2' \mathbf{u}_{i,1,r} \\ \vdots \\ u_{i,T} + \vartheta_{T-r}' \mathbf{u}_{i,1,r} \end{pmatrix} = \mathbf{u}_{i,\underline{r+1,T}} + \Xi \mathbf{u}_{i,\underline{1,r}},$$

with $\mathbf{u}_{i,\underline{1,r}} = (u_{i,1}, \dots, u_{i,r})'$ and $\mathbf{u}_{i,\underline{r+1,T}} = (u_{i,r+1}, \dots, u_{iT})'$.

Similarly, we have

$$\mathbf{H}(\vartheta)' \mathbf{y}_i = \begin{pmatrix} \mathbf{h}_1(\vartheta_1)' \mathbf{y}_i \\ \mathbf{h}_2(\vartheta_2)' \mathbf{y}_i \\ \vdots \\ \mathbf{h}_{T-r}(\vartheta_{T-r})' \mathbf{y}_i \end{pmatrix} = \begin{pmatrix} y_{i,r+1} + \vartheta_1' \mathbf{y}_{i,1,r} \\ y_{i,r+2} + \vartheta_2' \mathbf{y}_{i,1,r} \\ \vdots \\ y_{iT} + \vartheta_{T-r}' \mathbf{y}_{i,1,r} \end{pmatrix} = \mathbf{y}_{i,\underline{r+1,T}} + \Xi \mathbf{y}_{i,\underline{1,r}},$$

with $\mathbf{y}_{i,\underline{1,r}} = (y_{i,1}, \dots, y_{i,r})'$ and $\mathbf{y}_{i,\underline{r+1,T}} = (y_{i,r+1}, \dots, y_{iT})'$, and

$$\mathbf{H}(\vartheta)' \mathbf{y}_{i,-1} = \begin{pmatrix} \mathbf{h}_1(\vartheta_1)' \mathbf{y}_{i,-1} \\ \mathbf{h}_2(\vartheta_2)' \mathbf{y}_{i,-1} \\ \vdots \\ \mathbf{h}_{T-r}(\vartheta_{T-r})' \mathbf{y}_{i,-1} \end{pmatrix} = \begin{pmatrix} y_{i,r} + \vartheta_1' \mathbf{y}_{i,0,r-1} \\ y_{i,r+2} + \vartheta_2' \mathbf{y}_{i,0,r-1} \\ \vdots \\ y_{iT-1} + \vartheta_{T-r}' \mathbf{y}_{i,0,r-1} \end{pmatrix} = \mathbf{y}_{i,\underline{r,T-1}} + \Xi \mathbf{y}_{i,\underline{0,r-1}},$$

with $\mathbf{y}_{i,0,r-1} = (y_{i,0}, \dots, y_{i,r-1})'$ and $\mathbf{y}_{i,r,T-1} = (y_{i,r+1}, \dots, y_{iT})'$.

For the transformed model (A.18), the variance-covariance matrix is given by

$$\begin{aligned}\Omega_{TS} &= E[\mathbf{H}(\vartheta)' \mathbf{u}_i \mathbf{u}_i' \mathbf{H}(\vartheta)] = E[(\mathbf{u}_{i,r+1,T} + \Xi \mathbf{u}_{i,1,r})(\mathbf{u}_{i,r+1,T} + \Xi \mathbf{u}_{i,1,r})'] \\ &= \sigma_u^2 (\mathbf{I}_{T-r} + \Xi \Xi'). \quad (\text{A.19})\end{aligned}$$

In order to obtain an explicit solution of $(\mathbf{I}_{T-r} + \Xi \Xi')^{-1}$, assume Ξ has a singular value decomposition such that $\Xi = \mathbf{U} \mathbf{D} \mathbf{V}$ where \mathbf{U} and \mathbf{V} are $(T-r) \times (T-r)$ and $r \times r$ orthogonal matrices (Magnus & Neudecker, 1999), respectively, and \mathbf{D} is the $(T-r) \times r$ matrix of singular values of Ξ on its diagonal

$$\mathbf{D} = \begin{pmatrix} \text{diag}(d_1, d_2, \dots, d_r) \\ 0 \end{pmatrix},$$

where $d_1 \geq d_2 \geq \dots \geq d_r$. Then it is obvious that $\Xi \Xi' = \mathbf{U} \mathbf{D} \mathbf{D}' \mathbf{U}'$, and $\mathbf{D} \mathbf{D}'$ is the $(T-r) \times (T-r)$ diagonal matrix

$$\mathbf{D} \mathbf{D}' = \begin{pmatrix} \text{diag}(d_1^2, d_2^2, \dots, d_r^2) & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$\mathbf{I}_{T-r} + \Xi \Xi' = \mathbf{U} (\mathbf{I}_{T-r} + \mathbf{D} \mathbf{D}') \mathbf{U}',$$

thus

$$(\mathbf{I}_{T-r} + \Xi \Xi')^{-1} = \mathbf{U} \begin{pmatrix} \overset{\circ}{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix} \mathbf{U}',$$

with $\overset{\circ}{\mathbf{D}} = \text{diag}\left(\frac{1}{1+d_1^2}, \frac{1}{1+d_2^2}, \dots, \frac{1}{1+d_r^2}\right)$.

As a result, we have

$$\begin{aligned}\mathbf{H}(\vartheta)' (\mathbf{I}_{T-r} + \Xi \Xi')^{-1} \mathbf{H}(\vartheta) &= \begin{pmatrix} \mathbf{V}' \mathbf{D}' \mathbf{U}' \\ \mathbf{I}_{T-r} \end{pmatrix} \mathbf{U} \begin{pmatrix} \overset{\circ}{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix} \mathbf{U}' (\mathbf{U} \mathbf{D} \mathbf{V}, \mathbf{I}_{T-r}) \\ &= \begin{pmatrix} \mathbf{V}' \mathbf{D}' \\ \mathbf{U} \end{pmatrix} \begin{pmatrix} \overset{\circ}{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix} (\mathbf{D} \mathbf{V}, \mathbf{U}). \quad (\text{A.20})\end{aligned}$$

The quasi-log-likelihood function of (A.18) is given by

$$\log L(\gamma, \vartheta) = -\frac{N}{2} \log |\Omega_{TS}| - \frac{1}{2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{y}_{i,-1} \gamma)' \mathbf{H}(\vartheta) \Omega_{TS}^{-1} \mathbf{H}(\vartheta)' (\mathbf{y}_i - \mathbf{y}_{i,-1} \gamma). \quad (\text{A.21})$$

Conditional on ϑ , the FOC for γ is

$$\frac{\partial \log L(\gamma, \vartheta)}{\partial \gamma} = \sum_{i=1}^N \mathbf{y}_{i,-1}' \mathbf{H}(\vartheta) \Omega_{TS}^{-1} \mathbf{H}(\vartheta)' (\mathbf{y}_i - \mathbf{y}_{i,-1} \gamma) = 0, \quad (\text{A.22})$$

which yields

$$\hat{\gamma}_{MLE}^{TS,M} = \left(\sum_{i=1}^N \mathbf{y}_{i,-1}' \mathbf{H}(\vartheta) \Omega_{TS}^{-1} \mathbf{H}(\vartheta)' \mathbf{y}_{i,-1} \right)^{-1} \left(\sum_{i=1}^N \mathbf{y}_{i,-1}' \mathbf{H}(\vartheta) \Omega_{TS}^{-1} \mathbf{H}(\vartheta)' \mathbf{y}_i \right), \quad (\text{A.23})$$

where TS,M refers to using normalization (A.17) to remove the multiple interactive effects.

The denominator of (A.23) divided by NT can be shown to converge to a nonzero constant. For the numerator, we can observe that

$$\begin{aligned} E[\mathbf{y}_{i,-1}' \mathbf{H}(\vartheta) \Omega_{TS}^{-1} \mathbf{H}(\vartheta)' \mathbf{u}_i] &= tr\{\mathbf{H}(\vartheta) \Omega_{TS}^{-1} \mathbf{H}(\vartheta)' E[\mathbf{u}_i \mathbf{y}_{i,-1}']\} \\ &= tr\{\mathbf{H}(\vartheta) (\mathbf{I}_{T-r} + \Xi \Xi')^{-1} \mathbf{H}(\vartheta)' \dot{\Psi}\}, \end{aligned}$$

by using the derivation of (A.8) and

$$E[\mathbf{u}_i \mathbf{y}_{i,-1}'] = \sigma_u^2 \begin{pmatrix} 0 & 1 & \dots & \gamma^{T-2} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix} = \sigma_u^2 \dot{\Psi}. \quad (\text{A.24})$$

Using the results of (A.20) yields

$$\begin{aligned} tr\{\mathbf{H}(\vartheta) (\mathbf{I}_{T-r} + \Xi \Xi')^{-1} \mathbf{H}(\vartheta)' \dot{\Psi}\} &= tr\left[\begin{pmatrix} \mathring{\mathbf{D}} & 0 \\ 0 & \mathbf{I}_{T-2r} \end{pmatrix} (\mathbf{DV}, \mathbf{U}') \dot{\Psi} \begin{pmatrix} \mathbf{V}' \mathbf{D}' \\ \mathbf{U} \end{pmatrix} \right] \\ &\leq C tr\left[\dot{\Psi} \begin{pmatrix} \mathbf{V}' \mathbf{D}' \\ \mathbf{U} \end{pmatrix} (\mathbf{DV}, \mathbf{U}') \right] \\ &\leq C tr\left[\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{V}' \mathbf{D}' \mathbf{D}' \mathbf{V} & \mathbf{V}' \mathbf{D}' \mathbf{U}' \\ \mathbf{U} \mathbf{D}' \mathbf{V} & \mathbf{I}_{T-r} \end{pmatrix} \right] \end{aligned}$$

where we partition $\dot{\Psi}$ as follows

$$\Psi_{11} = \begin{pmatrix} 0 & 1 & \gamma & \dots & \gamma^{r-2} \\ 0 & 0 & 1 & \dots & \gamma^{r-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}_{r \times r}, \quad \Psi_{12} = \begin{pmatrix} \gamma^{r-1} & \gamma^r & \dots & \gamma^{T-2} \\ \gamma^{r-2} & \gamma^{r-1} & \dots & \gamma^{T-3} \\ \vdots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ 1 & \gamma & \dots & \gamma^{T-r-1} \end{pmatrix}_{r \times (T-r)},$$

$$\Psi_{21} = \mathbf{0}_{(T-r) \times r}, \Psi_{22} = \begin{pmatrix} 0 & 1 & \gamma & \cdots & \gamma^{T-r-2} \\ 0 & 0 & 1 & \cdots & \gamma^{T-r-3} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}_{(T-r) \times (T-r)},$$

then

$$\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ 0 & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V} & \mathbf{V}' \mathbf{D}' \mathbf{U}' \\ \mathbf{U} \mathbf{D} \mathbf{V} & \mathbf{I}_{T-r} \end{pmatrix} = \begin{pmatrix} \Psi_{11} \mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V} + \Psi_{12} \mathbf{U} \mathbf{D} \mathbf{V} & \Psi_{11} \mathbf{V}' \mathbf{D}' \mathbf{U}' + \Psi_{12} \\ \Psi_{22} \mathbf{U} \mathbf{D} \mathbf{V} & \Psi_{22} \end{pmatrix}.$$

Taking trace, we obtain

$$\begin{aligned} \text{tr} \left[\begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix} \begin{pmatrix} \mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V} & \mathbf{V}' \mathbf{D}' \mathbf{U}' \\ \mathbf{U} \mathbf{D} \mathbf{V} & \mathbf{I}_{T-r} \end{pmatrix} \right] &= \text{tr}(\Psi_{11} \mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V} + \Psi_{12} \mathbf{U} \mathbf{D} \mathbf{V}) + \text{tr}(\Psi_{22}) \\ &= \text{tr}(\Psi_{11} \mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V}) + \text{tr}(\Psi_{12} \Xi), \end{aligned}$$

since $\text{tr}(\Psi_{22}) = 0$.

Using the result of Fang et al. (1994, p. 2489), we have

$$\begin{aligned} |\text{tr}(\Psi_{11} \mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V})| &\leq (|\lambda_{\max}(\frac{\Psi_{11} + \Psi'_{11}}{2})| + |\lambda_{\min}(\frac{\Psi_{11} + \Psi'_{11}}{2})|) \text{tr}(\mathbf{V}' \mathbf{D}' \mathbf{D} \mathbf{V}) \\ &= (|\lambda_{\max}(\frac{\Psi_{11} + \Psi'_{11}}{2})| + |\lambda_{\min}(\frac{\Psi_{11} + \Psi'_{11}}{2})|) \sum_{i=1}^r d_i^2 \\ &= O(1), \end{aligned}$$

the last equation follows the fact $\frac{\Psi_{11} + \Psi'_{11}}{2}$ is a $r \times r$ matrix with bounded elements, so it has bounded smallest and largest eigenvalues. For the last term of (A.25), we notice that the diagonal element $a_{ii}, i = 1, 2, \dots, r$, of $\Psi_{12} \Xi$ is given by

$$a_{ii} = \frac{1}{\gamma^{i+1}} \sum_{t=1}^{T-r} \gamma^{r+t} \theta_{it},$$

thus

$$\begin{aligned} |a_{ii}| &\leq \frac{1}{\gamma^{i+1}} \sum_{t=1}^{T-r} \gamma^{r+t} |\theta_{it}| \leq \frac{1}{\gamma} \sum_{t=1}^{T-r} \gamma^t |\theta_{it}| \\ &= O(1), \end{aligned}$$

as $T \rightarrow \infty$, which in turn yields

$$\begin{aligned} |\text{tr}(\Psi_{12} \Xi)| &= \left| \sum_{i=1}^r a_{ii} \right| \leq \sum_{i=1}^r |a_{ii}| \\ &= O(1), \end{aligned}$$

as long as r is finite.

Combining these results we obtain

$$\text{tr} \left\{ \mathbf{H}(\vartheta) \left(\mathbf{I}_{T-r} + \Xi \Xi' \right)^{-1} \mathbf{H}(\vartheta)' \dot{\Psi} \right\} = O(1), \quad (\text{A.26})$$

as $T \rightarrow \infty$.

Substituting (A.26) into (A.23), we get $\frac{1}{\sqrt{NT}} \sum_{i=1}^N (\mathbf{y}'_{i,-1} \mathbf{H}(\vartheta) \Omega_{TS}(\vartheta)^{-1} \mathbf{H}(\vartheta)' \mathbf{u}_i) = O_p \left(\sqrt{\frac{N}{T}} \right)$. i.e., the QDMLE for γ using normalization (A.17) is asymptotically biased of order $\sqrt{\frac{N}{T}}$.

Alternatively, model (2.1) with (A.11) can be rewritten in vector form as

$$\mathbf{y}_t = \gamma \mathbf{y}_{t-1} + \Lambda \mathbf{f}_t + \mathbf{u}_t, \quad t=1, \dots, T, \quad (\text{A.27})$$

where $\mathbf{y}_t = (y_{1,t}, \dots, y_{N,t})'$, $\mathbf{y}_{t-1} = (y_{1,t-1}, \dots, y_{N,t-1})'$, $\mathbf{u}_t = (u_{1,t}, \dots, u_{N,t})'$ and $\Lambda = (\lambda_1, \dots, \lambda_N)'$ is a $N \times r$ matrix of factor loading.

For model (A.27), we can apply the quasi-difference pairwise proposed before. To this end, let the factor loading matrix Λ satisfies the normalization such that $\Lambda = (-\mathbf{I}_r, \Psi')'$, where Ψ is an $(N-r) \times r$ matrix of unrestricted parameters. Define the $N \times (N-r)$ matrix

$$\mathbf{G}(\psi) = (\Psi, \mathbf{I}_{N-r})' = [\mathbf{g}_1(\psi_1), \mathbf{g}_2(\psi_2), \dots, \mathbf{g}_{N-r}(\psi_{N-r})], \quad (\text{A.28})$$

where $\psi = \text{vec}(\Psi)$, ψ_j is the j th column of Ψ' , and $\mathbf{g}_j(\psi_j)$ is the j th column of $\mathbf{G}(\psi)$ for $j = 1, \dots, N-r$. Given the definition of $\mathbf{G}(\psi)$, we have

$$\mathbf{G}(\psi)' \Lambda = (\Psi, \mathbf{I}_{N-r})(-\mathbf{I}_r, \Psi')' = 0_{(N-r) \times r}. \quad (\text{A.29})$$

As a result, we can remove the interactive effects in (A.27) by multiplying $\mathbf{G}(\psi)'$ in both sides of (A.27), which in turn gives

$$\mathbf{G}(\psi)' \mathbf{y}_t = \gamma \mathbf{G}(\psi)' \mathbf{y}_{t-1} + \mathbf{G}(\psi)' \mathbf{u}_t, \quad t=1, \dots, T, \quad (\text{A.30})$$

or

$$\check{\mathbf{y}}_t(\psi) = \gamma \check{\mathbf{y}}_{t-1}(\psi) + \check{\mathbf{u}}_t(\psi), \quad t=1, \dots, T, \quad (\text{A.31})$$

where $\check{\mathbf{y}}_t(\psi) = \mathbf{G}(\psi)' \mathbf{y}_t$, $\check{\mathbf{y}}_{t-1}(\psi) = \mathbf{G}(\psi)' \mathbf{y}_{t-1}$ and $\check{\mathbf{u}}_t(\psi) = \mathbf{G}(\psi)' \mathbf{u}_t$.

For the transformed errors, we have

$$\check{\mathbf{u}}_t(\psi) = \mathbf{G}(\psi)' \mathbf{u}_t = \begin{pmatrix} \mathbf{g}_1(\psi_1)' \mathbf{u}_t \\ \mathbf{g}_2(\psi_2)' \mathbf{u}_t \\ \vdots \\ \mathbf{g}_{N-r}(\psi_{N-r})' \mathbf{u}_t \end{pmatrix} = \begin{pmatrix} u_{r+1,t} + \psi_1' \mathbf{u}_{1,r,t} \\ u_{r+2,t} + \psi_2' \mathbf{u}_{1,r,t} \\ \vdots \\ u_{N,t} + \psi_{N-r}' \mathbf{u}_{1,r,t} \end{pmatrix} = \underline{\mathbf{u}}_{r+1,N,t} + \Psi \underline{\mathbf{u}}_{1,r,t},$$

with $\mathbf{u}_{\underline{1},r,t} = (u_{1t}, \dots, u_{rt})'$ and $\mathbf{u}_{\underline{r+1},N,t} = (u_{r+1,t}, \dots, u_{Nt})'$.

Similarly, we have

$$\dot{\mathbf{y}}_t(\psi) = \mathbf{G}(\psi)' \mathbf{y}_t = \begin{pmatrix} \mathbf{g}_1(\psi_1)' \mathbf{y}_t \\ \mathbf{g}_2(\psi_2)' \mathbf{y}_t \\ \vdots \\ \mathbf{g}_{N-r}(\psi_{N-r})' \mathbf{y}_t \end{pmatrix} = \begin{pmatrix} y_{r+1,t} + \psi_1' \mathbf{y}_{1,r,t} \\ y_{r+2,t} + \psi_2' \mathbf{y}_{1,r,t} \\ \vdots \\ y_{N,t} + \psi_{N-r}' \mathbf{y}_{1,r,t} \end{pmatrix} = \mathbf{y}_{\underline{r+1},N,t} + \Psi \mathbf{y}_{\underline{1},r,t},$$

with $\mathbf{y}_{\underline{1},r,t} = (y_{1t}, \dots, y_{rt})'$ and $\mathbf{y}_{\underline{r+1},N,t} = (y_{r+1,t}, \dots, y_{Nt})'$, and

$$\dot{\mathbf{y}}_{t-1}(\psi) = \mathbf{G}(\psi)' \mathbf{y}_{t-1} = \begin{pmatrix} \mathbf{g}_1(\psi_1)' \mathbf{y}_{t-1} \\ \mathbf{g}_2(\psi_2)' \mathbf{y}_{t-1} \\ \vdots \\ \mathbf{g}_{N-r}(\psi_{N-r})' \mathbf{y}_{t-1} \end{pmatrix} = \begin{pmatrix} y_{r+1,t-1} + \psi_1' \mathbf{y}_{1,r,t-1} \\ y_{r+2,t-1} + \psi_2' \mathbf{y}_{1,r,t-1} \\ \vdots \\ y_{N,t-1} + \psi_{N-r}' \mathbf{y}_{1,r,t-1} \end{pmatrix} = \mathbf{u}_{\underline{r+1},N,t-1} + \Psi \mathbf{u}_{\underline{1},r,t-1},$$

with $\mathbf{y}_{\underline{1},r,t-1} = (y_{1,t-1}, \dots, y_{r,t-1})'$ and $\mathbf{y}_{\underline{r+1},N,t-1} = (y_{r+1,t-1}, \dots, y_{N,t-1})'$.

