How Transparency and Reproducibility Can Increase Credibility in Policy Analysis: A Case Study of the Minimum Wage Policy Estimate

Fernando Hoces de la Guardia

Dissertation

Pardee RAND Graduate School

Committee:

Kristine M. Brown (Chair)

Susan M. Paddock

Chapin White

Edward A. Miguel (Outside Reader)

Final Manuscript: May 9, 2017

Abstract

The analysis of public policies, even when performed by the best non-partisan agencies, often lacks credibility (Manski, 2013). This allows policy makers to cherry-pick between reports, or within a specific report, to select estimates that better match their beliefs. For example, in 2014 the Congressional Budget Office (CBO) produced a report on the effects of raising the minimum wage that was cited both by opponents and supporters of the policy, with each side accepting as credible only partial elements of the report. Lack of transparency and reproducibility (TR) in a policy report implies that its credibility relies on the reputation of the authors, and their organizations, instead of on a critical appraisal of the analysis.

This dissertation translates to policy analysis solutions developed to address the lack of credibility in a different setting: the reproducibility crisis in science. I adapt the Transparency and Openness Promotion (TOP) guidelines (Nosek et al, 2015) to the policy analysis setting. The highest standards from the adapted guidelines involve the use of two key tools: dynamic documents that combine all elements of an analysis in one place, and open source version control (git). I then implement these high standards in a case study of the CBO report mentioned above, and present the complete analysis in the form of an open-source dynamic document. In addition to increasing the credibility of the case study analysis, this methodology brings attention to several components of the policy analysis that have been traditionally overlooked in academic research, for example the distribution of the losses used to pay for the increase in wages. Increasing our knowledge in these overlooked areas may prove most valuable to an evidence-based policy debate.

Contents

1	Intr	roduction	8
	1.1	Definition Of The Problem	8
	1.2	Relevance	10
	1.3	Research Objectives	12
	1.4	Structure of the Dissertation	13
2	Bac	kground	15
	2.1	Transparency and Reproducibility in Research: Lessons For Policy Analysis	16
		2.1.1 The Reproducibility Crisis in Science and Why It Matters for Policy Analysis?	16
		2.1.1.1 The crisis	16
		2.1.1.2 The response: best practices for Open Science	18
	2.2	A General Framework For The Role Of Policy Analysis in Connecting Research And	
		Policy	20
		2.2.1 Manski's Critique And The Role of Low Transparency and Reproducibility .	22
		2.2.1.1 Policy makers can cherry-pick the facts	23

			2.2.1.2	Little knowledge exists regarding how research affects policy estimates	27
			2.2.1.3	Hard to automate or systematically improve policy reports	29
	2.3	Descri	ption of tl	he case study	29
3	Met	hods			33
	3.1	Adapt	TOP Gui	idelines to Policy Analysis	34
		3.1.1	Three sta	andards of TR in Policy Analysis	36
			3.1.1.1	Workflow	36
			3.1.1.2	Data	38
			3.1.1.3	Methods	38
	3.2	Apply	Guideline	es to the CBO Report and Reproduce it Using the Highest Standards	
		of TR			40
		3.2.1	A Dynan	nic Document for CBO's report	40
		3.2.2	Distribut	ted Version Control (Git)	41
		3.2.3	Difference	es Between Reproducing Research and Policy Analysis	42
	3.3	Sensiti	vity Anal	ysis	42
4	Resi	ults			45
	4.1	Descri	ption of tl	he Dynamic Document for the CBO Report	45
	4.2	Assess	ing TR in	Original CBO Report and in the DD	50
		4.2.1	Standard	l #1: Workflow	50
		4.2.2	Standard	#2: Data	58

		4.2.3 Standard #3: Methods	59
5	Sen	sitivity Analysis	61
	5.1	Sensitivity of a Few Parameters for all Policy Estimates	62
	5.2	Sensitivity of all Parameters on a Sample Welfare Function	64
6	Con	nclusions and Extensions	71
	6.1	Conclusions	71
	6.2	Extensions within the current case study (minimum wage)	73
	6.3	General extensions to TR in policy analysis	74
	6.4	Extensions to other possible policy analyses	75
7	Apj	pendices	77
	7.1	Summary of TOP Guidelines	78
	7.2	Checklist for TR in Policy Analysis	79
	7.3	Equations that define Model in CBO Report	81
		7.3.1 Employment	81
		7.3.2 Wages	83
		7.3.3 Income	84
	7.4	Key Definitions	86
	75	Printed version of the dynamic document	87

Evil is relatively rare, ignorance is an epidemic.