For the transformed model (A.30), the variance-covariance matrix is given by

$$\Omega_{CS} = \sigma_u^2 (\mathbf{I}_{N-r} + \Psi \Psi'), \quad (A.32)$$

and

$$\Omega_{CS}^{-1} = \frac{1}{\sigma_u^2} (\mathbf{I}_{N-r} - \Psi (\mathbf{I}_r + \Psi' \Psi)^{-1} \Psi'). \quad (A.33)$$

The quasi-log-likelihood function of (A.30) is given by

$$\log L(\gamma, \psi) = -\frac{N}{2} \log |\Omega_{CS}| - \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \gamma \mathbf{y}_{t-1})' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) (\mathbf{y}_t - \gamma \mathbf{y}_{t-1}). \quad (A.34)$$

Conditional on ψ , the FOC for γ is

$$\frac{\partial \log L(\gamma, \psi)}{\partial \gamma} = \sum_{t=1}^T \mathbf{y}_{t-1}' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) (\mathbf{y}_t - \gamma \mathbf{y}_{t-1}) = 0, \quad (A.35)$$

which yields

$$\hat{\gamma}_{MLE}^{CS,M} = \left(\sum_{t=1}^T \mathbf{y}_{t-1}' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{y}_{t-1} \right)^{-1} \sum_{t=1}^T (\mathbf{y}_{t-1}' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{y}_t), \quad (A.36)$$

where CS,M refers to using normalization (A.29) to remove the multiple interactive effects.

For this QDMLE (A.36), we have

$$\begin{aligned} \sqrt{NT}(\hat{\gamma}_{MLE}^{CS,M} - \gamma) &= \left(\frac{1}{NT} \sum_{t=1}^T \mathbf{y}'_{t-1} \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{y}_{t-1} \right)^{-1} \\ &\quad \times \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbf{y}'_{t-1} \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{u}_t \right). \end{aligned} \quad (\text{A.37})$$

For the asymptotic distribution of $\hat{\gamma}_{MLE}^{CS,M}$, we notice that the transformed system (A.31) is stationary under assumption that $|\gamma| < 1$. Thus, as $(N, T) \rightarrow \infty$, the denominator of (A.37) converges to

$$\begin{aligned} \frac{1}{NT} \sum_{t=1}^T E[\mathbf{y}'_{t-1} \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{y}_{t-1}] &= \frac{1}{NT} \sum_{t=1}^T E[\tilde{\mathbf{y}}'_{t-1}(\psi)' \Omega_{CS}^{-1} \tilde{\mathbf{y}}_{t-1}(\psi)] \\ &\rightarrow \frac{1}{1-\gamma^2}. \end{aligned} \quad (\text{A.38})$$

For the numerator, we can observe that

$$\begin{aligned} E[\mathbf{y}'_{t-1} \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{u}_t] &= \text{tr}\{E[\mathbf{y}'_{t-1} \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{u}_t]\} \\ &= \text{tr}\{\mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) E(\mathbf{u}_t \mathbf{y}'_{t-1})\} \\ &= 0, \end{aligned} \quad (\text{A.39})$$

by using the fact that \mathbf{u}_t is independent of \mathbf{y}_{t-1} . As a result, we can claim that the QDMLE (A.36) is asymptotically unbiased in the sense that

$$E[\sqrt{NT}(\hat{\gamma}_{MLE}^{CS,M} - \gamma)] = 0.$$

which suggests that the QDMLE (A.36) is asymptotically unbiased, and the asymptotical unbiasedness of QDMLE is independent of the way how (N, T) go to infinity.

Online supplementary material to “Estimation of Dynamic Panel Data Models with Interactive Effects: Quasi-differencing Over Time or Across Individuals?”

Cheng Hsiao, Qiankun Zhou

This online supplement includes the mathematical proofs of the QDMLE for Model with both Exogenous Regressors and Interactive Effects. In the main text, we consider the QDMLE for dynamic panels without exogenous variables. Here we show that the QDMLE can be easily extended for dynamic panels with exogenous variables. Consider the general model

$$\begin{aligned} y_{it} &= \gamma y_{it-1} + \mathbf{x}'_{it} \boldsymbol{\beta} + \lambda'_i \mathbf{f}_t + u_{it} \\ &= \mathbf{z}'_{it} \boldsymbol{\delta} + \lambda'_i \mathbf{f}_t + u_{it}, \end{aligned} \quad (\text{OA.1})$$

where $\mathbf{z}_{it} = (y_{i,t-1}, \mathbf{x}'_{it})'$, $\boldsymbol{\delta} = (\gamma, \boldsymbol{\beta})'$ and \mathbf{x}_{it} is a $k \times 1$ vector of strictly exogenous variables with respect to u_{it} .

For the general model (OA.1), it can be stacked in vector form as

$$\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\delta} + \mathbf{F} \boldsymbol{\lambda}_i + \mathbf{u}_i, \quad i=1, \dots, N, \quad (\text{OA.2})$$

where $\mathbf{Z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{iT})'$, \mathbf{y}_i , \mathbf{F} , $\boldsymbol{\lambda}_i$ and \mathbf{u}_i are defined in (A.14). Using the same normalization (A.17), if we multiply $\mathbf{H}(\vartheta)'$ in both sides of (OA.2), then we have

$$\mathbf{H}(\vartheta)' \mathbf{y}_i = \mathbf{H}(\vartheta)' \mathbf{Z}_i \boldsymbol{\delta} + \mathbf{H}(\vartheta)' \mathbf{u}_i, \quad i=1, \dots, N, \quad (\text{OA.3})$$

then it is obvious that both $\mathbf{H}(\vartheta)' \mathbf{y}_i$ and $\mathbf{H}(\vartheta)' \mathbf{u}_i$ have the similar structure as in (A.18), and

$$\mathbf{H}(\vartheta)' \mathbf{Z}_i = \begin{pmatrix} \mathbf{h}_1(\vartheta_1)' \mathbf{Z}_i \\ \mathbf{h}_2(\vartheta_2)' \mathbf{Z}_i \\ \vdots \\ \mathbf{h}_{T-r}(\vartheta_{T-r})' \mathbf{Z}_i \end{pmatrix} = \mathbf{Z}_{i,\underline{r+1,T}} + \Xi \mathbf{Z}_{i,\underline{1,r}},$$

with $\mathbf{Z}_{i,\underline{1,r}} = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{ir})'$ and $\mathbf{Z}_{i,\underline{r+1,T}} = (\mathbf{z}_{i,r+1}, \dots, \mathbf{z}_{iT})'$.

For model (OA.3), the variance-covariance matrix is given by (A.19), and the quasi-log-likelihood function of (OA.3) is given by

$$\log L(\boldsymbol{\delta}, \vartheta) = -\frac{N}{2} \log |\Omega_{TS}| - \frac{1}{2} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{Z}_i \boldsymbol{\delta})' \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) (\mathbf{y}_i - \mathbf{Z}_i \boldsymbol{\delta}). \quad (\text{OA.4})$$

Conditional on ϑ , the FOC for $\boldsymbol{\delta}$ is

$$\frac{\partial \log L(\boldsymbol{\delta}, \vartheta)}{\partial \boldsymbol{\delta}} = \sum_{i=1}^N \mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) (\mathbf{y}_i - \mathbf{Z}_i \boldsymbol{\delta}) = 0, \quad (\text{OA.5})$$

which yields

$$\hat{\delta}_{MLE}^{TS} = \left(\sum_{i=1}^N \mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{y}_i \right), \quad (\text{OA.6})$$

where $_{TS}$ refers to using normalization (A.17) to remove the multiple interactive effects.

For this QDMLE (OA.6), we have

$$\begin{aligned} \sqrt{NT} \left(\hat{\delta}_{MLE}^{TS} - \boldsymbol{\delta} \right) &= \left(\frac{1}{NT} \sum_{i=1}^N \mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{Z}_i \right)^{-1} \\ &\quad \frac{1}{\sqrt{NT}} \sum_{i=1}^N \left(\mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i \right), \end{aligned} \quad (\text{OA.7})$$

where the denominator can be shown to converge to a nonsingular matrix. For the numerator, we can observe that

$$\mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i = \begin{pmatrix} \mathbf{y}'_{i,-1} \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i \\ \mathbf{X}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i \end{pmatrix},$$

and $E[\mathbf{u}_i \mathbf{y}'_{i,-1}] = \sigma_u^2 \dot{\Psi}$ with $\dot{\Psi}$ is given by (A.24), then

$$E[\mathbf{y}'_{i,-1} \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i] = O(1),$$

by following the above derivation. It can also be show that

$$E(\mathbf{X}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i) = E \begin{pmatrix} \mathbf{x}'_{1i} \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i \\ \vdots \\ \mathbf{x}'_{ki} \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i \end{pmatrix} = 0,$$

under the strictly exogenous assumption of \mathbf{x}_{it} , where $\mathbf{x}_{ji} = (x_{j,i1}, x_{j,i2}, \dots, x_{j,iT})'$ for $j = 1, \dots, k$. Consequently, we have

$$E(\mathbf{Z}'_i \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_i) = \begin{pmatrix} O(1) \\ \mathbf{0} \end{pmatrix},$$

which in turn yields

$$\sqrt{NT} \left(\hat{\delta}_{MLE}^{TS} - \boldsymbol{\delta} \right) = O_p \left(\sqrt{\frac{N}{T}} \right),$$

i.e., the QDMLE of $\boldsymbol{\delta}$ is asymptotically biased of order $\sqrt{\frac{N}{T}}$, which is the same order as in the model without exogenous variables.

Alternatively, (OA.1) can be stacked in vector form as

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\delta} + \Lambda \mathbf{f}_t + \mathbf{u}_t, \quad t=1, \dots, T, \quad (\text{OA.8})$$

where $\mathbf{Z}_t = (\mathbf{z}_{1t}, \dots, \mathbf{z}_{Nt})'$, \mathbf{y}_t , \mathbf{u}_t , and Λ are defined in (A.27).

Using the same normalization (A.29), if we multiply $\mathbf{G}(\psi)'$ in both sides of (OA.8), then we have

$$\mathbf{G}(\psi)' \mathbf{y}_t = \mathbf{G}(\psi)' \mathbf{Z}_t \delta + \mathbf{G}(\psi)' \mathbf{u}_t, \quad t=1, \dots, T, \quad (\text{OA.9})$$

where $\mathbf{G}(\psi)' \mathbf{y}_t$ and $\mathbf{G}(\psi)' \mathbf{u}_t$ are defined in (A.30), and

$$\mathbf{G}(\psi)' \mathbf{Z}_t = \begin{pmatrix} \mathbf{g}_1(\psi_1)' \mathbf{Z}_t \\ \mathbf{g}_2(\psi_2)' \mathbf{Z}_t \\ \vdots \\ \mathbf{g}_{N-r}(\psi_{N-r})' \mathbf{Z}_t \end{pmatrix} = \mathbf{Z}_{\underline{r+1,N},t} + \Psi \mathbf{Z}_{\underline{1,r},t},$$

where $\mathbf{Z}_{\underline{1,r},t} = (\mathbf{z}_{1t}, \dots, \mathbf{z}_{rt})'$ and $\mathbf{Z}_{\underline{r+1,N},t} = (\mathbf{z}_{r+1,t}, \dots, \mathbf{z}_{Nt})'$.

For the transformed model (OA.9), the variance-covariance matrix is given by (A.32), and the quasi-log-likelihood function of (OA.9) is given by

$$\log L(\gamma, \psi) = -\frac{N}{2} \log |\Omega_{CS}| - \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{Z}_t \delta)' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) (\mathbf{y}_t - \mathbf{Z}_t \delta). \quad (\text{OA.10})$$

Conditional on ψ , the FOC for γ is

$$\frac{\partial \log L(\delta, \psi)}{\partial \delta} = \sum_{t=1}^T \mathbf{Z}_t' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) (\mathbf{y}_t - \mathbf{Z}_t \delta) = 0, \quad (\text{OA.11})$$

which yields

$$\hat{\delta}_{MLE}^{CS} = \left(\sum_{t=1}^T \mathbf{Z}_t' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{Z}_t \right)^{-1} \sum_{t=1}^T (\mathbf{Z}_t' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{y}_t), \quad (\text{OA.12})$$

where CS refers to using normalization (A.29) to remove the multiple interactive effects.

For this QDMLE (OA.12), we have

$$\begin{aligned} \sqrt{NT} (\hat{\delta}_{MLE}^{CS} - \delta) &= \left(\frac{1}{NT} \sum_{t=1}^T \mathbf{Z}_t' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{Z}_t \right)^{-1} \\ &\quad \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^T \sum_{i=1}^T \mathbf{Z}_t' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{u}_i \right). \end{aligned} \quad (\text{OA.13})$$

For the numerator of (OA.13), we can observe that

$$\mathbf{Z}_t' \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{u}_t = \begin{pmatrix} \mathbf{y}'_{t-1} \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_t \\ \mathbf{X}'_t \mathbf{H}(\vartheta)' \Omega_{TS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_t \end{pmatrix},$$

and thus

$$E\left[\mathbf{Z}'_t \mathbf{G}(\psi)' \Omega_{CS}^{-1} \mathbf{G}(\psi) \mathbf{u}_t\right] = E\left(\begin{array}{l} \mathbf{y}'_{t-1} \mathbf{H}(\vartheta)' \Omega_{CS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_t \\ \mathbf{X}'_t \mathbf{H}(\vartheta)' \Omega_{CS}^{-1} \mathbf{H}(\vartheta) \mathbf{u}_t \end{array}\right) = 0,$$

the last identity holds since

$$E(\mathbf{u}_t \mathbf{y}'_{t-1}) = 0 \text{ and } E(\mathbf{u}_t \mathbf{X}'_t) = 0.$$

by using the fact that \mathbf{u}_t is independent of \mathbf{y}_{t-1} and \mathbf{X}_t . As a result, we can claim that the QDMLE (OA.12) is asymptotically unbiased in the sense that

$$E\left[\sqrt{NT}(\hat{\boldsymbol{\delta}}_{MLE}^{CS} - \boldsymbol{\delta})\right] = 0.$$

which suggests that the QDMLE (A.36) is asymptotically unbiased, and the asymptotical unbiasedness of QDMLE is independent of the way how (N, T) go to infinity.

CHAPTER 14

INFORMATIONAL CONTENT OF FACTOR STRUCTURES IN SIMULTANEOUS BINARY RESPONSE MODELS

Shakeeb Khan^a, Arnaud Maurel^b and Yichong Zhang^c

^aBoston College, United States

^bDuke University, United States

^cSingapore Management University, Singapore

ABSTRACT

We study the informational content of factor structures in discrete triangular systems. Factor structures have been employed in a variety of settings in cross-sectional and panel data models, and in this chapter we formally quantify their identifying power in a bivariate system often employed in the treatment effects literature. Our main findings are that imposing a factor structure yields point-identification of parameters of interest, such as the coefficient associated with the endogenous regressor in the outcome equation, under weaker assumptions than usually required in these models. In particular, we show that a “non-standard” exclusion restriction that requires an explanatory variable in the outcome equation to be excluded from the treatment equation is no longer necessary for identification, even in cases where all of the regressors from the outcome equation are discrete. We also establish identification of the coefficient of the endogenous regressor in models with more general factor structures, in situations where one has access to at least two continuous measurements of the common factor.

Keywords: Factor structures; discrete choice; causal effects; identification; binary dependent variable; endogeneity

JEL Classification: C14; C31; C35

1. INTRODUCTION

Factor models see widespread and increasing use in various areas of econometrics. This type of structure has been employed in a variety of settings in cross-sectional, panel and time series models, and has proven to be a flexible way to model the behavior of and relationship between unobserved components of econometric models. The basic idea behind factor models is to assume that the dependence across the unobservables is generated by a low-dimensional set of mutually independent random factors. The applied and theoretical research employing factor structures in econometrics is extensive. In particular, these models are often used in the treatment effect literature as a way to identify the joint distribution of potential outcomes from the marginal distributions, and then recover the distribution of treatment effects from this joint distribution.¹ Factor models have been used in a number of different contexts in applied microeconomics. These include, among others, earnings dynamics (Abowd & Card, 1989; Bonhomme & Robin, 2010), estimation of returns to schooling and work experiences (Ashworth et al., 2021), as well as cognitive and non-cognitive skill production technology (Cunha et al., 2010). Heckman and Vytlacil (2007a, 2007b) provide various additional references. All of these papers, with the notable exception of Cunha et al. (2010), rely on linear factor models where the unobservables are assumed to be written as the sum of a linear combination of mutually independent factors and an idiosyncratic shock.

In this chapter, we bring together the literature on factor models with the literature on the identification and estimation of binary response models (Blundell & Powell, 2004; Klein & Spady, 1993; Lewbel, 2000; Park & Phillips, 2000), in particular triangular binary choice models (Chesher, 2005; Han & Vytlacil, 2017; Shaikh & Vytlacil, 2011; Vytlacil & Yildiz, 2007), by exploring the *informational content* of factor structures in this class of models.² Focusing on this class can be well motivated from both an empirical and theoretical perspective. From the former, many treatment effect models fit into this framework as treatment is typically a binary and endogenous variable in the system, whose effect on outcomes is often a parameter the econometrician wishes to conduct inference on. From a theoretical perspective, inference on this type of system can be complicated, if not impossible without strong parametric assumptions, which may not be reflected in the observed data. Imposing no restriction on the structure of endogeneity often fails to achieve identification of parameter, or at best only do so in sparse regions of the data, thus making inference impractical in practice. In this context, modeling the endogeneity between the selection and the outcome by a factor structure may be a useful “in-between” setting, which, at the very least, can be used to gauge the sensitivity of the parametric approach to their stringent assumptions.

We start our analysis by imposing a particular factor structure to the two unobservables in our system of binary equations described in further detail in the next section, and explore the informational content of this assumption. We assume that the unobservables from the treatment equation (V) and the outcome equation (U) are related through the following factor model:

$$U = \gamma_0 V + \Pi \quad (1.1)$$

where Π is an unobserved random variable assumed to be distributed independently of V and γ_0 is a scalar parameter. This structure generalizes the canonical case where the unobservables (U, V) are jointly normally distributed, for which this relationship always holds. Our main finding is that there is indeed informational content of factor structures in the sense that, in contrast to prior literature – notably [Vytlačil and Yıldız \(2007\)](#) – one no longer requires an additional “non-standard” exclusion restriction, nor the strong support conditions on the covariates entering the outcome equation that are generally needed for identification in these models. Our identification results are constructive and translate directly into a rank-based estimator of the coefficient associated with the binary endogenous variable, which we provide and study in a supplement to this chapter.

While an appealing feature of the structure considered in [Equation \(1.1\)](#) is that it is a natural extension of the bivariate Probit specification that has often been considered in the literature, this model does impose significant restrictions on the nature of the dependence between the unobservables U and V . In the chapter, we extend this baseline specification by considering a linear factor structure of the form:

$$U = \gamma_0 W + \eta_1 \quad (1.2)$$

$$V = W + \eta_2 \quad (1.3)$$

where (W, η_1, η_2) are mutually independent unobserved random variables. We study the informational content of this extended factor structure in the context of triangular binary choice models and establish identification, assuming access to at least two continuous noisy measurements of the unobserved factor W . This setup has been used in a number of applications, in particular in labor economics. In these applications, the unobserved factor is typically interpreted as latent individual ability, about which several continuous noisy measurements are available from the data. This is the case of, for instance, [Carneiro et al. \(2003\)](#), [Cunha et al. \(2010\)](#), [Heckman et al. \(2018\)](#) and [Ashworth et al. \(2021\)](#), who use components of the Armed Services Vocational Aptitude Battery test as measurements of cognitive ability.

The rest of the chapter is organized as follows. In Section 2, we formally describe the triangular system with our factor structure, and discuss our main identification results for the parameters of interest in this model. Section 3 explores identification in more general factor structure models which involve

multiple idiosyncratic errors, in a context where one has access to two continuous noisy measurements of the common unobserved factor. Section 4 concludes. We prove Theorems 2.1 and 3.1 in Sections A and B, respectively. In Section C, we establish the sharp identified set of α_0 when the support condition for point-identification is violated in the one-factor model and the necessary and sufficient condition for point-identification in the two-factor model with two continuous measurements of the common factor. The corresponding results are proved in Sections D and E. The Supplementary Material studies the asymptotic properties of a rank-based estimator for α_0 and explores its finite sample properties through some Monte Carlo simulation exercises.

Notation: throughout the chapter we write $\mathbf{1}\{A\}$ to denote the usual indicator function that takes value 1 if event A happens, and 0 otherwise. We also denote by $d(U)$ and $d(U|V)$ the lengths of the support of random variable U , and the conditional support of U given V , respectively.

2. TRIANGULAR BINARY MODEL WITH FACTOR STRUCTURE

2.1. Set-up and Main Identification Result

In this section, we consider the identification of the following triangular binary model:

$$Y_1 = \mathbf{1}\{Z_1'\lambda_0 + Z_3'\beta_0 + \alpha_0 Y_2 - U > 0\} \quad (2.4)$$

$$Y_2 = \mathbf{1}\{Z'\delta_0 - V > 0\} \quad (2.5)$$

where $Z \equiv (Z_1, Z_2)$ and (U, V) is a pair of random shocks. Z_2 and Z_3 provide the exclusion restrictions in the model, and the distribution of (Z_2, Z_3) is required to be nondegenerate conditional on $Z_1'\lambda_0 + Z_3'\beta_0$. We further assume that the error terms U and V are jointly independent of (Z_1, Z_2, Z_3) . The endogeneity of Y_2 in (2.4) arises when U and V are not independent.