JON STEWART, 2014

Sunlight is said to be the best of disinfectants.

LOUIS BRANDEIS, 1914

Acknowledgments:

I would like to thank to Kristine Brown, my doctoral advisor, who provided essential guidance, shared my enthusiasm for finding rigorous ways to connect policy and research, and constantly reminded me what my contribution was and why it matters. Edward Miguel, and his team at the Berkeley Initiative for Transparency in Social Sciences, played a similar role. Every time I thought of my work was a misfired moonshot, Ted and BITSS provided me with with essential feedback and a rich research environment around which to orbit.

I am also grateful to the other members of my committee, including Susan Paddock, whose line-by-line comments on the first draft of this manuscript helped improve its quality and organization significantly, and Chapin White, whose extensive experience doing policy analysis and shared enthusiasm toward increasing its transparency and reproducibility, helped to organize this project and orient its way forward. I would also like to thank Sean Grant as an honorary committee member. His enthusiasm about open science plus his expertise on guideline development was fundamental in shaping up the structure of my work.

Through the whole dissertation process I benefited from multiple conversations that helped me achieve this final work. I would like to acknowledge and say thank you to: Phillip Armour, Robert Brook, Mike Cui, Cheryl Damberg, Beth Ann Griffin, Italo Gutierrez, Dan Han, Andres Hojman, Jessica Hwang, Emmett Keeler, Stephanie Kovalchik, Kenneth Kuhn, Jodi Liu, Italo Lopez, Kathleen Mullen, Jeanne Ringel, Carolyn Rutter, Kanaka Shetty, and Timothy Smith, among others.

Finally, this dissertation would have not been possible without the extraordinary support of Erin Taylor, my wife. Extraordinary is an appropriate qualifier, but it hides how multidimensional her support was. She provided insightful comments on my dissertation throughout, worked as official copy-editor, was my financial backer, and on top of that, managed to absorb much more than her fair share of childcare responsibilities at home for long stretches of time.

Chapter 1

Introduction

1.1 Definition Of The Problem

Policy analysis aims to inform policy makers and stakeholders with an empirical assessment of consequences associated with a set of policy alternatives (Weimer and Vining, 2015; Williams, 1971; Dunn, 2015). Its purpose is to connect general purpose evidence with concrete policies, playing a fundamental link in the evidence-based policy world. Even before the term "evidence-based policy" existed, experts had been providing consequential information to policy makers. Among the earliest examples of such reports is the famous Einstein-Szilard letter sent in 1939 to then-US-president F.D. Roosevelt, assessing the consequences of new developments in Physics and its consequences for the ongoing conflict with Germany. That letter triggered a chain of policy responses that culminated in the Manhattan Project and the development of the first atomic bomb (Hewlett and Anderson, 1962). Roosevelt's response was based on the information provided (the policy analysis), and on the credibility of the analysis.

Today, scholars have noticed that the credibility of policy analysis is often low (Manski, 2013). With low credibility, different reports can be treated as equally valid. Parties differing in a normative dimension (their values or beliefs) of a policy issue tend to choose empirical evidence from different reports that systematically align well with the normative principles (Wesselink et al (2013) label the phenomena "report wars"). Even if an analysis is perceived as non-partisan, low credibility of

the whole report allows for selective validation of only components of the report that matches the readers' normative positions.

In an interview conducted in 2016, Douglas Elmendorf, former director of the Congressional Budget Office (CBO), provided a comment that illustrates this issue:

"When I was director of the CBO, I was very frustrated when we would write a policy report [saying] a certain policy would have these two advantages and these two disadvantages, and the advocates would quote only the part about the advantages, and the opponents would quote only the part about the disadvantages. That encourages the view that there are simple answers. There aren't generally simple answers. There are trade-offs." (Harvard Magazine, 2016)

When high credibility cannot be obtained by directly assessing the intrinsic quality of a report, the reader has to rely on reputation as a proxy for quality. This is consistent with the evidence found by Doberstein (2016) where the same policy brief receives drastically different credibility scores, depending on the reputation of the authoring institution placed on the cover. The problem of low credibility in policy analysis is important because the output of these analyses is among the main inputs that policy makers use as evidence when designing policy. (Nutley et al, 2007).

Manski (2013) describes the practice of policy analysis as one dominated by "Incredible Certitudes": the ubiquitous practice of reporting policy estimates with no uncertainty and strong but undisclosed assumptions. By developing a typology of analyses that presents policy estimates with little or no uncertainty, Manski describes examples of how even the most reputable agencies and research organizations performing policy analysis lack credibility. Manski suggests a menu of methodological improvements to policy analysis. In increasing order of desirability, these improvements are: (1) display standard errors; (2) bound estimated effects; (3) add [policy] decision criteria to the analysis (best).

An underlying theme in Manski's typology is that of characterizing the policy analysis process as a black box, where little is known about how each component of the analysis affects the final results. Using this underlying theme, the problem of low credibility can be understood as one of low transparency and reproducibility (TR).

Statement of the problem: There is low transparency and reproducibility in policy analysis, damaging the credibility of the empirical analysis behind it and its conclusions.

Lack of transparency means that it is hard or impossible to understand all of the details of a policy analysis, and lack of reproducibility means that is hard or impossible to obtain the same results using the same data, methodology and code.

Manski showed that incredible certitudes are a widespread problem in policy analysis by examining some of the most rigorous analyses available, produced by the most reputable organizations (the Congressional Budget Office, RAND, and others). The underlying rationale is that if the best reports suffer from a credibility problem, then we should expect to find similar, or worse, problems among less rigorous analyses¹. Using this same approach, this dissertation examines a case study that used among the best methodological practices and reporting standards in policy analysis. A detail examination of this analysis is provided with the purpose of illustrating a widespread challenge in policy analysis.

1.2 Relevance

The deficit in TR in policy analysis harms public policy for three reasons.

First, lack of TR implies that trust in policy analysis is based on reputation as opposed to sound reasoning. When discussing the Congressional Budget Office, Manski comments "[...]I worry that someday sooner or later the existing social contract to take CBO scores at face value will break down. Conventional certitudes that lack foundation cannot last indefinitely." (Manski, 2013, page 20) Low TR, or low credibility, in non-partisan estimates makes it easier for different parties to cherry-pick their facts from less neutral policy reports, generating the aforementioned "report wars" (Wesselink et al, 2013). In turn, overt cherry-picking provides a fertile ground for the surge of demagoguery and a general disregard for a scientific approach to policy.

¹Examples of less rigorous policy analysis are the plethora of cases where a law, regulation or policy is discussed based on an empirical estimate that is briefly -or obscurely- described, but nonetheless taken as a fact. Manski described two good examples when discussing the concept of dueling certitues: the estimated fiscal effects of the Affordable Care Act by CBO over a ten year period. One report by CBO and the Joint Committee on Taxation predicted a reduction in the deficit over a 10-year timeline of \$138 billions (US Congressional Budget Office, 2010). Days after its publication a former director of CBO wrote in a newspaper article that the deficit would increase in \$562 billions (Holtz-Eakin, 2010). In either case, almost no detail was provided on how those estimates were obtained.

Second, it makes it difficult to understand precisely how research affects the estimates produced by policy analysis (policy estimates hereafter). Research generates evidence that feeds into policy analysis. But little is known about how the emergence of new evidence might affect the final policy estimates (Vivalt, 2015; Nutley et al, 2007). If we were to understand how the current research is used in policy analysis, we could accurately assess the potential effect of new research on policy estimates. With this information, research resources could be allocated based on where the value of additional information is the highest, and researchers could identify the biggest gaps in knowledge related to specific policies (Snilstveit et al, 2013).