The above model, or minor variations of it, have often been considered in the recent literature. See for example, [Vytlacil and Yildiz \(2007\)](#), [Abrevaya et al. \(2010\)](#), [Klein et al. \(2015\)](#), [Vuong and Xu \(2017\)](#), [Khan and Nekipelov \(2018\)](#) and references therein. A key parameter of interest³ in our chapter as is in much of the literature is α_0 . In this chapter, we provide conditions under which the parameters of interest are point-identified. As such, our analysis complements alternative partial-identification approaches that have been proposed in the context of triangular binary models. See, in particular, [Chiburis \(2010\)](#), [Shaikh and Vytlacil \(2011\)](#), and [Mourifié \(2015\)](#).⁴ As discussed in the aforementioned papers, the parameter α_0 is difficult, if not impossible to point identify and estimate without imposing parametric restrictions on the unobserved variables in the model, (U, V) .

The difficulty of identifying α_0 in semi-parametric “distribution-free” models, and the sensitivity of its identification to misspecification in parametric models is what motivates the factor structure we add in this chapter to the above model. Specifically, to allow for endogeneity in the form of possible non-zero correlation between U and V , we augment the model with the following equation:

$$U = \gamma_0 V + \Pi \quad (2.6)$$

where Π is an unobserved random variable, assumed to be distributed independently of (V, Z_1, Z_2, Z_3) , and γ_0 is an additional unknown scalar parameter. Importantly, this type of factor structure always holds when the residuals of both equations are jointly normally distributed. Furthermore, this specification corresponds to the type of structure used in Independent Component Analysis (ICA), where V and Π are two mutually independent factors. This method has found many applications in various fields, including signal processing and image extraction; applications in economics include, e.g., [Hyvärinen and Oja \(2000\)](#), [Moneta et al. \(2013\)](#) and [Gourieroux et al. \(2017\)](#). While, in contrast to the ICA literature, the factors and the factor loadings are not the main objects of interest in our analysis, this dimension-reducing structure plays a key role in our identification results.

Our aim is to first explore identification of the parameters $(\alpha_0, \delta_0, \gamma_0, \beta_0, \lambda_0)$ under standard nonparametric regularity conditions on (V, Π) . Note that the parameter δ_0 in the selection equation can be identified up to scale in various ways. See, for example, [Klein and Spady \(1993\)](#) and [Han \(1987\)](#), among others. We then impose the usual condition that one of δ_0 's coordinates is equal to one to fix the scale. For simplicity, for the rest of the chapter, we denote $X \equiv Z' \delta_0$ and assume X is observed. We further define $X_1 \equiv Z_1' \lambda_0 + Z_3' \beta_0$. However, we cannot identify λ_0 and β_0 beforehand. We propose instead to identify them along with α_0 .

Our main identification result is based on the Assumptions **A1–A4** we state below:

- A1** The first coefficient of λ_0 is normalized to one so that $\lambda_0 = (1, \lambda_{0,-1}^T)^T$. The parameter $\theta_0 \equiv (\alpha_0, \gamma_0, \lambda_{0,-1}, \beta_0)$ is an element of a compact subset of $\Re^{d_1+d_3+1}$, where d_1 and d_3 are the dimensions of Z_1 and Z_3 , respectively.
- A2** The vector of unobserved variables, (U, V, Π) is continuously distributed with support on a subset of \Re^3 and independently distributed of the vector (Z_1, Z_2, Z_3) . Furthermore, we assume that the unobserved random variables Π, V are distributed independently of each other.
- A3** X is continuously distributed with absolute continuous density w.r.t. Lebesgue measure. Its density is bounded and bounded away from zero on any compact subset of its support.
- A4** Let $Z_{1,-1}$ be all the coordinates of Z_1 except the first one, and $d = d_1 + d_3 + 1$. There exist $2d$ vectors $\{z_1^{(l)}, z_3^{(l)}, x^{(l)}\}_{l=1}^d$ and $\{\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}, \tilde{x}^{(l)}\}_{l=1}^d$ in the joint support of (Z_1, Z_3, X) such that

$$\alpha_0 + (z_{1,-1}^{(l)} - \tilde{z}_{1,-1}^{(l)})' \lambda_{0,-1} + (z_3^{(l)} - \tilde{z}_3^{(l)})' \beta_0 - \gamma_0 (x^{(l)} - \tilde{x}^{(l)}) = \tilde{z}_{1,1}^{(l)} - z_{1,1}^{(l)}, \quad l = 1, \dots, d$$

and $\text{rank}(\mathcal{M}) = d$, where

$$\mathcal{M} = \begin{pmatrix} 1 & \dots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \dots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \dots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ x^{(1)} - \tilde{x}^{(1)} & \dots & x^{(d)} - \tilde{x}^{(d)} \end{pmatrix}.$$

Before turning to our main identification result, a couple of remarks are in order.

Remark 2.1. *The first part of Assumption A1 is a standard scale normalization. Assumption A2 is also standard in this literature. The assumption that the instruments are independent of the unobservables can also be found in, among others, Abrevaya et al. (2010), Vytlacil and Yildiz (2007), Klein et al. (2015), and Khan and Nekipelov (2018). The assumption of independence between Π and V is also made in Bai and Ng (2002) and Carneiro et al. (2003).*

Remark 2.2. *Assumptions A3 and A4 impose some restrictions on the distributions of the covariates entering the selection and outcome equations, respectively. Specifically, Assumption A3 requires one component of the covariates Z entering the selection equation to be continuously distributed, which is often required in models with discrete outcomes. In contrast, Assumption A4 only requires some variation of (Z_1, Z_3) . In particular, the distribution of (Z_1, Z_3) cannot be degenerate but is allowed to be discrete. This assumption can be interpreted as a full rank condition, which ensures that the system of linear equations that delivers point-identification has a unique solution.*

We now turn to our main identification result, Theorem 2.1, which concludes that under our stated conditions and our factor structure we can attain point-identification of the vector of parameters θ_0 .

Theorem 2.1. *Under Assumptions A1-A4, θ_0 is point identified.*

An important takeaway from this result, which we discuss further in Section 2.2 below, is that imposing the factor structure (2.6) yields point-identification under weaker support conditions when compared to the existing literature, and does not require a second exclusion restriction either. In particular, our model delivers point-identification of the parameters of interest even in situations where all of the regressors from the outcome equation are discrete. This indicates that, from the selection equation combined with the factor structure that we impose here, we can overturn the non-identification result of Bierens and Hartog (1988) which would apply to the outcome equation alone.

The proof of Theorem 2.1, which is reported in Section A in the Supplementary Appendix, relies on the fact that, for two observations (Z_1, Z_3, X) and $(\tilde{Z}_1, \tilde{Z}_3, \tilde{X})$,

$$\begin{aligned} \partial_x P^{11}(Z_1, Z_3, X) / f_V(X) + \partial_x P^{10}(\tilde{Z}_1, \tilde{Z}_3, \tilde{X}) / f_V(\tilde{X}) &= 0 \\ \Leftrightarrow \alpha_0 + (Z_1 - \tilde{Z}_1)' \lambda_0 + (Z_3 - \tilde{Z}_3)' \beta_0 - \gamma_0 (X - \tilde{X}) &= 0, \end{aligned} \tag{2.7}$$

where $f_V(\cdot)$ is the pdf of V , which is identified over the support of X , and $P^{ij}(z_1, z_3, x) \equiv \text{Prob}(Y_1 = i, Y_2 = j | Z_1 = z_1, Z_3 = z_3, X = x)(\partial_x P^{ij}(z_1, z_3, x))$ denote the choice probability (partial derivative of the ij -choice probability with respect to the third argument), which are both identified from the data.

Remark 1. This identification result can be extended to the case of a separable non-parametric factor model. Namely, consider the following relationship between unobserved components:

$$U = g_0(V) + \tilde{\Pi} \quad (2.8)$$

where $\tilde{\Pi}$ is an unobserved random variable assumed to be distributed independently of V and all instruments. $g_0(\cdot)$ is an unknown function assumed to satisfy standard smoothness conditions. The parameter of interest is $(\alpha_0, \lambda_0, \beta_0)$, but now the unknown nuisance parameter in the factor equation is infinite dimensional. By replacing $\gamma_0 X$ by $g_0(X)$ in (2.7), we have

$$\begin{aligned} \partial_x P^{11}(Z_1, Z_3, X) / f_V(X) + \partial_x P^{10}(\tilde{Z}_1, \tilde{Z}_3, \tilde{X}) / f_V(\tilde{X}) &= 0 \\ \Leftrightarrow \alpha_0 + (Z_1 - \tilde{Z}_1)' \lambda_0 + (Z_3 - \tilde{Z}_3)' \beta_0 - (g_0(X) - g_0(\tilde{X})) &= 0. \end{aligned} \quad (2.9)$$

One can then establish identification after modifying the rank condition A4 by replacing $\gamma_0(x^{(l)} - \tilde{x}^{(l)})$ by $g_0(x^{(l)}) - g_0(\tilde{x}^{(l)})$.

Remark 2. We assume rank invariance in (2.6). It is possible to relax such condition to rank similarity.⁵ Specifically, we can consider the following model:

$$\begin{aligned} Y_1 &= \mathbf{1}\{Z_1' \lambda_0 + Z_3' \beta_0 + \alpha_0 Y_2 - U(Y_2) > 0\} \\ Y_2 &= \mathbf{1}\{Z' \delta_0 - V > 0\}, \end{aligned}$$

where

$$U(y_2) = \gamma_0 V + \Pi(y_2), \text{ for } y_2 = 0, 1.$$

We further assume $(V, \Pi(1), \Pi(0))$ is continuously distributed with support on a subset of \Re^3 and independently distributed of the vector Z_1, Z_2, Z_3 , V and $(\Pi(1), \Pi(0))$ are independent, $P(\Pi(1) \leq \pi) = P(\Pi(0) \leq \pi)$ for $\pi \in \Re$, and Assumptions A1, A3, and A4 hold. Then, we can identify θ_0 by a similar argument as the proof of Theorem 2.1.

2.2. Connection with Prior Literature

We now discuss in detail how our setup and main identification result relates to the existing literature.

In a related work, [Han and Vytlacil \(2017\)](#) consider the identification of a generalized bivariate Probit model.⁶ Our linear factor structure and the one-parameter copula model considered in [Han and Vytlacil \(2017\)](#) are not nested by

each other. First, note that based on the factor structure, we can recover F_{Π} , the distribution of Π , as a function of (F_U, F_V, γ_0) by deconvolution. We can then write the copula of (U, V) as

$$\begin{aligned} F_{U,V}(F_U^{-1}(u), F_V^{-1}(v)) &= \int_{-\infty}^{F_U^{-1}(v)} F_{\Pi}(F_U^{-1}(u) - \gamma_0 w; F_U, F_V, \gamma_0) f_V(w) dw \\ &= C(u, v; F_U, F_V, \gamma_0). \end{aligned}$$

The copula depends not only on γ_0 but also on two infinite dimensional parameters (F_U, F_V) . Thus, unlike Han and Vytlacil (2017), our factor structure cannot be characterized by a one-parameter copula. In addition, in order to achieve identification, Han and Vytlacil (2017) first nonparametrically identify the two marginals by assuming the existence of a full support regressor that is common to both equations.⁷ In contrast, our approach does not rely on the existence of such a regressor. Under the factor structure assumed in our analysis, we bypass the nonparametric identification of the marginals as a whole and directly consider the identification of the structural parameters. It follows that our model cannot be nested by the one-parameter copula model considered by Han and Vytlacil (2017). On the other hand, there exist one-parameter copula models that cannot be decomposed into linear factor structures.⁸ This implies that our model does not nest Han and Vytlacil (2017) either.

Our analysis also relates to Vytlacil and Yildiz (2007) and Vuong and Xu (2017), who consider the identification of α_0 in a triangular binary model. Our identification result, however, differs from theirs in important ways. Namely, denote $X = Z'\delta_0 = Z_1'\delta_{1,0} + Z_2'\delta_{2,0}$. Then, Assumption A4 implies that we can find a pair of observations (z_1, z_2, z_3) and $(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$ such that

$$z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 - \gamma_0(z_1'\delta_{1,0} + z_2'\delta_{2,0}) = \tilde{z}_1'\lambda_0 + \tilde{z}_3'\beta_0 - \gamma_0(\tilde{z}_1'\delta_{1,0} + \tilde{z}_2'\delta_{2,0}). \quad (2.10)$$

In contrast, using our notation, Vytlacil and Yildiz (2007) require that one can find a pair of observations (z_1, z_2, z_3) and $(\tilde{z}_1, \tilde{z}_2, \tilde{z}_3)$ such that $z'\delta_0 = \tilde{z}'\delta_0$ and

$$z_1'\lambda_0 + z_3'\beta_0 + \alpha_0 = \tilde{z}_1'\lambda_0 + \tilde{z}_3'\beta_0. \quad (2.11)$$

Vuong and Xu (2017) do not assume the existence of Z_3 . In our binary outcome setup, the functions $h(0, x, \tau)$ and $h(1, x, \tau)$ defined in Vuong and Xu (2017) are equal to $1\{x + F_U^{-1}(\tau) \geq 0\}$ and $1\{x + \alpha + F_U^{-1}(\tau) \geq 0\}$, respectively, where $x = z_1'\lambda_0$ and F_U is the CDF of $-U$. Then, Vuong and Xu (2017), Assumption C(ii) requires that we can find z_1 and \tilde{z}_1 in the support of Z_1 so that for any τ_1, τ_2 , if $1\{\tilde{z}_1'\lambda_0 + F_U^{-1}(\tau_1) \geq 0\} = 1\{\tilde{z}_1'\lambda_0 + F_U^{-1}(\tau_2) \geq 0\}$, then $1\{z_1'\lambda_0 + \alpha_0 + F_U^{-1}(\tau_1) \geq 0\} = 1\{z_1'\lambda_0 + \alpha_0 + F_U^{-1}(\tau_2) \geq 0\}$. Provided that the support of U nests the supports of Z_1 and $Z_1'\lambda_0 + \alpha_0$, Vuong and Xu (2017), Assumption C(ii) is then equivalent to:⁹

$$z_1'\lambda_0 + \alpha_0 = \tilde{z}_1'\lambda_0. \quad (2.12)$$

Several remarks are in order. First, note that sufficient support conditions for the restrictions (2.10)–(2.12) are $d(Z_1'\lambda_0 + Z_3'\beta_0 - Z'\delta_0\gamma_0) \geq |\alpha_0|$, $d(Z_1'\lambda_0 + Z_3'\beta_0 | Z'\delta_0) \geq |\alpha_0|$, and $d(Z_1'\lambda_0 | Z'\delta_0) \geq |\alpha_0|$ with a positive probability, respectively, where $d(\cdot)$ denotes the “length” of its argument. These three support conditions are such that

$$d(Z_1'\lambda_0 + Z_3'\beta_0 - Z'\delta_0\gamma_0) \geq d(Z_1'\lambda_0 + Z_3'\beta_0 | Z'\delta_0) \geq d(Z_1'\lambda_0 | Z'\delta_0),$$

where the first and second inequalities are strict if Z_2 and Z_3 have at least one continuous component, respectively. Importantly, we show in Section C of the Supplement that for a version of the triangular binary model with univariate Z_2 and Z_3 and no common regressor Z_1 , the support condition $d(Z_1'\lambda_0 + Z_3'\beta_0 | Z'\delta_0) \geq |\alpha_0|$ is actually also necessary to the identification of the model without factor structure. This implies that by imposing our factor structure, one can identify values of α_0 in a region that cannot be identified in the model considered by [Vytlačil and Yıldız \(2007\)](#). Such region is characterized in Section C of the Supplement.

Second, it directly follows from these support conditions that, in the presence of a factor model and in contrast to both [Vytlačil and Yıldız \(2007\)](#) and [Vuong and Xu \(2017\)](#), variation in Z_2 helps in the identification of α_0 . In that sense, the factor model allows to restore the intuition from standard IV approaches in linear models that variation in the instrument Z_2 is critical to the identification of the parameters of the outcome equation. Related to this, the support of Z_2 plays an important role in our identification analysis. In particular, if Z_2 is discrete, our identification strategy requires sufficient variation in the variables in the outcome equation, namely Z_1 and Z_3 . In this case, our support requirement is equivalent to that assumed by [Vytlačil and Yıldız \(2007\)](#).

Third, another important aspect of Assumption A4 is that it does not impose any constraint on the variables from the outcome equation. Specifically, consider a case where the outcome equation does not contain a variable that is excluded from the selection equation (i.e., $\beta_0 = 0$), the regressor that is common to both equations, Z_1 , is scalar and binary, and where $\lambda_0 = 1$. In this case, one can show that the identifying support conditions associated with [Vytlačil and Yıldız \(2007\)](#) (2.11) and [Vuong and Xu \(2017\)](#) (2.12) generally fail to hold, except for a finite set of values $\alpha_0 \in \{-1, 0, 1\}$. In contrast, our support restriction (2.10) holds under more general conditions: without any restriction on α_0 if one element of Z_2 is continuous with large support, and on a continuum of possible values for α_0 if one element of Z_2 is continuous with bounded support. In that sense, the factor structure replaces the need for a continuous component in (Z_1, Z_3) in the outcome equation.

Finally, at a high level, our identification strategy shares similarities with the Local Instrumental Variable (LIV) approach that has been proposed by [Heckman and Vytlačil \(2005\)](#) and further discussed by [Carneiro and Lee \(2009\)](#). In particular, our identifying restriction (2.7) can be alternatively derived from a local IV strategy applied to a potential outcomes model characterized by $Y_1(y_2) = \mathbf{1}\{Z_1'\lambda_0 + Z_3'\beta_0 + \alpha_0 y_2 - U > 0\}$, with treatment given by

$Y_2 = \mathbf{1}\{Z'\delta_0 - V > 0\}$. In contrast to the LIV literature though, we focus in our analysis on the structural parameter α_0 rather than on the marginal treatment effects. Our identification result shows that, by leveraging the identifying power of the factor structure, one can identify α_0 under weaker support restrictions than in the prior literature. In particular, our strategy makes it possible to use variation in $X = Z'\delta_0$ to identify α_0 , even when all the components of Z_1 and Z_3 are discrete.¹⁰

3. EXTENDED FACTOR STRUCTURE IN THE PRESENCE OF CONTINUOUS MEASUREMENTS

Up until now we have proposed identification and estimation results for a triangular system with a particular factor structure. A disadvantage of this structure is that it only includes one idiosyncratic shock (Π). We consider below an extension that addresses this limitation.

Namely, we consider the following model:

$$\begin{aligned} Y_1 &= \mathbf{1}\{X_1 + \alpha_0 Y_2 - U \geq 0\} \\ Y_2 &= \mathbf{1}\{X - V \geq 0\}, \end{aligned} \tag{3.1}$$

where $X_1 = Z_1'\lambda_0 + Z_3'\beta_0$, $X = Z'\delta_0$, $U = \gamma_0 W + \eta_1$, $V = W + \eta_2$, and (W, η_1, η_2) are mutually independent. In this setup, W can be interpreted as an unobserved confounder that satisfies the matching-on-unobservables condition $(Y_1(0), Y_1(1)) \perp\!\!\!\perp Y_2 | W, X, X_1$ (Abbring & Heckman, 2007). Recall that, following the arguments in Section 2.1 above, we assume that X is observed. In addition, we assume two auxiliary continuous measurements

$$\begin{aligned} Y_3 &= \nu_0 W + \eta_3 \\ Y_4 &= \sigma_0 W + \eta_4, \end{aligned} \tag{3.2}$$

where $(W, \eta_1, \eta_2, \eta_3, \eta_4)$ are mutually independent, and $\nu_0 \neq 0$.¹¹

Our identification result is based on the following assumptions:

- B0** The first coefficient of λ_0 is normalized to one so that $\lambda_0 = (1, \lambda_{0,-1}^T)^T$. The parameter $\theta_0 \equiv (\alpha_0, \gamma_0, \lambda_{0,-1}, \beta_0, \nu_0, \sigma_0)$ is an element of a compact subset of $\Re^{d_1+d_3+3}$, where d_1 and d_3 are the dimensions of Z_1 and Z_3 , respectively. The vector of unobservables in the outcome and selection equations $(W, \eta_1, \eta_2, \eta_3)$ are independently distributed of the vector (Z_1, Z_2, Z_3) . Both η_1 and η_2 are continuously distributed.
- B1** $\gamma_0 \neq 0$. X is continuously distributed with absolute continuous density w.r.t. Lebesgue measure over the whole real line, conditionally on Z_1 and Z_3 . The unconditional density of X is bounded and bounded away from zero on any compact subset of its support.
- B2** W is not normally distributed or both η_3 and η_4 do not have a Gaussian component.