Third, lack of TR prevents automation and/or systematic improvements of reports. Translating evidence from general research into specific policy estimates is not an easy task. It requires modeling how multiple agents will react to a specific policy, assumptions regarding generalizability, estimation of key socio-demographic parameters for the target population, and other contextual elements (like relative prices of components in the model or rates needed to discount effects over time). All of these processes are usually encapsulated under the task of generating a micro-simulation study. Micro-simulation studies require a large amount of highly skilled labor, which tends to be expensive, and involve a large number of somewhat arbitrary decisions that need to be made under heavy time constraints. With high levels of TR, redoing this analysis or performing similar ones can be substantially less expensive. With the current lack of TR in policy analysis, such savings cannot be realized. Moreover, there is no reason to think that the somewhat arbitrary assumptions chosen for one report will be consistent with ones chosen for future versions of similar reports.

Increasing TR will ameliorate these issues across most policy discussions. Even in a hyper-partisan environment (popular examples in the current US policy discussions are healthcare reform and climate change) the framework developed here aspires to identify a minimal set of facts and to clearly identify the source of technical differences. With this framework it can still be the case that no agreement is reached for any facts, but even in this case the new framework would help to (i) identify such cases, and (ii) keep track of the precise differences between different stakeholders.

To increase the TR of policy analysis, this dissertation draws a connection between Manski's credibility critique and the response to the reproducibility crisis in science.

1.3 Research Objectives

As a complement to Manski's prescriptions to increase the credibility of policy analysis, this dissertation proposes a systematic approach to increase TR in policy analysis. This approach defines the three study objectives and their respective research questions:

Research Objective #1 Translate guidelines developed to increase TR in science to the
policy analysis setting.

Description: Summarize the current state of knowledge on transparency and reproducibility in science, and draw parallels to policy analysis. The common solutions are grouped into two categories: guidelines and tools. The most widely adopted guidelines in science are described and translated to the policy analysis setting. To achieve the highest standards of TR (both in science and policy analysis), new tools, like dynamic documents and open source version control (git), are described.

• Research Objective #2 Apply state of the art practices of TR to one policy analysis and develop a report that presents the elements involved in the analysis in a transparent and reproducible fashion.

Description: The draft guidelines are implemented in the case study of a minimum wage policy analysis. First, a TR assessment is briefly discussed for the current state of the report. Second, I implement the highest standards from the adapted guidelines to reproduce the analysis of the case study, and present it in the form of an open-source dynamic document. This document is compared to the original report to demonstrate the gains offered by TR.

• Research Objective #3 Using Sensitivity Analysis, identify the components of the policy analysis where additional research/knowledge is most informative.

Description: In addition to increasing the credibility of the case study analysis, this methodology brings attention to several components of the policy analysis that have been traditionally overlooked in academic research. The fully reproducible report allows for a comprehensive sensitivity analysis showing how several components of the analysis are as, or more important, than the much-investigated elasticity of labor demand, for example, the distribution of the losses used to pay for the increase in wages. Increasing our knowledge in these overlooked areas may prove most valuable to an evidence-based policy debate.

The case study used in here is the report produced in 2014 by the Congressional Budget Office (CBO) estimating the effects on employment and income of a potential raise in the federal minimum wage from \$7.25 per hour to \$10.10². The report estimated positive effects due to wage gains and negative effects due to job losses and income losses from those paying for the wage gains (US Congressional Budget Office, 2014). Supporters of the policy took the positive effects as given, and questioned heavily some of its negative effects (White House, 2014). Conversely, opponents of a raise in the minimum wage took the negative effects at face value, while neglecting to acknowledge its benefits (Smith et al, 2014). This example of selective reading of the analysis is common practice in policy, and is supported by the lack of credibility in policy analysis described previously.

Three criteria were used to select the case study: (i) feasibility, measured by access to data, codes, very detailed documentation or contact information of experts willing to discuss details of their approach; (ii) scalability, or the extent to which the lessons from the case study can be extrapolated to other policy analyses; and (iii) recurrence, defined as the likelihood that a similar analysis will be needed again in the future. The selected case study must have a high score in all three dimensions.

The CBO report meets those criteria. Even though CBO could not provide access to further documentation, the report explains in detail the performed analysis, and the analysis is based on publicly available data. The minimum wage debate is a well known policy issue, with a mature body of research behind it and strongly opposing views from policy makers. All of these elements can be found in other policy issues and make the results from the case study more generalizable. Finally, this issue has shown some recurrence in the policy debate and it is likely that an analysis, similar to the one already conducted, will be required again in the near future.

1.4 Structure of the Dissertation

This dissertation is presented in a monograph format. This first chapter described the problem, its relevance and the research questions. The second chapter provides background information: a general framework to understand the problem of low TR in policy analysis, a summary of the reproducibility crisis in science and its proposed solutions, and a description of the case study on the minimum wage policy estimates. The third chapter describes the adapted guidelines and

²The report explored two policy alternatives: \$10.10 indexed to inflation, and \$9 without indexation. For simplicity only the \$10.10 option is described, but the analysis of the dissertation applies to both.

the tools required for its implementation to the highest degree of TR. Chapter four describes the implementation of the guidelines to the case study. Chapter five presents the results of a sensitivity analysis performed to identify gaps in knowledge from a policy perspective. Finally, chapter six concludes by summarizing the contribution of this dissertation, its limitations and extensions for future research.

Chapter 2

Background

Contents				
2.1	Tran	nsparency	y and Reproducibility in Research: Lessons For Policy	
	Ana	lysis		16
	2.1.1	The Rep	roducibility Crisis in Science and Why It Matters for Policy Anal-	
		ysis?		16
		2.1.1.1	The crisis	16
		2.1.1.2	The response: best practices for Open Science	18
2.2	A G	eneral Fr	amework For The Role Of Policy Analysis in Connecting	
	Rese	earch An	d Policy	20
	2.2.1	Manski's	Critique And The Role of Low Transparency and Reproducibility	22
		2.2.1.1	Policy makers can cherry-pick the facts	23
		2.2.1.2	Little knowledge exists regarding how research affects policy es-	
			timates	27
		2.2.1.3	Hard to automate or systematically improve policy reports	29
2.3	Desc	cription o	of the case study	29

2.1 Transparency and Reproducibility in Research: Lessons For Policy Analysis

2.1.1 The Reproducibility Crisis in Science and Why It Matters for Policy Analysis?

2.1.1.1 The crisis

More than four decades ago, social scientists from different fields identified how the lack of transparency threatened the validity of scientific results. By not disclosing most of the research decisions made in studies, two problems emerge: First, Rosenthal (1979) identified the "File Drawer Problem", currently known as Publication Bias, where only the studies with strong results tend to get published. Second, Leamer (1983) suggests that the inability to observe all of the model specifications tested by the researcher invalidates most of the conclusions presented in empirical work. Both problems could only be solved by opening up the entire scientific process, which means providing access to the original raw data and detailed instructions for all of the analyses conducted as part of the study. Unfortunately these critiques were largely ignored by the scientific community for years.

In the last decade, research transparency has become an active field across multiple disciplines. Ioannidis (2005) provides a probabilistic argument for why most published research is false. Franco et al (2014) assess that among a sample of high quality awarded research proposals in social sciences, the majority of those that find null results are never written, and the likelihood of publication increased dramatically with the strength of the results. Gerber et al (2008) showed that in top psychology and political science journals the number of papers with p-values just below 0.05 were more frequent than those just above 0.05 by a factor of two and three respectively. Brodeur et al (2016) found similar irregular behavior for the distribution of p-values in top economic journals.

Parallel to this work and in similar fashion, multiple disciplines began to assess issues of replicability and reproducibility of previously published research. Replicability tests whether or not the same results could be obtained in a different setting (data) using the same procedures (methodology). Reproducibility tests if it is possible for a third party to obtain the same results using the same

data, methods and code¹. Replicability has been part of the scientific method for centuries, and reproducibility has become increasingly important with the predominance of computation in empirical work.

A large scale replication effort in Psychology attempted to replicate the results of 100 studies. They were able to replicate about 40 of them, but for 30 studies the evidence was inconclusive and 30 studies failed to replicate (Collaboration et al, 2015). In a similar exercise for Behavioral Economics, 11 studies were replicated out of a total of 18 (Camerer et al, 2016). Regarding reproducibility (same data, code, methods) Peng (2011) and Stodden et al (2016) describe the importance of improving current standards for computational science. In an exercise to assess the reproducibility of 67 papers in Macroeconomics, Chang and Li (2015) were able to obtain qualitatively similar results for 29 (and 6 papers could not provide proprietary data).