- B3** $E(\eta_3) = E(\eta_4) = 0$, $E(|\eta_3|) < \infty$, and $E(|\eta_4|) < \infty$.
- B4** $E(\exp(i\zeta\eta_2))$, $E(\exp(i\zeta\eta_3))$, and $E(\exp(i\zeta\eta_4))$ do not vanish for any $\zeta \in \mathfrak{R}$, where $i = \sqrt{-1}$.
- B5** $E(\exp(i\zeta W)) \neq 0$ for all ζ in a dense subset of \mathfrak{R} .
- B6** The distributions of W , η_2 , and η_3 admit uniformly bounded densities $f_W(\cdot)$, $f_{\eta_2}(\cdot)$, and $f_{\eta_3}(\cdot)$ with respect to the Lebesgue measure that are supported on an interval (which may be infinite), respectively.
- B7** Let $Z_{1,-1}$ be all the coordinates of Z_1 except the first one, and $d = d_1 + d_3 + 1$. There exist $2d$ vectors $\{z_1^{(l)}, z_3^{(l)}\}_{l=1}^d$ and $\{\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}\}_{l=1}^d$ in the joint support of (Z_1, Z_3) and $\{w^{(l)}\}_{l=1}^d, \{\tilde{w}^{(l)}\}_{l=1}^d$ such that

$$\alpha_0 + (z_{1,-1}^{(l)} - \tilde{z}_{1,-1}^{(l)})' \lambda_{0,-1} + (z_3^{(l)} - \tilde{z}_3^{(l)})' \beta_0 - \gamma_0 (w^{(l)} - \tilde{w}^{(l)}) = \tilde{z}_{1,l}^{(l)} - z_{1,l}^{(l)}, \quad l = 1, \dots, d$$

and $\text{rank}(\mathcal{M}) = d$, where

$$\mathcal{M} = \begin{pmatrix} 1 & \cdots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \cdots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \cdots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ w^{(1)} - \tilde{w}^{(1)} & \cdots & w^{(d)} - \tilde{w}^{(d)} \end{pmatrix}.$$

We now discuss these assumptions, before turning to the identification result. First, Assumption B0 is similar to Assumptions A1 and A2. We only need one of the idiosyncratic errors in the continuous measurements to be independent of the covariates because the other one is used to identify the distribution of the common factor W only. Second, as we assume in Assumption B1 that $\gamma_0 \neq 0$ and X has full support, the support condition

$$d(Z_1' \lambda_0 + Z_3' \beta_0 - \gamma_0 X) \geq |\alpha_0|.$$

holds automatically. The full support condition of X is necessary to identify the density of V , which is further used to identify the distribution of η_2 . Assumption B1 reinforces this condition by supposing that X has full support conditional on Z_1 and Z_3 , which is needed to identify the parameters from the outcome equation in a second step. Since $X = Z'\delta_0$ with $Z = (Z_1, Z_2)$, this is in turn equivalent to Z_2 having full support conditional on Z_1 and Z_3 . Third, Assumptions B2–B6 imply Assumptions 1 to 4 in [Hu and Schennach \(2013\)](#). In practice, we add the condition that the characteristic function of η_2 does not vanish, which is used for the deconvolution arguments in the proof of Theorem 3.1. We refer the reader to [Hu and Schennach \(2013\)](#) for more discussions of these assumptions.¹²

Theorem 3.1. *If (3.1)–(3.2) and Assumptions B0–B7 hold, then θ_0 are identified.*

The proof of Theorem 3.1 can be found in Section B of the Supplement. Several remarks are in order. First, while we allow for a more general factor structure on the unobservables U and V , we also depart from our baseline specification by supposing that we have access to two continuous noisy measurements

of the common factor W . This is a standard requirement in the nonparametric measurement error literature (Hu & Schennach, 2008). Besides, assuming access to a set of (selection-free) noisy measurements of the unobserved factors is also very standard in the evaluation literature. See, among many others, Carneiro et al. (2003), Heckman and Navarro (2007), Heckman and Vytlacil (2007a), and Cunha et al. (2010).

For instance, in applications in labor economics, the unobserved factor W often captures individual ability. This would apply, for example, to the evaluation of the effect of employment while in college (Y_2) on college graduation (Y_1). In this example, natural candidates for Z_2 are local labor market variables, including average wages and unemployment rate, while candidates for Z_1 include, among others, eligibility to financial aid programs providing tuition subsidy to students who maintain a minimum level of academic achievement.¹³ In this context, cognitive skill measurements, such as the ASVAB test components that are available in the NLSY79 and NLSY97 surveys, are natural and often used candidates for the continuous measurements (Y_3 , Y_4) (Ashworth et al., 2021).

Second, as is clear from the proof of Theorem 3.1, the key purpose of the continuous measurements is to identify the distribution of the common factor W . While we assume in this section that the measurement equations are linear, it is possible to identify θ_0 with a more general nonlinear system of continuous measurements, provided that the researcher has access to at least three such measurements. One can then combine Theorem 2 in Cunha et al. (2010) (Section 3.3, pp. 894–895), that yields identification of the distribution of W , with the proof of Theorem 3.1 in order to show identification of θ_0 for the case of nonlinear auxiliary measurements. Assuming access to a set of at least three measurements also makes it possible to relax the non-normality requirement imposed in Assumption B2.

Third, under the previous set of assumptions, the average treatment effect (ATE) is also identified. Key to this identification result is the full support condition on X given Z_1 and Z_3 (Assumption B1). Note that the conditional ATE given $X_1 = x_1$ is equal to $F_U(x_1 + \alpha_0) - F_U(x_1)$. In addition,

$$P(Y_1 = 1, Y_2 = 1 | X_1 = x_1, X = x) = F_{U,V}(x_1 + \alpha_0, x).$$

One can let $x \rightarrow \infty$ so that

$$\lim_{x \rightarrow \infty} P(Y_1 = 1, Y_2 = 1 | X_1 = x_1, X = x) = F_U(x_1 + \alpha_0).$$

Similarly,

$$\lim_{x \rightarrow -\infty} P(Y_1 = 1, Y_2 = 0 | X_1 = x_1, X = x) = F_U(x_1).$$

This identifies the conditional and unconditional ATE.

Fourth, similar to the earlier discussions in Remark 2.2 and Section 2.2, Assumption B7 may still hold even when Z_3 is an empty set and Z_1 is discrete, since W is assumed to have full support. In such a case, identification primarily

relies on the factor structure and the variation of the covariates in the selection equation, rather than that in the outcome equation. In this respect, this identification result is similar in spirit to Theorem 2.1 and different from the existing identification results in the literature for triangular binary models, e.g., [Vytlačil and Yıldız \(2007\)](#) and [Vuong and Xu \(2017\)](#). More generally, in Section C.2 in the supplement we establish that the factor model provides identification restrictions that are not otherwise available.¹⁴

Finally, we can relax the rank invariance condition to rank similarity by replacing $Y_1 = \mathbf{1}\{X_1 + \alpha_0 Y_2 - U \geq 0\}$ by $Y_1 = \mathbf{1}\{X_1 + \alpha_0 Y_2 - U(Y_2) \geq 0\}$. We then require $U(y_2) = \gamma_0 W + \eta_1(y_2)$ for $y_2 = 0, 1$. If Assumptions **B0–B7** hold with η_1 replaced by $(\eta_1(1), \eta_1(0))$ and $P(\eta_1(1) \leq e) = P(\eta_1(0) \leq e)$ for $e \in \mathbb{R}$, then we can still identify θ_0 by a similar argument as the proof of Theorem 3.1.

4. CONCLUSION

In this chapter, we explore the identifying power of linear factor structures in the context of simultaneous binary response models. We impose two alternative types of factor structures on the unobservables of the model. The first setup is a natural distribution-free extension of the bivariate Probit model, while the second model corresponds to a standard linear factor model with one common factor and two equation-specific idiosyncratic shocks. We establish that both factor models have identifying power in that they make it possible to relax some of the exclusion and support conditions typically required for identification in this class of models ([Vytlačil & Yıldız, 2007](#)). Overall, our analysis adds to our understanding of the identifying power of factor models, beyond their well-known usefulness to recover the joint distribution of potential outcomes from the marginal distributions.

The work here opens areas for future research. The factor structure we assume could prove useful in more general nonlinear models. For instance, non-triangular discrete systems have shown to be an effective way to model entry games in the empirical industrial organization literature – see, for example, [Tamer \(2003\)](#). However, as shown in [Khan and Nekipelov \(2018\)](#), identification of structural parameters in these models can be even more challenging than for the triangular model considered in this chapter, and furthermore, as shown recently in [Khan and Nekipelov \(2021\)](#), conducting valid uniform interest in all these models is very difficult. It would be useful to determine if factor structures on the unobservables could alleviate this problem. We leave this open question to future work.

A. Proof of Theorem 2.1

Note that

$$\begin{aligned} P^{11}(z_1, z_3, x) &= \int_{-\infty}^x F_{\Pi}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \gamma_0 v) f_V(v) dv \\ P^{10}(\tilde{z}_1, \tilde{z}_3, \tilde{x}) &= \int_x^{+\infty} F_{\Pi}(\tilde{z}_1' \lambda_0 + \tilde{z}_3' \beta_0 - \gamma_0 v) f_V(v) dv. \end{aligned}$$

Taking derivatives w.r.t. the third argument of the LHS function, we obtain

$$\begin{aligned}\partial_x P^{11}(z_1, z_3, x) / f_V(x) &= F_{\Pi}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \gamma_0 x) \\ -\partial_x P^{10}(\tilde{z}_1, \tilde{z}_3, \tilde{x}) / f_V(\tilde{x}) &= F_{\Pi}(\tilde{z}_1' \lambda_0 + \tilde{z}_3' \beta_0 - \gamma_0 \tilde{x}).\end{aligned}$$

By Assumption **A4**, we know that there exists pairs such that

$$Z_1' \lambda_0 + Z_3' \beta_0 + \alpha_0 - \gamma_0 X = \tilde{Z}_1' \lambda_0 + \tilde{Z}_3' \beta_0 - \gamma_0 \tilde{X}.$$

Because $F_{\Pi}(\cdot)$ is monotone increasing, we have

$$\begin{aligned}\partial_x P^{11}(Z_1, Z_3, X) / f_V(X) + \partial_x P^{10}(\tilde{Z}_1, \tilde{Z}_3, \tilde{X}) / f_V(\tilde{X}) &= 0 \\ \Leftrightarrow \alpha_0 + (Z_1 - \tilde{Z}_1)' \lambda_0 + (Z_3 - \tilde{Z}_3)' \beta_0 - \gamma_0 (X - \tilde{X}) &= 0\end{aligned}$$

Note the LHS of the above display is identified from data. Denote $Z_{1,1}$ as the first element of Z_1 , whose coefficient is set to one. The rest of Z_1 is denoted as $Z_{1,-1}$, whose coefficient is denoted as $\lambda_{0,-1}$. Then, we have

$$\alpha_0 + (Z_{1,-1} - \tilde{Z}_{1,-1})' \lambda_{0,-1} + (Z_3 - \tilde{Z}_3)' \beta_0 - \gamma_0 (X - \tilde{X}) = \tilde{Z}_{1,1} - Z_{1,1}.$$

Then, by Assumption **A4**, we can find $(z_1^{(l)}, z_3^{(l)}, x^{(l)})_{l=1}^d$ and $(\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}, \tilde{x}^{(l)})_{l=1}^d$ such that

$$\text{rank} \begin{pmatrix} 1 & \dots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \dots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \dots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ x^{(1)} - \tilde{x}^{(1)} & \dots & x^{(d)} - \tilde{x}^{(d)} \end{pmatrix} = d.$$

Then, we can identify $(\alpha_0, \lambda_0, \beta_0, \gamma_0)$ by solving the linear system that

$$\begin{aligned}\alpha_0 + (z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)})' \lambda_{0,-1} + (z_3^{(1)} - \tilde{z}_3^{(1)})' \beta_0 - \gamma_0 (x^{(1)} - \tilde{x}^{(1)}) &= \tilde{z}_{1,1}^{(1)} - z_{1,1}^{(1)}, \\ \vdots \\ \alpha_0 + (z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)})' \lambda_{0,-1} + (z_3^{(d)} - \tilde{z}_3^{(d)})' \beta_0 - \gamma_0 (x^{(d)} - \tilde{x}^{(d)}) &= \tilde{z}_{1,1}^{(d)} - z_{1,1}^{(d)}.\end{aligned}$$

This concludes the proof.

B. Proof of Theorem 3.1

For notation simplicity, we write $\tilde{W} = \nu_0 W$, $\tilde{\sigma}_0 = \sigma_0 / \nu_0$, $\tilde{\nu}_0 = 1 / \nu_0$, and

$$\begin{aligned}Y_2 &= \mathbf{1}\{X \geq \tilde{\nu}_0 \tilde{W} + \eta_2\} \\ Y_3 &= \tilde{W} + \eta_3 \\ Y_4 &= \tilde{\sigma}_0 \tilde{W} + \eta_4.\end{aligned}$$

Because Assumptions B2–B6 hold, by applying [Hu and Schennach \(2013, Theorem 1\)](#) to Y_3 and Y_4 , we can identify the densities for $\nu_0 W = \tilde{W}$, η_3 , and η_4 as well as $\sigma_0 / \nu_0 = \tilde{\sigma}_0$.

Then, we have

$$\begin{aligned} \partial_{y_3} P(Y_2 = 1, Y_3 \leq y_3 | X = x) &= \partial_{y_3} \int F_{\eta_2}(x - \tilde{\nu}_0 w) F_{\eta_3}(y_3 - w) f_{\tilde{W}}(w) dw \\ &= \int F_{\eta_2}(x - \tilde{\nu}_0 w) f_{\eta_3}(y_3 - w) f_{\tilde{W}}(w) dw. \end{aligned}$$

Applying Fourier transform w.r.t. y_3 on both sides, we have

$$\mathcal{F}(\partial_{y_3} P(Y_2 = 1, Y_3 \leq \cdot | X = x))(t) = \mathcal{F}(F_{\eta_2}(x - \tilde{\nu}_0 \cdot) f_{\tilde{W}}(\cdot))(t) \mathcal{F}(f_{\eta_3}(\cdot))(t),$$

where for a generic function $g(w)$,

$$\mathcal{F}(g(\cdot))(t) = \frac{1}{\sqrt{2\pi}} \int \exp(-2\pi i t w) g(w) dw.$$

Therefore,

$$\frac{\mathcal{F}^{-1} \left(\frac{\mathcal{F}(\partial_{y_3} P(Y_2 = 1, Y_3 \leq \cdot | X = x))(\cdot)}{\mathcal{F}(f_{\eta_3}(\cdot))(\cdot)} \right)(w)}{f_{\tilde{W}}(w)} = F_{\eta_2}(x - \tilde{\nu}_0 w), \quad (\text{B.3})$$

where for a generic function $g(w)$,

$$\mathcal{F}^{-1}(g(\cdot))(t) = \frac{1}{\sqrt{2\pi}} \int \exp(2\pi i t w) g(w) dw.$$

Note the LHS of (B.3) can be identified from data. We choose two pairs (x, w) and (x', w') such that $w \neq w'$ and

$$\begin{aligned} &\frac{\mathcal{F}^{-1} \left(\frac{\mathcal{F}(\partial_{y_3} P(Y_2 = 1, Y_3 \leq \cdot | X = x))(\cdot)}{\mathcal{F}(f_{\eta_3}(\cdot))(\cdot)} \right)(w)}{f_{\tilde{W}}(w)} \\ &= \frac{\mathcal{F}^{-1} \left(\frac{\mathcal{F}(\partial_{y_3} P(Y_2 = 1, Y_3 \leq \cdot | X = x'))(\cdot)}{\mathcal{F}(f_{\eta_3}(\cdot))(\cdot)} \right)(w')}{f_{\tilde{W}}(w')}. \end{aligned}$$

Then, given the monotonicity of F_{η_2} , we have

$$x - \tilde{\nu}_0 w = x' - \tilde{\nu}_0 w',$$

or

$$\tilde{\nu}_0 = (x - x') / (w - w'),$$

which is identified. Given the identification of $\tilde{\nu}_0$ and the distribution of \tilde{W} , we can identify the distribution of $W = \tilde{\nu}_0 \tilde{W}$. Recall $F_{\eta_1}(\cdot)$ and $f_{\eta_2}(\cdot)$ are the CDF and PDF of η_1 and η_2 , respectively. Then, we have

$$P(Y_2 = 1 | X = x) = P(W + \eta_2 \leq x).$$

Because X has full support, we can identify the distribution of $W + \eta_2$. Then, it follows from standard deconvolution argument and the fact that the distribution of W is identified that we can identify the distribution of η_2 . In addition, note that

$$\begin{aligned} P^{11}(z_1, z_3, x) &= P(Y_1 = 1, Y_2 = 1 | Z_1 = z_1, Z_3 = z_3, X = x) \\ &= \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \gamma_0 w) F_{\eta_2}(x - w) f_W(w) dw \end{aligned}$$

and

$$\begin{aligned} P^{10}(z_1, z_3, x) &= P(Y_1 = 1, Y_2 = 0 | Z_1 = z_1, Z_3 = z_3, X = x) \\ &= \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 - \gamma_0 w) (1 - F_{\eta_2}(x - w)) f_W(w) dw. \end{aligned}$$

Taking derivatives of $P^{11}(z_1, z_3, x)$ and $P^{10}(z_1, z_3, x)$ w.r.t. x , we have

$$\partial_x P^{11}(z_1, z_3, x) = \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - w) f_{\eta_2}(x - w) f_W(w) dw \quad (\text{B.4})$$

and

$$-\partial_x P^{10}(z_1, z_3, x) = \int F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 - \gamma_0 w) f_{\eta_2}(x - w) f_W(w) dw. \quad (\text{B.5})$$

Applying Fourier transform on both sides of (B.4) and (B.5), we have

$$\mathcal{F}(\partial_x P^{11}(z_1, z_3, \cdot)) = \mathcal{F}(F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \cdot) f_W(\cdot)) \mathcal{F}(f_{\eta_2}(\cdot)) \quad (\text{B.6})$$

and

$$\mathcal{F}(-\partial_x P^{10}(z_1, z_3, \cdot)) = \mathcal{F}(F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 - \gamma_0 \cdot) f_W(\cdot)) \mathcal{F}(f_{\eta_2}(\cdot)).$$

Then, by (B.6), we can identify $F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \cdot)$ by

$$F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 + \alpha_0 - \gamma_0 \cdot) = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\partial_x P^{11}(z_1, z_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))} \right)(\cdot) / f_W(\cdot).$$

Similarly, we can identify

$$F_{\eta_1}(z_1' \lambda_0 + z_3' \beta_0 - \gamma_0 \cdot) = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(-\partial_x P^{10}(z_1, z_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))} \right)(\cdot) / f_w(\cdot).$$

Because $F_{\eta_1}(\cdot)$ is monotone increasing, we have

$$\begin{aligned} \mathcal{F}^{-1} \left(\frac{\mathcal{F}(\partial_x P^{11}(z_1, z_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))} \right)(w) / f_w(w) &= \mathcal{F}^{-1} \left(\frac{\mathcal{F}(-\partial_x P^{10}(\tilde{z}_1, \tilde{z}_3, \cdot))}{\mathcal{F}(f_{\eta_2}(\cdot))} \right)(\tilde{w}) / f_w(\tilde{w}) \\ \Leftrightarrow \alpha_0 + (z_1 - \tilde{z}_1)' \lambda_0 + (z_3 - \tilde{z}_3)' \beta_0 - \gamma_0 (w - \tilde{w}) &= 0 \end{aligned}$$

Then, by Assumption B7, we can find $(z_1^{(l)}, z_3^{(l)}, w^{(l)})_{l=1}^d$ and $(\tilde{z}_1^{(l)}, \tilde{z}_3^{(l)}, \tilde{w}^{(l)})_{l=1}^d$ such that

$$\text{rank} \begin{pmatrix} 1 & \cdots & 1 \\ z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)} & \cdots & z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)} \\ z_3^{(1)} - \tilde{z}_3^{(1)} & \cdots & z_3^{(d)} - \tilde{z}_3^{(d)} \\ w^{(1)} - \tilde{w}^{(1)} & \cdots & w^{(d)} - \tilde{w}^{(d)} \end{pmatrix} = d.$$

Then, we can identify $(\alpha_0, \lambda_0, \beta_0, \gamma_0)$ by solving the linear system that

$$\begin{aligned} \alpha_0 + (z_{1,-1}^{(1)} - \tilde{z}_{1,-1}^{(1)})' \lambda_{0,-1} + (z_3^{(1)} - \tilde{z}_3^{(1)})' \beta_0 - \gamma_0 (w^{(1)} - \tilde{w}^{(1)}) &= \tilde{z}_{1,1}^{(1)} - z_{1,1}^{(1)}, \\ \vdots \\ \alpha_0 + (z_{1,-1}^{(d)} - \tilde{z}_{1,-1}^{(d)})' \lambda_{0,-1} + (z_3^{(d)} - \tilde{z}_3^{(d)})' \beta_0 - \gamma_0 (w^{(d)} - \tilde{w}^{(d)}) &= \tilde{z}_{1,1}^{(d)} - z_{1,1}^{(d)}. \end{aligned}$$

This concludes the proof.

C. Identification With and Without Factor Structure

C.1. Identification Without Auxiliary Measurements

In this section, we discuss the information content of factor structure. For illustration purpose, we focus on the “condensed” model:

$$\begin{aligned} Y_1 &= \mathbf{1}\{X_1 + \alpha_0 Y_2 - U \geq 0\} \\ Y_2 &= \mathbf{1}\{X - V \geq 0\}. \end{aligned} \tag{C.7}$$

Assumption 1.

1. $(X_1, X) \perp (U, V)$.

2. (X_1, X) are continuously distributed with absolute continuous joint density w.r.t. Lebesgue measure. The conditional support of X_1 given X is $[a, b]$.

3. V is continuously distributed over \Re and its density w.r.t. Lebesgue measure exist.

Theorem C.1. If Assumption 1 holds, then $|\alpha_0| \leq b - a$ is necessary and sufficient for α_0 to be identified.

We note that under Assumption 1, $|\alpha_0| \leq b - a$ is equivalent to the fact that we can find x_1 and \tilde{x}_1 in the support of X_1 such that $\alpha_0 = x_1 - \tilde{x}_1$.

Next, we assume, in addition to Assumption 1, the factor structure, i.e., (2.6) in Section 2. Our rank estimator can be written as an M-estimator

$$\hat{\theta} = \arg \max_{\theta} Q_n(\theta) \equiv \sum_{i \neq j} \hat{g}_{i,j}(\theta)$$

in which

$$\begin{aligned} \hat{g}_{i,j}(\theta) = & [1\{\partial_2 \hat{P}^{11}(X_{1,i}, X_i) / \hat{f}_V(X_i) + \partial_2 \hat{P}^{10}(X_{1,j}, X_j) / \hat{f}_V(X_j) \geq 0\} \\ & + 1\{\Phi(X_{1,i}, X_i, X_{1,j}, X_j; \theta) \geq 0\} \\ & + 1\{\partial_2 \hat{P}^{11}(X_{1,i}, X_i) / \hat{f}_V(X_i) + \partial_2 \hat{P}^{10}(X_{1,j}, X_j) / \hat{f}_V(X_j) < 0\} \\ & + 1\{\Phi(X_{1,i}, X_i, X_{1,j}, X_j; \theta) < 0\}], \end{aligned}$$

with

$$\Phi(x_1, x, \tilde{x}_1, \tilde{x}; \theta) = x_1 + \alpha - \gamma x - (\tilde{x}_1 - \gamma \tilde{x}).$$

We will study the asymptotic properties of this estimator in the supplementary material.

The information content explored by the M-estimator can be summarized as follows:

$$\begin{aligned} \mathcal{A}_2(\theta) = & \{(X_1, \tilde{X}_1, X, \tilde{X}), \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta_0) \geq 0 > \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta) \\ & \text{or } \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta_0) < 0 \leq \Phi(X_1, X, \tilde{X}_1, \tilde{X}; \theta)\}. \end{aligned}$$

Then we cannot distinguish, from the true parameter θ_0 , all impostors in

$$\overline{\mathcal{A}}_2 = \{\theta : P(\mathcal{A}_2(\theta)) = 0\}.$$

In the condensed model, if $\text{Supp}(X_1, X) = [a, b] \times [c, d]$, then θ_0 is identified if $|\alpha_0| < b - a + |\gamma_0|(d - c)$. Recall Theorem C.1, without imposing factor structure, the necessary and sufficient condition for achieving identification is $|\alpha_0| \leq b - a$. Therefore, the blue area in the Fig. 1 is the additional parts of parameter space that are identified with factor structure but not otherwise.

Theorem C.2. Assumption 1 holds. When $|\alpha_0| > b - a$, the sharp identified set for α_0 is

$$\mathcal{A}^* = \{\alpha : \alpha > b - a \text{ if } \alpha_0 > 0 \text{ and } \alpha < a - b \text{ if } \alpha_0 < 0\}.$$

Theorem C.2 highlights that, in the case without the factor structure and α_0 does not satisfy the parameter restriction, except for the fact that the sign of α_0 is

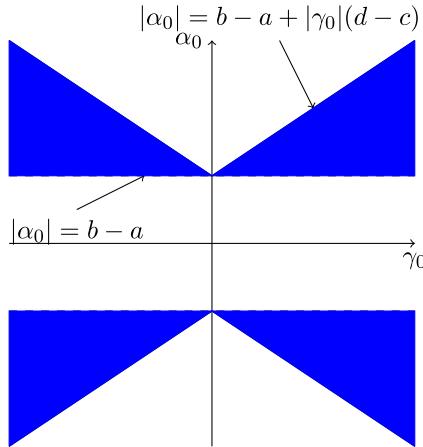


Fig. 1 Identifying Power of Factor Structure.

identified, we actually cannot say much about the value of $|\alpha_0|$. When we assume the factor structure, the parameter is still not identified if $|\alpha_0| > b - a + |\gamma_0|(d - c)$. In addition, suppose $\alpha_0 > 0$. In this case, if we do not impose factor structure, by Theorem C.2, the sharp identified set is $\{\alpha : \alpha > b - a\}$ while with the factor structure, the identified set (not necessarily sharp) is $\alpha > b - a + |\gamma_0|(d - c)$. This implies, when identification fails in both cases, the blue area is also the extra identifying power on the identified set given by the factor structure.

C.2. Identification with Two Auxiliary Measurements

Next, we expand our condensed model to include two continuous measurements. We show in this case, without the factor structure, α_0 is not identified. This is in contrast with the identification result established in Theorem 3.1.

Suppose in addition to (C.7), we also observe two continuous measurements of W denoted as Y_3 and Y_4 . One example of such Y_3 and Y_4 are described in (3.2).

Assumption 2.

1. $(X_1, X) \perp (U, V, Y_3, Y_4)$.
2. (X_1, X) are continuously distributed with absolute continuous joint density w.r.t. Lebesgue measure. The conditional support of X_1 given X is $[a, b]$.
3. V is continuously distributed over \mathbb{R} and its density w.r.t. Lebesgue measure exist.

Theorem C.3. If Assumption 2 holds, then $|\alpha_0| \leq b - a$ is necessary and sufficient for α_0 to be identified.

The proof of Theorem C.3 is similar to that of Theorem C.1, and thus, is omitted. In the proof of Theorem C.1, we show that when $|\alpha_0| > b - a$, we can find an imposter $\alpha \neq \alpha_0$ and \tilde{U} such that for any $x_1 \in [a, b]$ and any $v \in \text{Supp}(V)$, we have

$$\begin{aligned} P(\tilde{U} \leq x_1 + \alpha | V = v) &= P(U \leq x_1 + \alpha_0 | V = v) \\ P(\tilde{U} \leq x_1 | V = v) &= P(U \leq x_1 | V = v). \end{aligned}$$

This implies the conditional CDF of (Y_1, Y_2) given (X_1, X) under the DGPs (U, V, α_0) and (\tilde{U}, V, α) are the same, and thus, α_0 is observationally equivalent to the impostor α . Similarly, with the two continuous measurements, we can use the exact same construction of \tilde{U} and α to show that, for any $x_1 \in [a, b]$ and $(v, y_3, y_4) \in \text{Supp}(V, Y_3, Y_4)$, we have

$$\begin{aligned} P(\tilde{U} \leq x_1 + \alpha | V = v, Y_3 = y_3, Y_4 = y_4) &= P(U \leq x_1 + \alpha_0 | V = v, Y_3 = y_3, Y_4 = y_4) \\ P(\tilde{U} \leq x_1 | V = v, Y_3 = y_3, Y_4 = y_4) &= P(U \leq x_1 | V = v, Y_3 = y_3, Y_4 = y_4). \end{aligned}$$

This implies the conditional CDF of (Y_1, Y_2, Y_3, Y_4) given (X_1, X) under the DGPs $(U, V, Y_3, Y_4, \alpha_0)$ and $(\tilde{U}, V, Y_3, Y_4, \alpha)$ are the same too. Such non-identification result holds even when X has full support.

D. Proof of Theorem C.1

Denote $P^{ij}(x_1, x) = \text{Prob}(Y_1 = i, Y_2 = j | X_1 = x_1, X = x)$. Then

$$\begin{aligned} P^{11}(x_1, x) &= \int_{-\infty}^x F_U(x_1 + \alpha_0 | V = v) f(v) dv \\ P^{10}(\tilde{x}_1, x) &= \int_x^{+\infty} F_U(\tilde{x}_1 | V = v) f(v) dv. \end{aligned} \tag{D.8}$$

Taking derivatives w.r.t. the second argument of the LHS function, we have

$$\begin{aligned} \partial_2 P^{11}(x_1, x) &= F_U(x_1 + \alpha_0 | V = x) f(x) \\ \partial_2 P^{10}(\tilde{x}_1, x) &= -F_U(\tilde{x}_1 | V = x) f(x). \end{aligned}$$

If $|\alpha_0| \leq b - a$, then there exists a pair (x_1, \tilde{x}_1) such that $x_1 + \alpha_0 = \tilde{x}_1$. This pair can be identified by checking the equation below:

$$\partial_2 P^{11}(x_1, x) / f(x) + \partial_2 P^{10}(\tilde{x}_1, x) / f(x) = 0.$$

This concludes the sufficient part.

When $\alpha_0 < a - b$, for any $\alpha < \alpha_0$, we can define

$$\begin{aligned} \tilde{U} &= U + \alpha - \alpha_0 \text{ if } U \leq b + \alpha_0 \\ \tilde{U} &= U \text{ if } U > b + \alpha_0 \end{aligned}$$

Then for any $x_1 \in [a, b]$,

$$\begin{aligned} P(\tilde{U} \leq x_1 + \alpha | V = v) &= P(\tilde{U} \leq x_1 + \alpha, U \leq b + \alpha_0 | V = v) \\ &\quad + P(\tilde{U} \leq x_1 + \alpha, U > b + \alpha_0 | V = v) \\ &= P(U \leq x_1 + \alpha_0 | V = v) \end{aligned}$$

$$\begin{aligned}
P(\tilde{U} \leq x_1 | V = v) &= P(\tilde{U} \leq x_1, U \leq b + \alpha_0 | V = v) \\
&\quad + P(\tilde{U} \leq x_1, U > b + \alpha_0 | V = v) \\
&= P(U \leq b + \alpha_0, U \leq x_1 + \alpha_0 - \alpha | V = v) \\
&\quad + P(b + \alpha_0 < U \leq x_1 | V = v) \\
&= P(U \leq b + \alpha_0 | V = v) + P(b + \alpha_0 < U \leq x_1 | V = v) \\
&= P(U \leq x_1 | V = v),
\end{aligned}$$

where the third equality holds because, since $\alpha_0 < a - b$ and $\alpha < \alpha_0$, $b + \alpha_0 \leq x_1 + \alpha_0 - \alpha$ for $x_1 \in [a, b]$. Let $G_{U,V}$ and $G_{\tilde{U},V}$ be the joint distribution of (U, V) and (\tilde{U}, V) respectively. Then the above calculation with (D.8) imply that $(\alpha_0, G_{U,V})$ and $(\alpha, G_{\tilde{U},V})$ produce the identical pair $(P^{11}(x_1, x), P^{10}(x_1, x))$. In addition, the distribution of V is unchanged so that $P(Y_2 = 1 | X = x)$ is identified from data. Therefore, $(\alpha_0, G_{U,V})$ and $(\alpha, G_{\tilde{U},V})$ are observationally equivalent.

Similarly, when $\alpha_0 > b - a$, for any $\alpha > \alpha_0$, we can define

$$\begin{aligned}
\tilde{U} &= U + \alpha - \alpha_0 \text{ if } U > a + \alpha_0 \\
\tilde{U} &= U \text{ if } U \leq a + \alpha_0
\end{aligned}$$

Then for any $x_1 \in [a, b]$,

$$\begin{aligned}
P(\tilde{U} \leq x_1 + \alpha | V = v) &= P(\tilde{U} \leq x_1 + \alpha, U \leq a + \alpha_0 | V = v) \\
&\quad + P(\tilde{U} \leq x_1 + \alpha, U > a + \alpha_0 | V = v) \\
&= P(U \leq a + \alpha_0 | V = v) \\
&\quad + P(a + \alpha_0 < U \leq x_1 + \alpha_0 | V = v) \\
&= P(U \leq x_1 + \alpha_0 | V = v).
\end{aligned}$$

$$\begin{aligned}
P(\tilde{U} \leq x_1 | V = v) &= P(\tilde{U} \leq x_1, U \leq a + \alpha_0 | V = v) \\
&\quad + P(\tilde{U} \leq x_1, U > a + \alpha_0 | V = v) \\
&= P(U \leq x_1 | V = v),
\end{aligned}$$

where we use the facts that $x_1 \leq a + \alpha_0$ and $x_1 - a < \alpha$ for $x_1 \in [a, b]$. So again, $(\alpha_0, G_{U,V})$ and $(\alpha, G_{\tilde{U},V})$ are observationally equivalent.

E. Proof of Theorem C.2

The sign of α_0 is identified by the data. In the following, we focus on deriving the results when $\alpha_0 > b - a$. By the proof of Theorem C.1, we have already shown that all $\alpha > \alpha_0$ is in the identified set. Now we consider $\frac{b - a + \alpha_0}{2} \leq \alpha < \alpha_0$.

$$\begin{aligned}
\tilde{U} &= U + \alpha - \alpha_0 \text{ if } U > a + \alpha \\
\tilde{U} &= U \text{ if } U \leq a + \alpha
\end{aligned}$$

Then for any $x_1 \in [a, b]$,

$$\begin{aligned} P(\tilde{U} \leq x_1 + \alpha | V = v) &= P(\tilde{U} \leq x_1 + \alpha, U \leq a + \alpha | V = v) \\ &\quad + P(\tilde{U} \leq x_1 + \alpha, U > a + \alpha | V = v) \\ &= P(U \leq a + \alpha | V = v) \\ &\quad + P(a + \alpha < U \leq x_1 + \alpha_0 | V = v) \\ &= P(U \leq x_1 + \alpha_0 | V = v). \end{aligned}$$

$$\begin{aligned} P(\tilde{U} \leq x_1 | V = v) &= P(\tilde{U} \leq x_1, U \leq a + \alpha | V = v) \\ &\quad + P(\tilde{U} \leq x_1, U > a + \alpha | V = v) \\ &= P(U \leq x_1 | V = v) \\ &\quad + P(U \leq x_1 + \alpha_0 - \alpha, U > a + \alpha | V = v) \\ &= P(U \leq x_1 | V = v). \end{aligned}$$

Here note that the last equality is because $x_1 + \alpha_0 - \alpha \leq b + \alpha_0 - \alpha \leq a + \alpha$ if $\alpha \geq \frac{b-a+\alpha_0}{2}$. Denote $\alpha^{(1)} = \frac{b-a+\alpha_0}{2}$. Then we have shown that there exists $U^{(1)}(\alpha)$ which only depends on α such that for any $x_1 \in [a, b]$, any v and any $\alpha_0 > \alpha \geq \alpha^{(1)}$

$$\begin{aligned} P(U^{(1)}(\alpha) \leq x_1 + \alpha | V = v) &= P(U \leq x_1 + \alpha_0 | V = v) \\ P(U^{(1)}(\alpha) \leq x_1 | V = v) &= P(U \leq x_1 | V = v). \end{aligned}$$

In particular, there exists $U^{(1)}(\alpha^{(1)})$ such that

$$\begin{aligned} P(U^{(1)}(\alpha^{(1)}) \leq x_1 + \alpha^{(1)} | V = v) &= P(U \leq x_1 + \alpha_0 | V = v) \\ P(U^{(1)}(\alpha^{(1)}) \leq x_1 | V = v) &= P(U \leq x_1 | V = v). \end{aligned}$$

Now repeating the above construction but replacing U with $U^{(1)}$ and α_0 with $\alpha^{(1)}$, we have for any $\alpha^{(1)} > \alpha \geq \alpha^{(2)} \equiv \frac{b-a+\alpha^{(1)}}{2}$, there exists $U^{(2)}(\alpha)$ such that for any $x_1 \in [a, b]$, any v and any $\alpha^{(1)} > \alpha \geq \alpha^{(2)}$,

$$\begin{aligned} P(U^{(2)}(\alpha) \leq x_1 + \alpha^{(2)} | V = v) &= P(U^{(1)}(\alpha^{(1)}) \leq x_1 + \alpha^{(1)} | V = v) \\ &= P(U \leq x_1 + \alpha_0 | V = v) \\ P(U^{(2)}(\alpha) \leq x_1 | V = v) &= P(U^{(1)}(\alpha^{(1)}) \leq x_1 | V = v) \\ &= P(U \leq x_1 | V = v). \end{aligned}$$

This concludes that any α such that $\alpha_0 > \alpha \geq \alpha^{(2)}$ is in the identified set. In general, by repeating the procedure k times, we have that any α such that

$$\alpha_0 > \alpha \geq \alpha^{(k)} = \left(1 - \frac{1}{2^k}\right)(b-a) + \frac{\alpha_0}{2^k}$$

is in the identified set. For any $\alpha > b-a$, there exists some finite k such that $\alpha > \left(1 - \frac{1}{2^k}\right)(b-a) + \frac{\alpha_0}{2^k}$. This concludes the result that $\alpha > b-a$ is in the identified set.

Finally, since if $\alpha > b-a$, $\partial_2 P^{11}(x_1, x) + \partial_2 P^{10}(\tilde{x}_1, x) > 0$ for all pairs of (x_1, x) and (\tilde{x}_1, x) while, if $\alpha \leq b-a$, at least there exists one pair (x_1, x) and (\tilde{x}_1, x) such that $\partial_2 P^{11}(x_1, x) + \partial_2 P^{10}(\tilde{x}_1, x) \leq 0$. This implies $\alpha \leq b-a$ is not in the identified set. Therefore, the sharp identified set when $\alpha_0 > b-a$ is $(b-a, \infty)$.

When $\alpha_0 < a-b$, a symmetric argument implies that the identified set is $(-\infty, a-b)$.

NOTES

1. See also [Abbring and Heckman \(2007\)](#) for an extensive discussion of factor structures and prior studies using these models in the context of treatment effect estimation.

2. See also recent work by [Lewbel et al. \(2020\)](#), who study the identification of a triangular linear model assuming that the disturbances are related through a factor model.

3. As is always the case in models with binary outcomes, both the interpretation and the usefulness of regression coefficients warrant explanation. In the model considered here the coefficient on the treatment variable and the coefficients on exogenous variables in the binary outcome equation enable us to construct “equivalence classes” to answer important policy questions. For example consider the case where the dummy endogenous variable is job training, the exogenous regressor is years’ experience and the outcome variable is employment status. Knowing all coefficients would be informative on how many additional years of experience would be needed to compensate for a lack of training so the probability of being employed stays the same.

4. In Section C in the supplement, we establish the sharp identified set of α_0 when the support condition for point-identification is violated. This result highlights that, except for the fact that the sign of α_0 is identified, we generally cannot say much about the value of $|\alpha_0|$. Related work by [Shaikh and Vytlacil \(2011\)](#) also provides partial identification results for a triangular binary model. That the bounds for α_0 are generally tighter in their analysis reflects the identifying power of the additional support restrictions that they impose.

5. We thank the referee for pointing this out.

6. See also recent work by [Han and Lee \(2019\)](#) who study semiparametric estimation and inference in the framework considered by [Han and Vytlacil \(2017\)](#).

7. [Han and Vytlacil \(2017\)](#) establish their identification of the coefficient on the endogenous regressor (Theorems 4.2 and 5.1) under the assumption that the marginal distributions F_ε and F_ν are known. Then, they verify this condition by showing the identification of these two marginal distributions using large support common regressors.

8. For instance, suppose that (U, V) has a Gaussian copula with correlation ρ , and that the marginal distributions of U and V are uniform $[0, 1]$. It then follows that, denoting by $\Phi(\cdot)$ the standard normal cdf, $(\Phi^{-1}(U), \Phi^{-1}(V))$ is bivariate normal with correlation ρ , which in turn yields the following non-linear relationship between U and V : $U = \Phi(\rho\Phi^{-1}(V) + W)$, where W is normally distributed and independent from V .

9. To see this, note that if, say, $z_1' \lambda_0 + \alpha_0 > \tilde{z}_1' \lambda_0$, then we can find τ_1, τ_2 such that $-z_1' \lambda_0 - \alpha_0 \leq F_{-U}^{-1}(\tau_1) < -\tilde{z}_1' \lambda_0$ and $F_{-U}^{-1}(\tau_2) < -z_1' \lambda - \alpha_0 < -\tilde{z}_1' \lambda_0$. This violates the above requirement, and thus, shows that Vuong and Xu (2017, Assumption C(ii)) implies (2.12). On the other hand, if $z_1' \lambda_0 + \alpha_0 = \tilde{z}_1' \lambda_0$, then Vuong and Xu (2017, Assumption C(ii)) holds trivially.

10. An alternative approach to identifying this parameter can be found in Lewbel (2000). In his approach, a second equation to model the endogenous variable is not needed, nor is the factor structure we impose. However, he imposes a strong support condition on a variable like Z_3 requiring that it exceeds the length of the unobservable U .

11. In practice, the continuous measurements might also depend on some observable characteristics. Our analysis goes through in this case after residualizing Y_3 and Y_4 .

12. Note that Hu and Schennach (2013, Assumptions 5 and 6) hold automatically in our model with $\nu_0 \neq 0$.

13. See Scott-Clayton (2011) for an evaluation of a program of this kind (PROMISE scholarship in West Virginia), and for a discussion of similar merit-based scholarship programs in place in other states.

14. Specifically, we consider a version of the model (3.1), where we do not impose the factor structure and allow for an arbitrary (unknown to econometricians) dependence structure across the unobservables of the model. In this case, we show non-identification of α_0 as long as $|\alpha_0| > b - a$, where $[a, b]$ denotes the conditional support of X_1 given X and, consistent with our Assumption B1, X has full support on the real line. However, by imposing the factor structure (and other conditions implied by B0–B7), Theorem 3.1 shows that α_0 is identified for this model even when $|\alpha_0| > b - a$.