The issues described above are a subset of what some authors refer to as the reproducibility crisis (Baker, 2016), or the credibility crisis (Stodden, 2014), in science. From now on I will use the label of *open science* to refer to any effort aimed at documenting or improving the issues of transparency, replicability and reproducibility in science as described above.

One emblematic case where more open science could have saved billions of dollars is the initial studies that analyzed the effects of the drug Tamiflu, which is used to treat seasonal flu. Initial evidence suggested that Tamiflu could ameliorate several harms related to the flu, and the governments of the US and the UK bought more than \$20 billion in stockpiles. However once more researchers were able to access the original data and to review other studies that had been ignored in the initial review of the literature, all the benefits from Tamiflu disappeared. The British Medical Journal (BMJ) has emphasized that these financial resources would have been saved if protocols for TR had been in place (Abbasi, 2014).

This is one example for one scientific discipline, but across fields different types of initiatives are taking place. The BMJ and other top medical journals have adopted new standards for publication, which has increased TR substantially. In the social sciences, initiatives like the Center For Open Sciences (COS) and the Berkeley Initiative for Transparency in the Social Sciences (BITSS) have created tools, conducted research and gathered resources to promote TR. Efforts in this direction

 $^{^{1}}$ It is worth noting that there are many definitions of reproducibility and replicability. See Goodman et al (2016); Clemens (2015b)

have also found strong support from the fast-growing field of Data Science. Both in the public and private sectors, this field has made a hallmark of their work the promotion of open access to code, analysis, and data (Peng, 2011).

To summarize, at a high level there are two general problems that open science has identified. The first is that sharing only a small fraction of knowledge generated in a study² is in direct contradiction with the most basic goals of science (Merton, 1973). Second, the lack of disclosure allows researchers to, consciously or not, make discretionary choices that maximize the chances of obtaining desired results. Simmons et al (2011) label this last problem the issue of researcher degrees of freedom.

2.1.1.2 The response: Best practices for Open Science

Miguel et al (2014) argue for three guiding norms to promote transparency in social sciences: (i) disclosure of key details involved in the analysis and collection of the data; (ii) registration of pre-analysis plans that contain information on the outcome variable, independent variable(s) of interest, model specifications and other analytic choices before the data is collected; and (iii) open access to data, code, and additional documentation. Nosek et al (2015) operationalize these norms into a set of guidelines that identify different levels of compliance.

To apply the norms described above, Nosek et al (2015) developed the the Transparency and Openness Promotion (TOP) guidelines. These guidelines group the problems described in the previous section into eight dimensions, and for each dimension different degrees of compliance are identified.

The eight dimensions or standards are³

- 1. Citation Standards: the definition of a unit of scholarly work to cite should include, in addition to peer-reviewed articles, data and code generated by others.
- 2. **Data Transparency:** How accessible is the data. From all raw data to final analytic files, documentation and access should be clearly classified as accessible from the authors or from

²Stodden (2014) presents a quote from Buckheit and Donoho (1995) that describes the problem and is worth repeating here: "The idea is: An article about computational science in a scientific publication is not the scholar-ship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures."

³More details can be found at https://osf.io/9f6gx/ and a summary table is provided in the appendix

third parties. When access to sections of the data is not possible for propietary or sensitive reasons, those sections should be clearly labeled.

- 3. Analytic Methods (Code) Transparency: Same as data.
- 4. Research Materials Transparency: Same as data. The reason why dimensions 2-4 are presented separately; although they are assessed by same rubric, is that research can comply differently across these three dimensions.
- 5. Design and Analysis Transparency: Clear disclosure to readers and reviewers of the reporting guidelines (structure of the paper) that the paper is following. Most reporting guidelines can be found at the EQUATOR network (http://www.equator-network.org/), but if no guideline fits the specific type of study performed in a paper, then default to field-specific guidelines.
- Preregistration of Studies: Pre-registration of study design, sample, and principal outcomes and independent variables of interest.
- 7. Preregistration of Analysis Plans: Pre-registration of detailed analytic approach.
- 8. **Replication:** Journals incorporate replications as a component in the scientific output, encouraging their submission for publication.