ACKNOWLEDGMENTS

We thank the Co-Editor, Simon Lee, an anonymous referee, Arthur Lewbel, Serena Ng, and seminar participants at Arizona State University, Emory, Michigan State, Shanghai University of Finance and Economics, University of Arizona, as well as conference participants at the 2015 SEA meetings for helpful comments. We also thank Zhangchi Ma and Qingsong Yao for excellent research assistance. Zhang acknowledges the financial support from Singapore Ministry of Education Tier 2 grant under grant MOE2018-T2-2-169 and the Lee Kong Chian fellowship.

REFERENCES

- Abbring, J., & Heckman, J. (2007). Econometrics evaluation of social programs, Part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of econometrics* (Vol. 6B). North Holland.
- Abowd, J. M., & Card, D. (1989). On the covariance structure of earnings and hours changes. *Econometrica*, 57(2), 411–445.
- Abrevaya, J., Hausman, J., & Khan, S. (2010). Testing for causal effects in a generalized regression model with endogenous regressors. *Econometrica*, 78(6), 2043–2061.
- Ashworth, J., Hotz, V. J., Maurel, A., & Ransom, T. (2021). Changes across cohorts in wage returns to schooling and early work experiences. *Journal of Labor Economics*, 39(4), 931–964.
- Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1), 191–221.

- Bierens, H., & Hartog, J. (1988). Non-linear regression with discrete explanatory variables, with an application to the earnings function. *Journal of Econometrics*, 38(3), 269–299.
- Blundell, R., & Powell, J. (2004). Endogeneity in binary response models. *Review of Economic Studies*, 71(3), 655–679.
- Bonhomme, S., & Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies*, 77(2), 491–533.
- Carneiro, P., Hansen, K., & Heckman, J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2), 361–422.
- 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(2), 191–208.
- Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, 73(5), 1525–1550.
- Chiburis, R. (2010). Semiparametric bounds on treatment effects. *Journal of Econometrics*, 159(2), 267–275.
- Cunha, F., Heckman, J. J., & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3), 883–931.
- Gourieroux, C., Monfort, A., & Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics*, 196(1), 111–126.
- Han, A. (1987). Non-parametric analysis of a generalized regression model: The maximum rank correlation estimator. *Journal of Econometrics*, 35(2–3), 303–316.
- Han, S., & Lee, S. (2019). Estimation in a generalization of a bivariate probit models with dummy endogenous regressors. *Journal of Applied Econometrics*, 34(6), 994–1015.
- Han, S., & Vytlacil, E. J. (2017). Identification in a generalization of bivariate probit models with endogenous regressors. *Journal of Econometrics*, 199(1), 63–73.
- Heckman, J., & Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136(2), 341–396.
- Heckman, J. J., Humphries, J., & 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., & Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3), 669–738.
- Heckman, J. J., & Vytlacil, E. J. (2007a). Econometric evaluation of social programs, Part I: Causal models, structural models and econometric policy evaluation. In *Handbook of econometrics* (Vol. 6, pp. 4779–4874). North Holland.
- Heckman, J. J., & Vytlacil, E. J. (2007b). 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 *Handbook of econometrics* (Vol. 6, pp. 4875–5143). North Holland.
- Hu, Y., & Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1), 195–216.
- Hu, Y., & Schennach, S. M. (2013). Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association*, 108(501), 177–186.
- Hyvärinen, A., & Oja, E. (2000). Independent component analysis: *Algorithms and applications* (Vol. 13). Elsevier.
- Khan, S., & Nekipelov, D. (2018). Information structure and statistical information in discrete response models. *Quantitative Economics*, 9(2), 995–1017.
- Khan, S., & Nekipelov, D. (2021). On uniform inference in nonlinear models with endogeneity. *Journal of Econometrics*, forthcoming. <https://doi.org/10.1016/j.jeconom.2021.07.016>
- Klein, R., Shan, C., & Vella, F. (2015). Estimation of marginal effects in semiparametric selection models with binary outcomes. *Journal of Econometrics*, 185(1), 82–94.
- Klein, R., & Spady, R. (1993). An efficient semiparametric estimator for binary response model. *Econometrica*, 61(2), 387–421.

- Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1), 145–177.
- Lewbel, A., Schennach, S. M., & Zhang, L. (2020). *Identification of a triangular two equation system without instruments* [Working Paper].
- Moneta, A., Hoyer, D. E. P. O., & Coad, A. (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics*, 75(5), 705–730.
- Mourifié, I. (2015). Sharp bounds on treatment effects in a binary triangular system. *Journal of Econometrics*, 187(1), 74–81.
- Park, J., & Phillips, P. (2000). Nonstationary binary choice. *Econometrica*, 68(5), 1249–1280.
- Scott-Clayton, J. (2011). Discussion of “simple estimators for invertible index models” by Ahn et al. *Journal of Human Resources*, 46(3), 614–646.
- Shaikh, A. M., & Vytlacil, E. (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica*, 79(3), 949–955.
- Tamer, E. (2003). Incomplete bivariate discrete response model with multiple equilibria. *Review of Economic Studies*, 70(1), 147–167.
- Vuong, Q., & Xu, H. (2017). Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics*, 8(2), 589–610.
- Vytlacil, E. J., & Yıldız, N. (2007). Dummy endogenous variables in weakly separable models. *Econometrica*, 75(3), 757–779.

PART V

RETROSPECTIVE

This page intentionally left blank

CHAPTER 15

FORTY YEARS OF *ADVANCES IN ECONOMETRICS*

Asli Ogunc^a and Randall C. Campbell^b

^a*Department of Management and Economics, Texas A&M University-Commerce, Commerce, Texas, United States*

^b*Department of Finance and Economics, Mississippi State University, Starkville, Mississippi, United States*

ABSTRACT

Advances in Econometrics is a series of research volumes first published in 1982 by JAI Press. The authors present an update to the history of the *Advances in Econometrics* series. The initial history, published in 2012 for the 30th Anniversary Volume, describes key events in the history of the series and provides information about key authors and contributors to *Advances in Econometrics*. The authors update the original history and discuss significant changes that have occurred since 2012. These changes include the addition of five new Senior Co-Editors, seven new AIE Fellows, an expansion of the AIE conferences throughout the United States and abroad, and the increase in the number of citations for the series from 7,473 in 2012 to over 25,000 by 2022.

Keywords: *Advances in Econometrics*; history; AIE conference; AIE editors; AIE contributors; AIE citation count

1. INTRODUCTION

Advances in Econometrics is a series of research volumes first published in 1982 by JAI Press. Beginning in 1999 the series has been published online by Emerald Insight

Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications

Advances in Econometrics, Volume 45B, 413–435

Copyright © 2023 by Asli Ogunc and Randall C. Campbell

Published under exclusive licence by Emerald Publishing Limited

ISSN: 0731-9053/doi:[10.1108/S0731-90532023000045B017](https://doi.org/10.1108/S0731-90532023000045B017)

(<https://www.emerald.com/insight/publication/issn/0731-9053>). The *Advances in Econometrics* website (<https://advancesineconometrics.com>) gives the mission of the publication as,

To annually publish original econometrics papers on designated topics with the intention of expanding the use of developed and emerging econometric techniques by disseminating ideas on the theory and practice of econometrics throughout empirical economic, business and social literature.

Throughout its 40 years of existence, this series certainly achieved its mission with outstanding econometricians contributing impactful papers on an impressive variety of topics. Through the first 43 volumes, 880 different authors have contributed at least one paper to *Advances in Econometrics* including five Nobel Prize winners; most recently, UC Berkeley's David E. Card who was awarded the Nobel Prize in 2021 and contributed to *Advances in Econometrics*, Volume 38. In addition, 60 of 150 members of the "Econometricians' Hall of Fame" [see [Baltagi, 2003](#); [Table 5](#)] have contributed at least one paper to *Advances in Econometrics*. During the period from 1982 to 2022, JAI Press and Emerald Insight published 44 Volumes of *Advances in Econometrics*. These volumes featured 585 papers with over 25,700 citations.¹ The number of citations has more than tripled in the 10 years since [Campbell and Ogunc \(2012\)](#) wrote the initial history, increasing from 7,473 citations from 1982 to 2012 to the current total of 25,717.² Volume 44, which is scheduled for publication in September 2022, consists of Essays in Honor of Fabio Canova. [Table 1](#) lists the titles and editors for the first 44 volumes.

The rest of this chapter is organized as follows. Sections 2–6 summarize the history of the volume and are structured according to important events. Section 7 summarizes recent developments, lists key contributors, and concludes the chapter.

Table 1. AIE Titles and Editors: Volumes 1–44.

Volume	Year	Title	Editors
1	1982	Studies of Consumer and Worker Behavior	R.L. Basmann, George F. Rhodes, Jr.
2	1983	Exact Distribution Analysis in Linear Simultaneous Equation Models	R.L. Basmann, George F. Rhodes, Jr.
3	1984	Economic Inequality: Measurement and Policy	R.L. Basmann, George F. Rhodes, Jr.
4	1985	Economic Inequality: Survey Methods and Measurement	R.L. Basmann, George F. Rhodes, Jr.
5	1986	Innovations in Quantitative Economics: Essays in Honor of R.L. Basmann	Daniel J. Slottje
6	1987	Computation and Simulation	Thomas B. Fomby, G. F. Rhodes, Jr.

(Continued)

Table 1. (Continued)

Volume	Year	Title	Editors
7	1988	Nonparametric and Robust Inference	Thomas B. Fomby, G. F. Rhodes, Jr.
8	1990	Co-Integration, Spurious Regressions, and Unit Roots	Thomas B. Fomby, G. F. Rhodes, Jr.
9	1991	Econometric Methods and Models for Industrial Organizations	Thomas B. Fomby, G. F. Rhodes, Jr.
10	1994	Simulating and Analyzing Industrial Structure	Thomas B. Fomby, G. F. Rhodes, Jr.
11A	1996	Bayesian Computational Methods and Applications	Thomas B. Fomby, R. Carter Hill
11B	1996	Bayesian Methods Applied to Time Series Data	Thomas B. Fomby, R. Carter Hill
12	1997	Applying Maximum Entropy to Econometric Problems	Thomas B. Fomby, R. Carter Hill
13	1998	Messy Data – Missing Observations, Outliers, and Mixed-Frequency Data	Thomas B. Fomby, R. Carter Hill
14	2000	Applying Kernel and Nonparametric Estimation to Economic Topics	Thomas B. Fomby, R. Carter Hill
15	2001	Nonstationary Panels, Panel Cointegration, and Dynamic Panels	B. Baltagi, T. Fomby, R. Carter Hill
16	2002	Econometric Models in Marketing	P. H. Franses, A. L.Montgomery
17	2003	Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later	Thomas B. Fomby, R. Carter Hill
18	2004	Spatial and Spatiotemporal Econometrics	James P. Lesage, R. Kelley Pace
19	2004	Applications of Artificial Intelligence in Finance and Economics	J. Binner, G. Kendall, S. Chen
20A	2006	Econometric Analysis of Financial and Economic Time Series	Thomas B. Fomby, Dek Terrell
20B	2006	Econometric Analysis of Financial and Economic Time Series	Thomas B. Fomby, Dek Terrell
21	2008	Modeling and Evaluating Treatment Effects in Econometrics	T. Fomby, R.C. Hill, D. Millimet, J. Smith, and E. Vytlacil
22	2008	Econometrics and Risk Management	J-P. Fouque, T. Fomby, K. Solna
23	2008	Bayesian Econometrics	S.Chib, W.Griffiths, G.Koop, D.Terrell
24	2009	Measurement Error: Consequences, Applications and Solutions	J. Binner, D. Edgerton, T. Elger
25	2009	Nonparametric Econometric Methods	Qi Li, Jeffrey S. Racine
26	2010	Maximum Simulated Likelihood Methods and Applications	William Greene, R. Carter Hill
27A	2011	Missing Data Methods: Cross-Sectional Methods and Applications	David M. Drukker
27B	2011	Missing Data Methods: Time-Series and Applications	David M. Drukker
28	2012	DSGE Models in Macroeconomics-Estimation, Evaluation, and New Develop.	Balke, Canova, Milani, Wynne
29	2012	Essays in Honor of Jerry Hausman	B. Baltagi, R.C. Hill, W. Newey,

(Continued)

Table 1. (Continued)

Volume	Year	Title	Editors
30	2012	30th Anniversary Edition	H. White
31	2013	Structural Econometric Models	Dek Terrell, Daniel Millimet
32	2013	VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims	Eugene Choo, Matthew Shum Thomas B. Fomby, L. Killian, A. Murphy
33	2014	Essays in Honor of Peter C.B. Phillips	Y. Chang, T. Fomby, J. Park
34	2014	Bayesian Model Comparison	Ivan Jeliazkov, Dale Poirier
35	2016	Dynamic Factor Models	Eric Hillebrand, Siem Jan Koopman
36	2016	Essays in Honor of Aman Ullah	G. Gonzalez-Rivera, R.C. Hill, T. Lee
37	2016	Spatial and Spatiotemporal Econometrics	Balke, Canova, Milani, Wynne
38	2017	Regression Discontinuity Designs: Theory and Applications	M.D. Cattaneo, J.C. Escanciano
39	2019	The Econometrics of Complex Survey Data: Theory and Applications	K. Huynh, D. Jacho-Chávez, G. Tripathi
40A	2019	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part A	Ivan Jeliazkov, Justin L. Tobias
40B	2019	Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling: Part B	Ivan Jeliazkov, Justin L. Tobias
41	2020	Essays in Honor of Cheng Hsiao	T. Li, M.H. Pesaran, D. Terrell
42	2020	The Econometrics of Networks	Á. De Paula, E. Tamer, M. Voia
43A	2022	Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling	A. Chudik, C. Hsiao, A. Timmermann
43B	2022	Essays in Honor of M. Hashem Pesaran: Panel Modeling, Micro Applications, and Econometric Methodology	A. Chudik, C. Hsiao, A. Timmermann
44A	2022	Essays in Honor of Fabio Canova: Part A	J. Dolado, L. Gambetti, C. Matthes
44B	2022	Essays in Honor of Fabio Canova: Part B	J. Dolado, L. Gambetti, C. Matthes

2. BEGINNINGS OF THE VOLUME: 1982–1986

The founding co-editors of *Advances in Econometrics* (hereinafter AIE) were the late Robert L. Basmann of Texas A&M University and George F. Rhodes, Jr. of Colorado State University. Professor Basmann received his Ph.D. from Iowa State University. He was well known for his pioneering work on simultaneous equations and two-stage least squares estimators and was named a Fellow of the Econometric Society in 1966. Professor Rhodes received his Ph.D. from Ohio State University. His early work also involved simultaneous equations. In addition, his empirical work in the field of Labor Economics is well known.

The first volume of AIE was published in 1982. The editors declared the criteria for papers published in AIE as “Papers shall make original contributions in economics and econometrics that lay major new foundations for continued study,” and “Papers shall be sufficiently long and writing style sufficiently mature and complete that they are self-contained. They should be readable and should not require more than occasional reference to previous works.” A later criterion added, “Papers should, where possible, contain empirical economic examples that illustrate the usefulness of described econometric techniques.”

A few key statistics for the AIE volumes, papers, and authors are provided in [Tables 2–5](#). [Table 2](#) lists the top 10 volumes by total number of citations (Panel A) and by number of new citations since the 30th Anniversary edition (Panel B). [Table 3](#) lists the top 10 AIE volumes by number of papers. [Table 4](#), Panel A, lists the top 10 most cited papers in the history of AIE while [Table 4](#), Panel B, lists the 10 most cited papers since the 30th Anniversary edition. [Table 5](#) lists all

Table 2. Panel A. Top 10 AIE Volumes by Number of Citations.

Volume	Title	Citations
15	Nonstationary Panels, Panel Cointegration, and Dynamic Panels	11,507
8	Co-Integration, Spurious Regressions, and Unit Roots	1,340
32	VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims	1,146
3	Economic Inequality: Measurement and Policy	1,056
38	Regression Discontinuity Designs: Theory and Applications	741
21	Modeling and Evaluating Treatment Effects in Econometrics	683
16	Econometric Models in Marketing	579
18	Spatial and Spatiotemporal Econometrics	579
20B	Econometric Analysis of Financial and Economic Time Series	506
1	Studies of Consumer and Worker Behavior	482

Table 2. Panel B. Top 10 AIE Volumes by Number of New Citations Since 30th Anniversary Edition.

Volume	Title	Citations
15	Nonstationary Panels, Panel Cointegration, and Dynamic Panels	8,431
32	VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims	1,146
38	Regression Discontinuity Designs: Theory and Applications	741
8	Co-Integration, Spurious Regressions, and Unit Roots	719
21	Modeling and Evaluating Treatment Effects in Econometrics	451
3	Economic Inequality: Measurement and Policy	446
35	Dynamic Factor Models	379
16	Econometric Models in Marketing	370
18	Spatial and Spatiotemporal Econometrics	349
25	Nonparametric Econometric Methods	349

contributors who have authored or co-authored at least four papers published in AIE; eight of these authors are AIE Fellows.

The first volume of the series includes papers on consumer and worker behavior in addition to papers on general econometric theory. The total number of

Table 3. Top 10 AIE Volumes by Number of Papers.

Volume	Title	Number of Papers
23	Bayesian Econometrics	21
33	Essays in Honor of Peter C.B. Phillips	20
36	Essays in Honor of Aman Ullah	19
29	Essays in Honor of Jerry Hausman	18
5	Innovations in Quantitative Economics: Essays in Honor of R.L. Basmann	16
35	Dynamic Factor Models	16
41	Essays in Honor of Cheng Hsiao	16
25	Nonparametric Econometric Methods	15
12	Applying Maximum Entropy to Econometric Problems	14
14	Applying Kernel and Nonparametric Estimation to Economic Topics	14
20B	Econometric Analysis of Financial and Economic Time Series	14
30	30th Anniversary Edition	14
43A	Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling	14

Table 4. Panel A. Top 10 AIE Papers by Number of Citations.

Author(s)	Paper	Volume	Citations
Jorg Breitung	The Local Power of some Unit Root Tests for Panel Data	15	3,219
Peter Pedroni	Fully Modified OLS for Heterogeneous Cointegrated Panels	15	3,215
Chihwa Kao and Min-Hsien Chiang	On the Estimation and Inference of a Cointegrated Regression in Panel Data	15	2,698
Richard Blundell, Stephen Bond, and Frank Windmeijer	Estimation in Dynamic Panel Data Models: Improving the Performance of the Standard GMM Estimator	15	1,287
Badi H. Baltagi and Chihwa Kao	Nonstationary Panels, Cointegration in Panels, and Dynamic Panels: A Survey	15	911
Fabio Canova and Matteo Ciccarelli	Panel Vector Autoregressive Models: A Survey	32	673
James E. Foster	On Economic Poverty: A Survey of Aggregate Measures	3	456
Dale W. Jorgenson, Lawrence J. Lau, and Thomas M. Stoker	The Transcendental Logarithmic Model of Aggregate Consumer Behavior	1	380
Joon Y. Park	Testing for Unit Roots and Cointegration by Variable Addition	8	356
Nanak Kakwani	On the Measurement of Tax Progressivity and Redistributive Effect of Taxes with Applications to Horizontal and Vertical Equity	3	299

Table 4. Panel B. Top 10 AIE Papers by Number of New Citations Since 30th Anniversary Edition.

Author(s)	Paper	Volume	Citations
Jorg Breitung	The Local Power of some Unit Root Tests for Panel Data	15	2,522
Peter Pedroni	Fully Modified OLS for Heterogeneous Cointegrated Panels	15	2,490
Chihwa Kao and Min-Hsien Chiang	On the Estimation and Inference of a Cointegrated Regression in Panel Data	15	2,007
Richard Blundell, Stephen Bond, and Frank Windmeijer	Estimation in Dynamic Panel Data Models: Improving the Performance of the Standard GMM Estimator	15	853
Fabio Canova and Matteo Ciccarelli	Panel Vector Autoregressive Models: A Survey	32	673
Badi H. Baltagi and Chihwa Kao	Nonstationary Panels, Cointegration in Panels, and Dynamic Panels: A Survey	15	477
David S. Lee and Justin McCrary	The Deterrence Effect of Prison: Dynamic Theory and Evidence	38	255
Clive W.J. Granger and Tae-Hwy Lee	Multicointegration	8	223
Alexander Chudik, Kamiar Mohaddes	Long-run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors	36	211
M. Hashem Pesaran, Mehdi Raissi			
Brigham R. Frandsen	Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable Is Discrete	38	192

citations for volume 1 is 482, which is the 10th highest number of total citations among AIE volumes, as shown in Table 2, Panel A. Volume 1 includes the Jorgenson, Lau, and Stoker article entitled “The Transcendental Logarithmic Model of Aggregate Consumer Behavior.” This article has 380 total citations, which makes it the 8th most cited paper in the history of AIE (Table 4, Panel A). The first volume includes papers by two authors who have 3 or more AIE publications, Dale W. Jorgenson and James J. Heckman. The paper “Models of Analysis of Labor Force Dynamics” by Christopher J. Flinn and James J. Heckman appeared in the first volume. With this paper, James J. Heckman (Nobel Prize in Economics, 2000) became the first of five current or future Nobel Prize winners to publish a paper in AIE. Professor Heckman also published papers in volumes 2 and 5 of AIE.