These standards are promoted across the scientific community, but the key enforcers are funders and journals. Compliance with each standard is defined in four different levels. Level 0 (lowest) represents the status quo where not much is mentioned (by journals or funders) for a given standard. Level 1 is achieved when authors of the research are required to disclose how much are they following the standards. Level 2 is attained when the journal can provide a verification that standards are being followed. And level 3 (highest) is met when some third party (different from authors or journals) can guarantee that the authors comply with a standard up to the highest degree.

For further detail on how to achieve the highest level and follow best practices, Christensen and Soderberg (2015) provide a manual for best practices in research transparency, and Kitzes, J., Turek, D., and Deniz, F. (Eds.) (2017) present a set of 31 case studies of computational reproducible research.

The lessons from the reproducibility (or credibility) crisis in science, and its response, can be applied to the policy analysis setting in three areas. First, it has brought to light core scientific principles that encourage replication, reproduction and openness (Merton, 1973) and it has also shed light on how the scientific community simultaneously accepts these principles but does not practice them on a regular basis (Anderson et al, 2007). Second, the response articulated in the TOP guidelines will be a key starting point to promote these ideas in policy analysis. And third, the response provided several key concepts that help to describe similar problems found in policy analysis.

Policy analysis and science have different goals but share key similarities. Both seek to use the best knowledge available to the advancement of society, and both use knowledge as their main input. Science uses knowledge as an input to generate more knowledge. Policy analysis uses knowledge to generate condensed information designed to brief policy makers. Hence, the lack of TR harms both Science and Policy Analysis, but only the consequences of the former are currently documented. I now examine how the scientific community is addressing these problems and in subsequent chapters extract lessons for the practice of policy analysis.

In the next section I develop a conceptual framework that describes in more detail the connection between research, policy analysis, and policy making. Within this framework I discuss which elements of open science can be translated into policy analysis, arguing for an emphasis on transparency and reproducibility.

2.2 A GENERAL FRAMEWORK FOR THE ROLE OF POLICY ANALYSIS IN CONNECTING RESEARCH AND POLICY

Across the multiple definitions of policy analysis⁴, the common denominator is the role that it plays in synthesizing research and other sources of empirical information to brief policy makers about the consequences of a particular policy issue. This section presents a model developed to discuss the role that policy analysis plays in connecting evidence with policy. Figure 2.2.1.

⁴For example: "Policy analysis is client-oriented advice to public decision and informed by social values" Weimer and Vining (2015). "Policy analysis is a means of synthesizing information including research results to produce a format for policy decision" Williams (1971). "Policy analysis is an applied social science discipline which uses multiple methods of inquiry and argument to produce and transform policy relevant information that may be utilized in political settings to resolve policy problems" Dunn (2015)

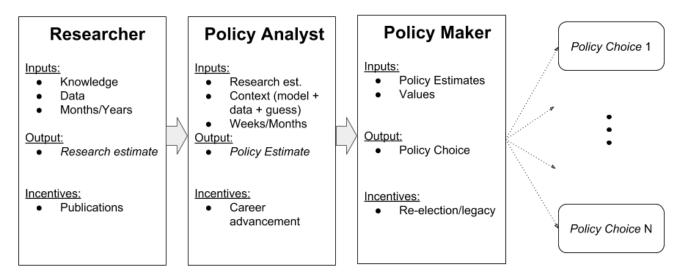


Figure 2.2.1: A simple model for the connection between evidence and policy

In this model the role of policy analysis is to synthesize general purpose research to a specific context related to a particular policy issue. This condensed information about all the empirical (or positive) aspects of the policy issue is provided to the policy maker who combines this information with her own normative beliefs to make a choice among the alternatives regarding the policy issue.

There are three agents in the model: researchers, policy analysts and policy makers.