The second volume, entitled “Exact Distribution Analysis in Linear Simultaneous Equation Models,” was edited by Robert L. Basmann, who also contributed one of his sole authored articles. George Rhodes, Jr., who has the 3rd highest number of total AIE publications with 7, contributed two articles to the volume.

Volume 3, “Economics Inequality: Measurement and Policy,” again edited by Professors Basmann and Rhodes, was published in 1984. In his introduction, Professor Rhodes explains the role of the disciplines and fields that are

Table 5. Top Authors by Number of Papers Published in AIE.

Author	Number of Papers	Volumes	Total Citations
Badi H. Baltagi^a	14	11 (Vol. 14, 15, 17, 18, 29, 30, 33, 36, 37, 41, 43)	993
Aman Ullah^a	9	8 (Volumes 6, 7, 14, 15, 25, 33, 40, 43)	126
Tae-Hwy Lee^a	7	5 (Volumes 8, 20, 30, 40, 43)	283
George F. Rhodes, Jr.	7	4 (Volumes 2, 5, 6, 10)	9
R. Carter Hill^b	6	6 (Volumes 5, 6, 7, 17, 26, 29)	77
Cheng Hsiao	6	5 (Volumes 16, 33, 36, 41, 43)	62
Qi Li	6	6 (Volumes 13, 14, 15, 25, 27, 36)	62
Georges Bresson	5	4 (Volumes 17, 29, 41, 43)	30
Thomas B. Fomby^b	5	5 (Volumes 5, 6, 7, 17, 41)	14
Gary Koop	5	3 (Volumes 23, 34, 40)	119
James P. LeSage	5	3 (Volumes 18, 30, 37)	271
Christopher F. Parmeter^a	5	4 (Volumes 21, 25, 29, 36)	119
Peter C.B. Phillips	5	5 (Volumes 2, 5, 8, 36, 41)	315
Daniel J. Slottje	5	4 (Volumes 3, 5, 6, 12)	8
Yong Bao	4	3 (Volumes 20, 33, 36)	14
Jane M. Binner	4	2 (Volumes 19, 24)	49
Siddhartha Chib	4	3 (Volumes 11, 16, 23)	161
Peter H. Egger	4	4 (Volumes 29, 37, 42, 43)	14
William Griffiths	4	4 (Volumes 11, 23, 30, 40)	9
Daniel J. Henderson^a	4	4 (Volumes 21, 25, 29, 36)	112
Kim P. Huynh^a	4	3 (Volumes 25, 27, 42)	26
David T. Jacho-Chavez^{a,b}	4	3 (Volumes 25, 27, 42)	26
Ivan Jeliazkov^b	4	4 (Volumes 23, 26, 32, 40)	133
Chihwa Kao	4	3 (Volumes 15, 30, 33)	3,627
Esfandiar Maasoumi	4	4 (Volumes 5, 33, 36, 43)	11
Ron C. Mittelhammer	4	4 (Volumes 11, 18, 29, 30)	46
R. Kelley Pace	4	3 (Volumes 30, 37, 41)	70
Jeffrey S. Racine	4	4 (Volumes 14, 16, 25, 36)	27
Matthew Shum	4	3 (Volumes 31, 34, 41)	39
Liangjun Su	4	3 (Volumes 25, 29, 36)	66
Donggyu Sul	4	3 (Volumes 33, 41, 43)	9
Justin L. Tobias	4	3 (Volumes 21, 23, 40)	51
Tiemens Woutersen^a	4	3 (Volumes 17, 27, 29)	31

^aAIE Fellow.^bSenior Co-Editor.

represented by the papers in this volume in the evolution of welfare economics. The 1,056 citations of this volume, which is the 4th highest total in the series, are evidence of the accomplishment of the stated purpose. In addition, this volume contains two of the top 10 most cited papers in AIE: “On Economic Poverty: A Survey of Aggregate Measures,” by James E. Foster, which ranks 7th with 456 citations, and “On Measurement of Tax Progressivity and Redistributive Effect of Taxes with Applications to Horizontal and Vertical Equity,” by Nanak Kakwani, which ranks 10th with 299 citations.

Volume 4, edited by Professors Basmann and Rhodes in 1985, continued to develop welfare economics. In the introduction to Volume 4, which is titled

“Economic Inequality: Survey Methods and Measurements,” Professor Rhodes talks more about the policy implications than the evolution of the field.

Daniel J. Slottje edited the 5th volume, titled “Innovations in Quantitative Economics: Essays in Honor of Robert L. Basman.” The fifth edition is noteworthy for several reasons. First, this is the first volume that honors a great econometrician, Robert L. Basman, who was co-founder of AIE. More importantly, this was the volume in which R. Carter Hill and Thomas B. Fomby made their first mark on AIE with their article, “Improved Confidence Sets in a Non-Utopian Setting.” Volume 5 is also noteworthy in that five of the fourteen authors with 5 or more AIE publications, as shown in [Table 5](#), have papers in this volume. Volume 5 ranks 5th in total number of papers with 16, as shown in [Table 3](#).

3. THOMAS B. FOMBY BECOMES CO-EDITOR: 1987–1994

A notable change occurred in 1987 when Robert J. Basman decided to step down as co-editor of AIE. Daniel J. Slottje, a former student of Professor Basman’s, was teaching at Southern Methodist University with Thomas B. Fomby. He asked Professor Fomby, an excellent econometrician who had published a paper in Volume 5 of AIE, if he would be interested in working with George F. Rhodes, Jr. as co-editor of AIE. Professor Fomby agreed to replace Professor Basman and remains active as senior co-editor today, 35 years later. When asked what motivated him to become co-editor, Professor Fomby said that he saw the volume as something that could continue well into the future since there was no research annual of this type. In addition, he said that working on a research volume for practitioners was one of the factors that motivated him to become an editor of the volume. He described AIE as a research volume where new methodologies could be quickly disseminated to practitioners and saw it becoming more methods oriented. This became evident in future volumes in diverse areas such as Bayesian methods, time series, marketing, financial econometrics, industrial organization, survey data, and spatial econometrics to name a few. Thomas B. Fomby has recently been recognized by the SMU Board of Trustees with a Career Achievement Award for his leadership and contributions to academia (<https://www.smu.edu/News/2021/Research/Economics-professor-and-campus-leader-Tom-Fomby-receives-SMU-Faculty-Career-Achievement-Award>).

The topic for volume 6 is “Computation and Simulation.” In response to rapid advances in high-speed computers during the 1980s, this volume focused on computer intensive techniques in a variety of econometric areas. This was the first of five volumes edited by Thomas B. Fomby and George F. Rhodes, Jr.

Volume 7 is titled “Nonparametric and Robust Inference.” This volume is divided into three sections: Robust Estimation, Robust Inference, and Consumer Demand Studies. The purpose is to connect microeconomic theories to testing and implementation. There are several well-cited papers throughout the volume although it no longer ranks among the top 10 most cited AIE volumes as it did at the 30th anniversary edition.

Volume 8 is titled “Co-Integration, Spurious Regressions, and Unit Roots.” This volume is divided into four sections: Survey and the New Concept of Multi-Co-Integration, Developments in Testing, Developments in Theory, and Applications. The volume has 1,340 total citations, indicating the quality of the papers and high-level of interest in this topic and ranking 2nd among AIE volumes. Volume 8 includes three of the most cited papers in AIE. The volume begins with a survey paper by Francis X. Diebold and Marc Nerlove, which has generated 277 citations (12th highest in AIE). The most frequently cited paper in the volume is “Testing for Unit Roots and Cointegration by Variable Addition,” by Joon Y. Park, which has been cited 356 times (9th highest in AIE). Other frequently cited papers in the volume include Pierre Perron (69 citations), Peter Schmidt (76 citations) and Bruce E. Hansen and Peter C. B. Phillips (256 citations, 13th highest in AIE history). Finally, volume 8 included the paper “Multicointegration” by Clive W. J. Granger and Tae-Hwy Lee. Clive W. J. Granger (Nobel Prize in Economics, 2003) became the second current or future Nobel Prize winner to publish in AIE (as of this writing). This paper ranks 16th overall with 235 citations. However, 223 of these citations have occurred since 2012 making it the 8th most cited paper over the past decade as shown in [Table 4](#), Panel B.

Volume 9 is titled “Econometric Methods and Models for Industrial Organization.” This is the first of two volumes focused on econometrics of industrial organization. Papers are divided between the econometric analysis of structure and incentives, regulation, and advertising. This topic was chosen in response to changes in the regulatory environment and other areas during the 1980s and early 1990s.

Volume 10, titled “Simulating and Analyzing Industrial Structure,” is the follow-up volume on econometric analysis of industrial organization. Volume 10 is shorter than previous volumes with only 6 contributed papers. Volume 10 was the final volume edited by Thomas B. Fomby and George F. Rhodes, Jr. as it marked the last volume for which founding co-editor George F. Rhodes, Jr. served as co-editor.

4. R. CARTER HILL BECOMES CO-EDITOR: 1996–2001

After George F. Rhodes, Jr. stepped down as co-editor of AIE in 1991, Thomas B. Fomby asked R. Carter Hill, who had published articles in volumes 6 and 7, to join him as co-editor for volume 11. Professors Fomby and Hill had already known each other for 20 years, when they were both students of Stanley R. Johnson at the University of Missouri-Columbia. Starting with the publication of volume 11 of AIE in 1996, they have worked together as co-editors of AIE for another 26 years and counting. Professor Fomby noted that since Professor Hill came aboard as co-editor, they have focused on papers that include more case examples and which have computer code that is fully accessible to readers upon request, allowing for quicker adoption of new methodologies. R. Carter Hill is currently an Emeritus Professor of Econometrics after he retired from Economics Department at LSU in May 2018.

Volume 11, which was split into two parts, focused on Bayesian methods. The popularity of the Bayesian approach to econometrics has grown rapidly due to new computational techniques and fast computers. Markov Chain Monte Carlo (MCMC) techniques have revolutionized the Bayesian methods that were considered impractical. The Gibbs sampling and Metropolis-Hastings algorithms made Bayesian methods very operational. Those applications were the focus of volume 11, part A, entitled “Bayesian Computational Methods and Applications.” Part B, “Bayesian Methods Applied to Time Series Data,” focused on the time series applications of Bayesian methods.

Following the volume on Bayesian methods, Thomas B. Fomby and R. Carter Hill edited volume 12, entitled “Applying Maximum Entropy to Econometric Problems.” In his work “The Bayesian Method of Moments: Theory and Applications,” Arnold Zellner explained entropy as a method that can achieve the goals of Bayesian methods with fewer assumptions.

The objective for volume 13 entitled “Messy Data” and edited by Professors Fomby and Hill was to address the very frequent issue of data problems. The ten articles in this volume tackled several types and levels of “messy” data and ways to diagnose and/or deal with these data problems for time series, cross-sectional and panel platforms.

In 2001, Professors Fomby and Hill edited Volume 14 “Applying Kernel and Nonparametric Estimation to Economic Topics.” The volume had two parts, Methodology and Applications. Volume 14 ranks 9th among AIE volumes with 14 contributed papers. Badi H. Baltagi, whose tremendous contributions to AIE over the coming years would lead to his becoming one of the inaugural AIE Fellows, published his first AIE paper in volume 14. As shown in [Table 5](#), Professor Baltagi has currently authored the most total AIE publications with 14.

Badi H. Baltagi, Thomas B. Fomby, and R. Carter Hill served as editors for Volume 15. This volume, entitled “Nonstationary Panels, Panel Cointegration, and Dynamic Panels,” includes 11 refereed papers on nonstationary and dynamic panel data models written by 20 authors. Within this context, Blundell et al. discuss the estimation of dynamic panel models using generalized method of moments (GMM) in the dynamic error components framework and reports efficiency gains resulting from the system GMM estimator. This paper has received 1,287 citations (4th highest in the AIE series). Jörg Breitung, in the most frequently cited AIE paper (3,219 citations) studies the local power of panel unit root test statistics against a sequence of local power alternatives. Peter Pedroni, in the 2nd most frequently cited AIE paper (3,215 citations), develops methods for estimating and testing hypotheses for cointegrating vectors in dynamic panels. Chihwa Kao and Min-Hsien Chiang, in the 3rd most frequently cited paper (2,698 citations) study the limiting distributions of ordinary least squares (OLS), fully modified OLS (FMOLS) and dynamic OLS (DOLS) estimators in a panel cointegrated regression model, and show that the OLS, FMOLS and DOLS estimators are all asymptotically normally distributed. Finally, co-editor and frequent contributor to AIE, Badi H. Baltagi contributed a joint paper with Chihwa Kao that has been cited 911 times (5th highest). Overall, Volume 15 contains the top five

most cited papers in AIE history. Each of the top three most cited papers have received more citations than any other *volume* of AIE. Due to the substantial number of heavily cited papers, Volume 15 has a total of 11,507 citations, easily the most of any volume in the series with over eight times the number of citations as the next most cited volume. Although it was published in 2001, Volume 15 also contains five of the six most cited papers over the past decade (2012–2022), since the 30th Anniversary Edition.

Philip Hans Franses and Alan L. Montgomery served as guest editors for volume 16, titled “Econometric Models in Marketing.” This volume focused on the application of econometric methods on marketing theory. The most frequently cited paper in volume 16 is “Discrete choice models incorporating revealed preferences and psychometric data,” by Take Morikawa, Moshe Ben-Akiva, and Daniel L. McFadden. This is also noteworthy in that Daniel L. McFadden (Nobel Prize in Economics, 2000) became the third current or future Nobel Prize winner to publish in AIE.

5. ADVANCES IN ECONOMETRICS CONFERENCES HELD AT LSU AND SMU: 2002–2013

Volume 17, “Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later,” edited by Thomas B. Fomby and R. Carter Hill, was in recognition of Halbert L. White’s seminal work in robust estimation. R. Carter Hill had for years wanted to have a small conference where contributors to the volume could present their work and receive strong feedback from other contributors. His goal was realized in November 2002, when authors were finally able to present their work in a unique workshop-type setting on the campus of Louisiana State University. The 1st AIE Conference was held in the Cook Conference Center on the campus of Louisiana State University, and presented papers were published in AIE Volume 17. Halbert White attended the conference and contributed a paper to the volume, as did many other outstanding econometricians. The AIE conference is still going strong with 24 conferences held through 2019. The conference for AIE Volume 43 was originally scheduled for April 2020 in Dallas, Texas. Unfortunately, the Volume 43 conference was cancelled due to Covid-19. Additionally, the conference for the AIE Volume titled “Essays in Honor of Joon Y. Park” was originally to be held in Bloomington, Indiana in September 2020. This AIE conference was postponed due to Covid and is tentatively scheduled for September 2023 as of this writing. [Table 6](#) provides the dates and locations of all AIE conferences that have been held from 2002 to 2022.

The 18th volume was edited by James P. LeSage and R. Kelley Pace and is titled “Spatial and Spatiotemporal Econometrics.” Professors LeSage and Pace contributed a lengthy introduction to the volume, which describes spatial regression models and estimation. The contributed papers from this volume generated 579 total citations, making it the 7th most cited volume in AIE history. Authors presented their research at the 2nd Annual AIE Conference, which was held at

Table 6. AIE Conference Dates, Locations, and Hosts: Conferences 1–24.

Conference	Volume	Conference Date	Location	Host
1	17	November 2002	Baton Rouge, LA	Louisiana State University
2	18	November 2003	Baton Rouge, LA	Louisiana State University
3	20	November 2004	Baton Rouge, LA	Louisiana State University
4	21	October 2005	Dallas, TX	Southern Methodist University
5	22	November 2006	Baton Rouge, LA	Louisiana State University
6	23	November 2007	Baton Rouge, LA	Louisiana State University
7	25	November 2008	Baton Rouge, LA	Louisiana State University
8	26	November 2009	Baton Rouge, LA	Louisiana State University
9	27	October 2010	Dallas, TX	Southern Methodist University
10	28	November 2011	Dallas, TX	Southern Methodist University
11	29	February 2012	Baton Rouge, LA	Louisiana State University
12	30	March 2012	Baton Rouge, LA	Louisiana State University
13	32	November 2012	Dallas, TX	Southern Methodist University
14	31	March 2013	Baton Rouge, LA	Louisiana State University
15	33	November 2013	Dallas, TX	Southern Methodist University
16	34	February 2014	Irvine, CA	University of California, Irvine
17	35	November 2014	Aarhus, Denmark	Aarhus University
18	36	March 2015	Riverside, CA	University of California, Riverside
19	37	October 2015	Baton Rouge, LA	Louisiana State University
20	38	May 2016	Ann Arbor, MI	University of Michigan
21	39	October 2017	Ottawa, Canada	Bank of Canada
22	40	June 2018	Irvine, CA	University of California, Irvine
23	41	October 2018	Baton Rouge, LA	Louisiana State University
24	42	May 2019	Paltinis, Romania	Bank of Romania

Louisiana State University in November 2003. Professors Fomby and Hill continued to serve as series editors, however as seen in Table 1 there were frequent guest editors beginning with volume 18.

The 19th volume was titled “Applications of Artificial Intelligence in Finance and Economics.” Papers for this volume were presented at the 2003 International Conference on Artificial Intelligence in Las Vegas, NV. The volume was edited by Jane M. Binner, Graham Kendall, and Shu-Heng Chen.

Volume 20 of AIE honored the econometric contributions of 2003 Nobel Prize winners Robert F. Engle III and Sir Clive W. J. Granger. The volume is titled “Econometric Analysis of Financial and Economic Time Series.” Milton Dekalb (“Dek”) Terrell joined Thomas B. Fomby as editor for this volume. Volume 20 was split into two parts. To date, part 1 has 310 citations and part 2 has 506 citations, making Volume 20B the 9th most cited volume. Contributed papers in Volume 20 were presented at the 3rd Annual AIE Conference at Louisiana State University in November 2004. Professor Granger was not able to attend the conference as he was in New Zealand at the time. However, he did send a “reflections” piece on his career that was included in the volume. Robert F. Engle III (Nobel Prize in Economics, 2003) attended the conference and gave comments on his career. This piece was also included in the volume, marking the fourth Nobel Prize winner to publish in AIE.

Dek Terrell's involvement in AIE for Volume 20 is noteworthy for two reasons. First, Professor Terrell would later be recognized as one of two inaugural AIE Fellows. In addition, Professor Terrell holds the Freeport-McMoRan Endowed Chair of Economics at Louisiana State University and is Director of the LSU Division of Economic Development. Generous funding provided through the Freeport-McMoRan Chair has been instrumental in allowing the AIE Conference to be held as well as attracting excellent contributors to the volume and conference participants. Additional funding for the AIE conferences held at LSU was provided by the Department of Economics, Department of Agricultural Economics & Agribusiness, Department of Finance, Department of Experimental Statistics, Real Estate Research Institute, and the Ourso Distinguished Chair of Economics held by R. Carter Hill at that time. Subsequent funding by numerous academic departments and donors led to AIE conferences being held at various universities in the United States and internationally as discussed below. Although there are too many to list, the authors want to thank all the donors for their generous support that has allowed for the AIE conferences to be held for the past 20 years.

To examine the impact of the AIE Conferences on the volume, we regress the number of papers per volume on a constant and an indicator variable equal to one for volumes with a conference and zero for volumes without a conference.³ We obtain the following estimated regression:

$$\widehat{\text{Papers}} = 10.889 + 3.747 \text{Conference}, \quad R^2 = 0.1584$$

(s.e.)	(1.04)	(1.40)
--------	--------	--------

The above regression shows that AIE conference (and its funding) is associated with a significant increase in the number of contributed papers. Overall, the average number of contributed papers has increased from 10.9 to 14.6 from the pre-conference to post-conference periods, an increase of 34%.⁴ Admittedly, this regression suffers from omitted variables bias; it is plausible that the growing reputation of the AIE volume has jointly led to more paper submissions and the creation of the AIE conference. Nevertheless, the funding allows the conference host to bring contributors to the conference. In addition, the ability to receive feedback in a workshop setting is attractive to authors. Thus far, LSU has hosted the most AIE conferences with twelve followed by SMU with five. Of the remaining conferences, only University of California, Irvine has hosted more than one conference. Thus far, three conferences have been held outside the United States (see [Table 6](#) for the complete list).

Volume 21, "Modeling and Evaluating Treatment Effects in Econometrics," was edited by Thomas B. Fomby, R. Carter Hill, Daniel L. Millimet, Jeffrey A. Smith, and Edward J. Vytlacil. This volume addressed the estimation of the effects of treatments, which engaged a tremendous amount of discussion during the early 2000s. Thirteen contributing papers and 23 contributing writers had 683 citations (6th highest among AIE volumes) for the volume that covered most key components of the current evolution in this literature. The papers were presented in the 4th Annual AIE Conference in October 2005. This was the first AIE conference to be held at Southern Methodist University.

The subject of Volume 22, entitled “Econometrics and Risk Management,” was credit risk and credit derivatives. This volume, which was edited by Jean-Pierre Fouque, Thomas B. Fomby, and Knut Solna, became especially relevant following the subprime crisis in August 2008. The 5th Annual AIE Conference took place in November 2006 at Cook Conference Center at Louisiana State University.

Following the success of the two parts of Volume 11, Siddhartha Chib, William Griffiths, Gary Koop, and Dek Terrell edited another volume of AIE on “Bayesian Econometrics.” Volume 23 focuses on the scope and diversity of Bayesian applications as well as the Bayesian inference and computations, both on cross-section and time-series data. There are 21 contributed papers in this volume, ranking it as the volume with the highest number of papers (see [Table 3](#)). Volume 23 has 420 total citations, but 326 of these have occurred since 2012 making this the 12th most cited volume during the last decade. Authors presented their papers at the 6th Annual AIE Conference at Louisiana State University in November 2007.

Volume 24, “Measurement Error: Consequences, Applications and Solutions,” was edited by Jane M. Binner, David L. Edgerton, and Thomas Elger. This volume focused on the measurement error in macroeconomic data and its impact on empirical work which impacts macroeconomic policy at the highest levels.

Volume 25 was edited by Qi Li and Jeffrey S. Racine. This volume focused on nonparametric estimation methods. Volume 25 has also been heavily cited in recent years. 349 of its 416 citations have occurred since 2012 making it the 9th most cited volume over the last decade. Authors in the 25th volume presented their research at the 7th Annual AIE Conference at Louisiana State University in November 2008. Volume 25 was also noteworthy in that this was the first volume for which editors selected a paper to receive the Emerald Literati Outstanding Author Contribution award. The first recipients of this award in the AIE series were Yanqin Fan and Sang Soo Park for their paper “Partial Identification of the Distribution of Treatment Effects and Its Confidence Sets.” [Table 7](#) provides a list of recipients of this prestigious award for other AIE volumes.

Volume 26 was titled “Maximum Simulated Likelihood Methods and Applications.” This volume was edited by William H. Greene and R. Carter Hill. Papers appearing in the 26th volume were presented at the 8th Annual AIE Conference at Louisiana State University in November 2009.

Volume 27 was edited by David M. Drukker, then Director of Econometrics at Stata. The volume is titled “Missing Data Methods: Cross-Sectional Methods and Applications.” Due to its length, Volume 27 is split into two parts. Papers appearing in the 27th volume were presented at the 9th Annual AIE Conference at Southern Methodist University in October 2010.

Volume 28, title “DSGE Models in Macroeconomics – Estimation, Evaluation, and New Developments,” was edited by Nathan Balke, Fabio Canova, Fabio Milani, and Mark A. Wynne. Dynamic Stochastic General Equilibrium (DSGE) models combine micro- and macroeconomic theory. This volume was split between papers that examine estimation practice and papers that present new methods in econometric methodology. Papers for

Table 7. Emerald Literati Outstanding Author Contribution Award Recipients.

Author(s)	Paper	Volume	Citations
Yanqin Fan and Sang Soo Park	Partial Identification of the Distribution of Treatment Effects and Its Confidence Sets	25	47
Florian Heiss	The Panel Probit Model: Adaptive Integration on Sparse Grids	26	22
Daniel L. Millimet	The Elephant in the Corner: A Cautionary Tale About Measurement Error in Treatment Effects Models	27A	53
Massimo Guidolin	Markov Switching in Portfolio Choice and Asset Pricing Models: A Survey	27B	27
Denis Tkachenko and Zhongjun Qu	Frequency Domain Analysis of Medium Scale DSGE Models with Application to Smets and Wouters	28	13
Matthew Harding, Carlos Lamarche	Quantile Regression Estimation of Panel Duration Models with Censored Data	29	7
George Judge and Ron Mittelhammer	A Risk Superior Semiparametric Estimator for Overidentified Linear Models	30	2
Federico Echenique, Ivana Komunjer	A Test for Monotone Comparative Statics	31	3
Fabio Canova and Matteo Ciccarelli	Panel Vector Autoregressive Models: A Survey	32	673
Yixiao Sun	Fixed-smoothing Asymptotics and Asymptotic <i>F</i> and <i>t</i> Tests in the Presence of Strong Autocorrelation	33	15
Garland Durham and John Geweke	Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments	34	78
Jens H. E. Christensen and Glenn D. Rudebusch	Modeling Yields at the Zero Lower Bound: Are Shadow Rates the Solution?	35	133
Alexander Chudik, Kamiar Mohaddes	Long-Run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors	36	211
M. Hashem Pesaran, Mehdi Raissi	Fast Simulated Maximum Likelihood Estimation of the Spatial Probit Model Capable of Handling Large Samples	37	32
R. Kelley Pace and James P. LeSage	The Deterrence Effect of Prison: Dynamic Theory and Evidence	38	255
David S. Lee and Justin McCrary	Model-selection Tests for Complex Survey Samples	39	0
Iraj Rahmani, Jeffrey M. Wooldridge	Macroeconomic Nowcasting Using Google Probabilities	40A	69
Gary Koop and Luca Onorante	Identifying Global and National Output and Fiscal Policy Shocks Using a GVAR	41	3
Alexander Chudik, M. Hashem Pesaran, and Kamiar Mohaddes			

Volume 28 were presented at the 10th Annual AIE Conference at Southern Methodist University in November 2011.

Volume 29 consists of “Essays in Honor of Jerry Hausman.” Co-editors for Volume 29 were Badi H. Baltagi and Whitney Newey. Papers for Volume 29 were presented at the 11th Annual AIE Conference at Louisiana State University in February 2012. The conference concluded with a discussion of the papers and remarks by Jerry Hausman himself.

AIE celebrated its 30th year in 2012 with the “30th Anniversary Edition.” In March 2012, Dek Terrell organized the 12th AIE Conference in honor of Thomas B. Fomby and R. Carter Hill. Fellow econometricians around the world showed up to honor these outstanding econometricians whose arduous work and dedication have been instrumental in the tremendous success of the AIE volumes. Dek Terrell and Daniel L. Millimet edited Volume 30, which includes several brilliant econometric papers. The authors of this history were extremely fortunate to have been students of R. Carter Hill and to have gotten to know Thomas B. Fomby and many of the contributors to the volume through the conferences. We were honored to contribute “A History of the Advances in Econometrics Series” to this volume, and to be able to update this history 10 years later.

Volume 31 was titled “Structural Econometric Models” and was edited by Eugene Choo and Matthew Shum. Volume 31 contains sections on structural dynamic models, structural models of games, and applications of structural economic models. Contributors presented their work at the 14th AIE Conference at LSU in March 2013.

Volume 32 was titled “VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims.” Thomas B. Fomby, Lutz Kilian, and Anthony Murphy were co-editors for this volume. Volume 32 was among the most cited volumes with 1,146 total citations, which places it as the 3rd most cited volume overall and 2nd most cited in the last decade. Volume 32 contains the paper “Panel Vector Autoregressive Models: A Survey,” by Fabio Canova and Matteo Ciccarelli, which has 673 total citations ranking 6th among all papers published in AIE. Papers for Volume 32 were presented at the 13th AIE Conference at SMU in November 2012.

Volume 33 consists of “Essays in Honor of Peter C. B. Phillips.” Co-editors for Volume 33 were Yoosoon Chang, Thomas B. Fomby, and Joon Y. Park. Volume 33 contained twenty total papers, which is the 2nd highest total among all AIE volumes. Papers for Volume 33 were presented at the 15th AIE Conference at SMU in November 2013.

6. ADVANCES IN ECONOMETRICS CONFERENCES EXPANSION: 2014-PRESENT

The AIE conferences have further advanced the field of econometrics, allowing a unique forum for authors to share and receive feedback on their research. When

we interviewed Professor Fomby in 2012, he said that he expected the series to become geographically more dispersed and that he “would like to see the conference be held outside the U.S. sometime in the coming years.” The addition of co-editors in 2009 expanded the number of people working on the volumes and led to the AIE conference being held at a variety of locations both inside and outside the United States.

In 2009, Ivan Jeliazkov (University of California, Irvine) joined Professors Fomby and Hill as Senior Co-Editor. Professor Jeliazkov earned his Ph.D. at Washington University in St. Louis. He has published extensively, particularly in Bayesian econometrics and MCMC estimation. Professor Jeliazkov had previously published papers in Volumes 23 and 26 of AIE. Dr Jeliazkov is an international scholar who worked in academic and research institutions around the world including China, Korea, and Australia. He is also the Associate Editor of the *International Journal of Mathematical Modelling and Numerical Optimisation*. Professor Jeliazkov and Dale J. Poirier served as co-editors for Volume 34, titled “Bayesian Model Comparison.” Contributors to Volume 34 presented their research at the 16th AIE conference held at the University of California, Irvine in February 2014. This was the first conference that was not held at either LSU or SMU.

Volume 35 was titled “Dynamic Factor Models” and was edited by Eric Hillebrand and Siem Jan Koopman. Professor Hillebrand was named an AIE Senior Co-Editor in 2012. Papers for Volume 35 were presented at the 17th AIE conference at Aarhus University in November 2014. This conference in Aarhus, Denmark was the first AIE conference held outside the United States. Between 2014 and 2022, the AIE conference was held at nine separate locations and in four countries, bringing to fruition Professor Fomby’s prediction that the series would become international.

Volume 36 consists of “Essays in Honor of Aman Ullah.” Professor Ullah ranks 2nd in AIE publications with nine (see Table 5) and was named an AIE Fellow in 2018. Volume 36, edited by Gloria González-Rivera, R. Carter Hill, and Tae-Hwy Lee, is the only AIE volume in honor of an AIE Fellow thus far. This volume ranks 3rd in number of papers with 19. Contributors presented their research at the 18th AIE conference at the University of California, Riverside in March 2015.

Volume 37 was the second volume on “Spatial and Spatiotemporal Econometrics,” following Volume 18. Badi H. Baltagi served as co-editor along with James P. LeSage and R. Kelley Pace who were also co-editors for Volume 18. Researchers presented at the 19th AIE conference, which returned to LSU in October 2015.

In 2010, Juan Carlos Escanciano (Indiana University) also became a Senior Co-Editor for AIE. Professor Escanciano earned his Ph.D. at Universidad Carlos III de Madrid, Spain. He has published numerous articles on specification tests in semiparametric and nonparametric models and many other areas of econometric theory. Dr Escanciano also served as the Associate Editor of *Economic Bulletin* and is affiliated with numerous international econometric, statistical, and mathematical associations. Professor Escanciano has since

returned to Universidad Carlos III de Madrid as Research Chair and Professor of Economics.

Mattias D. Cattaneo and Juan Carlos Escaniciano served as co-editors for Volume 38, “Regression Discontinuity Design: Theory and Applications.” Volume 38 is noteworthy as it ranks 3rd in citations from 2012–2022 (and 5th overall) with 741 citations. In addition, Volume 38 included the paper “Regression Kink Design: Theory and Practice,” by David E. Card, David S. Lee, Zhuan Pei, and Andrea Weber, which has received 67 citations to date. With this paper, David E. Card (Nobel Prize in Economics, 2021) became the fifth and final (so far) Nobel Prize winner to publish in paper in AIE. Volume 38 also contains two of the ten most cited papers during the last decade, “The Deterrence Effect of Prison: Dynamic Theory and Evidence” by David S. Lee and Justin McCrary (255 citations) and “Party Bias in Union Representation Elections: Testing for Manipulation in the Regression Discontinuity Design When the Running Variable Is Discrete” by Brigham R. Frandsen (192 citations). Papers for this volume were presented at the 20th AIE conference at the University of Michigan in May 2016.

Volume 39 was titled “The Econometrics of Complex Survey Data: Theory and Applications.” Kim Huynk, David Jacho-Chávez (a Senior co-editor and AIE Fellow), and Gautam Tripathi served as co-editors for Volume 39. Contributors presented at the 21st AIE conference in Ottawa, Canada hosted by the Bank of Canada. This marked the 2nd international AIE conference.

Volume 40 was edited by Ivan Jeliazkov and Justin L. Tobias. The volume was titled “Topics in Identification, Limited Dependent Variables, Partial Observability, Experimentation, and Flexible Modeling.” Volume 40 was split into two parts: Part A contains 13 papers with a focus on analysis of time-series and panel data while Part B contains 10 papers and focuses on nonparametric and semiparametric estimation. Papers were presented at the 22nd AIE conference at the University of California, Irvine, which became the first school aside from LSU and SMU to host more than one AIE conference.

Volume 41 consists of “Essays in Honor of Cheng Hsiao” and was edited by Tong Li, M. Hashem Pesaran, and Dek Terrell. Professor Hsiao ranks 5th in total AIE publications with 6 contributed papers. This marked the 5th volume where Professor Terrell served as an editor. The 23rd AIE conference returned to LSU in October 2018.

Volume 42 was titled “The Econometrics of Networks” and was edited by Áureo de Paula, Elie Tamer, and Marcel Voia. Papers were presented at the 24th AIE conference in Paltinis, Romania hosted by the Bank of Romania. This marked the 3rd country outside the United States to host an AIE conference.

Volume 43, titled “Essays in Honor of M. Hashem Pesaran” was edited by Alexander Chudik, Cheng Hsiao, and Allan Timmermann. Papers for this volume were scheduled to be presented at the 25th AIE conference in Dallas, Texas on April 4–5, 2020. Although the conference was cancelled due to Covid-19, Volume 43 was published in January 2022. Volume 43 contains 27 papers and is split into two parts. Part A contains papers related to prediction and macro modeling and Part B contains papers related to micro applications and panel modeling.

Volume 44, titled “Essays in Honor of Fabio Canova” is scheduled for publication in September 2022. Volume 44 is edited by Juan J. Dolado, Luca Gambetti, and Christian Matthes and is also split into Part A, which contains 6 papers, and Part B, which contains 5 papers.

The AIE conference for the present volume, “Essays in Honor of Joon Y. Park,” is scheduled to begin on Friday, September 29, 2023, in Bloomington, Indiana.

7. RECENT DEVELOPMENTS: SENIOR CO-EDITORS AND AIE FELLOWS

During the past 40 years, *Advances in Econometrics* has certainly succeeded in “expanding the use of newly vintaged econometric techniques throughout the empirical economic literature and beyond.” Numerous outstanding econometricians, including five Nobel Prize winners, have published in the volume. Combined, the papers in AIE have been cited over 25,000 times. Professor Fomby stated that AIE has been able to establish a reputation for quality due to the outstanding papers submitted by its authors and their editorial comments and suggestions for improving the publication. Among the almost 900 contributors over the years, we note a few who deserve special mention.

First, the series editors, George F. Rhodes, Jr. and Robert L. Basmann, who started the series, and Thomas B. Fomby and R. Carter Hill who have been Senior Co-Editors for 36 and 27 years, respectively. In addition, to Professors Fomby and Hill, the following individuals currently serve as Senior Co-editors for AIE:

Juan Carlos Escanciano (Universidad Carlos III de Madrid)
Eric Hillebrand (Aarhus University)
David Jacho-Chávez (Emory University)
Ivan Jeliazkov (University of California, Irvine)
Daniel L. Millimet (Southern Methodist University)
Alicia Rambaldi (University of Queensland)
Rodney Strachan (University of Queensland)

In February 2012, the election of *Advances in Econometrics* Fellows was introduced, and the two inaugural AIE fellows were Badi H. Baltagi and M. Dek Terrell. They were followed by Aman Ullah, Tae-Hwy Lee, Christopher F. Parmeter, Daniel J. Henderson, Kim P. Huynh, David Jacho-Chávez, and Tiemen Woutersen. Contributors are eligible to for election as AIE Fellow provided they have:

- i) Served as editor or co-editor for 3 or more volumes of *Advances in Econometrics*, OR
- ii) Contributed 4 papers to the series as author or co-author.
- iii) A contributor who meets (i) or (ii) and is recommended by an AIE co-editor and approved by a unanimous vote of the Board of AIE co-editors will be named an AIE Fellow.

As shown in [Table 5](#), Professor Baltagi of Syracuse University ranks first in total articles contributed to the series with 14 articles, which have appeared in Volumes 14, 15, 17, 18, 29, 30, 33, 36, 37, 41, and 43. In addition, he served as editor for Volume 15, which has the highest number of total citations in the series (see [Table 2](#)). His contributed paper in Volume 15 has the fifth highest number of total citations (see [Table 4](#)). Professor Baltagi also served as co-editor for Volume 29.

Professor Terrell contributed papers to Volumes 23 and 37. He served as editor for Volumes 20A, 20B, 23, 30, and 41. Volume 20B ranks 9th in total citations (see [Table 2](#)). In addition, [Table 3](#) shows that Volume 23 ranks first in total number of papers while Volume 20B ranks in the top 10 with 14 papers. Volume 20A and 20B combined rank first with 27 total papers. In addition, Professor Terrell's generous contributions via the Freeport-McMoRan Endowed Chair of Economics have provided necessary funding for the AIE Conference.

Professor Ullah was awarded his Doctorate in Economics by the Delhi School of Economics at Delhi University. Before he joined the University of California, Riverside in 2008, Ullah completed a post-doctoral fellowship at SMU in 1973. He is currently a distinguished professor at University of California, Riverside. He contributed 9 papers across 8 volumes of AIE. His contributions to volumes 6, 7, 14, 15, 25, 33, 40, and 43 generated a total of 126 citations. *Advances in Econometrics* dedicated Volume 36 to honor Dr Ullah who has extensive work and interest in panel data, non-parametric econometrics, and information theoretic econometrics.

Tae-Hwy Lee was named an AIE fellow in 2018. He has been co-editor of AIE Volume 36: Essays in Honor of A. Ullah in 2017. He contributed papers to Volumes 8, 20, 30, 40, and 43. His 7 contributed papers have received a total of 283 citations. Professor Lee received his Ph.D. in Economics from University of California, San Diego in June 1990 with a thesis entitled "Essays on Multicointegration and Nonlinearity." His thesis committee included Sir Clive W. J. Granger and Halbert White Jr. He is currently Professor of Economics at the University of California, Riverside where he has been employed since 2004. Professor Lee's research topics include forecasting, financial econometrics, and machine learning.

Professor Christopher F. Parmeter was also elected as an AIE fellow in 2018 for his work on Volumes 21, 25, 29, and 36, which have received a total of 119 citations. Professor Parmeter received his PhD from Binghamton University in 2006. He is currently at University of Miami, Department of Economics. His research interests include applied nonparametric econometrics, hedonic modeling, growth empirics, and efficiency analysis.

Daniel J. Henderson contributed to Volumes 21, 25, 29, and 36 where he focused on his interests of nonparametric econometrics, applied micro and panel data econometrics. He earned his Ph.D. in Economics at University of California, Riverside in Spring 2003. Professor Henderson is currently a Professor of Economics at the Department of Economics, Finance and Legal Studies at the University of Alabama where he has been working since Fall 2015.

Tiemen Woutersen was selected as an AIE fellow in 2019. Professor Woutersen received his Ph.D. in Economics from Brown University and is currently a Professor of Economics at University of Arizona. He has done a great deal of work on IV models with heteroskedasticity and many instruments, publishing

numerous articles in top econometrics journals including *Econometrica*, *Journal of Econometrics*, and *Econometric Theory*. He has contributed four articles in AIE in volumes 17, 27, and 29 (2 articles). His sole-authored paper in volume 27, “Consistent Estimation and Orthogonality”, is his most cited work in AIE with 17 citations to date.

The most recent AIE fellows are Kim P. Huynh and David Jacho-Chávez. Professor Huynh received his Ph.D. from Queen’s University at Kingston. He is currently a Senior Research Advisor at the Bank of Canada and Adjunct Research Professor at Emory University. Professor Huynh’s research interest includes applied econometrics, firm dynamics, and industrial economics. Professor Jacho-Chávez received his Ph.D. from the London School of Economics & Political Science. He is currently Professor of Economics and Director of Graduate Studies at Emory University. His research interests include nonparametric and semiparametric estimation methods. Professors Huynh and Jacho-Chávez have co-authored four papers in AIE for volumes 25, 27 (2 papers), and 42. In addition, they served as co-editors with Gautam Tripathi for AIE volume 39. Professor Jacho-Chávez also serves as a Senior Co-editor for AIE. Due to their significant contributions to AIE, professors Huynh and Jacho-Chávez were both elected as AIE fellows in 2020.

AIE celebrates its 40th year in 2022 and, given the continuing contributions of many including the newest Senior Co-Editors and AIE Fellows, will be “expanding the use of newly vintage econometric techniques” for many years to come.

NOTES

1. Citation counts were obtained from Google Scholar (<https://scholar.google.com>) as of July 28, 2022.

2. 25,717 is the citation count for articles. We counted an additional 267 citations for the volume introductions, led by the introduction to Volume 38 which has 63 citations. Including these gives a total of 25,984 citations.

3. We ran the regression using Volumes 1–40.

4. We ran a similar regression with citations as the dependent variable and found the coefficient on *conference* to be insignificant. This is not surprising since the conference is recent and older papers have had more time to be cited. In addition, as shown in Table 2, a few volumes, especially Vol. 15, tend to dominate in terms of citations.

ACKNOWLEDGMENTS

The authors would like to thank Tom Fomby for taking time to talk with us about AIE and for his comments on the paper. We also thank Carter Hill for his comments. Any remaining errors are our own.

REFERENCES

- Baltagi, B. H. (2003). Worldwide institutional and individual rankings in econometrics over the period 1989–1999: An update. *Econometric Theory*, 19, 165–224.
- Campbell, R. C., & Ogunç, A. (2012). A history of the advances in econometrics series. *Advances in Econometrics*, 30, 3–24.