- Researchers produce knowledge. Throughout this document I will refer to their output, either a new piece of knowledge or a synthesis of previous work, as a research estimate. They rely mainly on previous research, relevant data, and significant time as resources (months or years). Researchers' key incentive is to maximize their output, commonly measured by peer reviewed publications.
- Policy analysts produce quantification of costs, benefits and/or distributional effects of a specific policy. From now on I will refer to their output as *policy estimates*. The production of policy estimates happens in a much shorter time frame than that of the researcher (weeks or a few months). Good examples of policy analyses range in complexity from back-of-the envelope calculation (GiveWell, 2015), to micro-simulation models (US Congressional Budget Office, 2014; Agency for Healthcare Research and Quality, 2015), to dynamic choice models (Rothstein, 2015). In addition to the modeling technique, key inputs of the policy analysis

are research estimates and specific characteristics of the population affected by each potential policy. I assume that policy analysts' key incentive is to maximize career advancement possibilities, which I assume depends directly on the amount of policy-relevant information produced given the time constraint⁵

• Policy makers' output is the support of a specific policy alternative, labeled from now on as a policy choice. The two inputs used to produce a policy choice are the policy estimates, as a representation of the empirical or positive information (the facts), and the policy maker's personal values. In this last category I include everything that is not commonly accepted as empirical information. This includes how different policy estimates are weighed across different populations and any additional effect (benefit or costs) that one policy maker sees as valid to the exclusion of others. Following standard social choice theory (Mas-Colell et al, 1995) I assume that the policy maker's goal is to maximize chances of career advancement (e.g. re-election, promotion to more influential positions, legacy).

This model is a highly simplified version of the connection between evidence and policy. It is used here only to study the role of policy analysis, as the sythesizer/producer of commonly agreed empirical effects of a specific policy. It abstracts from direct links between researchers and policy makers, from policy analysis designed to advocate for a specific policy, and many other elements of the link between evidence and policy. For an extensive survey of models that explore other dimensions of the connection between research and policy, see Nutley et al (2007).

2.2.1 Manski's Critique And The Role of Low Transparency and Reproducibility

Using this model I draw a parallel between Manski's critique of the credibility of policy analysis (Manski, 2013), and the problems that arise from low levels of transparency and reproducibility (TR).

⁵ I define policy relevant information as a policy estimate that increases the relative merits of a particular policy alternative, that passes basic reality checks (do all the proportions add to a 100%?, are predictions within a sensible range?, etc.), and that appears to have good predictive power. There are two important dimensions that qualify this type of information: ambiguity or the degree to which a statement can be falsified; and an uncertainty dimension, particularly how it is acknowledged in the analysis. Ambiguity can increase the perceived policy relevancy of a report, but does not have any intrinsic informational content (i.e. cheap talk). Properly reported uncertainty, on the other hand, is an essential input for good decision making (Tetlock and Gardner, 2015).

Manski's central argument is that policy analysis, like any empirical analysis, is a combination of data and assumptions, and the results of policy analysis are usually framed in terms of complete certainty. Such certainty can only be based on very strong assumptions that are typically not disclosed, making the results not credible.

In terms of figure 2.2.1, what Manski defines as assumptions I separate into inputs from research (research estimates), modeling decisions, and additional guess work needed to complete a policy analysis (more on the role of guess work in section 2.2.1.3). Lack of disclosure of assumptions, in Manski's terminology, is analogous to the lack of transparency discussed above for the case of scientific research. Lack of reproducibility prevents us from understanding how sensitive the final results are to the different inputs used in the analysis, producing what Manski labels *Incredible Certitudes*.

With this framework it is possible to describe three reasons why the problem low transparency and reproducibility harms the credibility of policy analysis: it is easier for policy makers to cherry-pick evidence; the effects of new research on policy analysis are unknown; and it hinders automation or a systematic approach to recurrent policy analyses.

2.2.1.1 Policy makers can cherry-pick the facts

Policy makers and stakeholders debate policy issues often based on different empirical policy reports (for the same issue). This is the phenomena that Wesselink et al (2013) labeled as "report wars". There are two troubling aspects of this phenomena worth highlighting: (i) the multiplicity of *empirical* reports, and (ii) how predictably the "chosen" report (in quotations as strictly speaking there should be no choice regarding empirical issues) by a group of stakeholders makes their preferred alternative seem like the dominant one.

Underlying this multiplicity of reports are low levels of TR. To explain why, I start from a simplified example, based on figure 2.2.1 where there is high TR both in research and in policy analysis. In figure 2.2.2 there is consensus on what research is the best representation of some specific phenomena (truth), and what policy analysis is the best representation of gains and losses associated with a specific policy issue (using research as one of the inputs). In this ideal context different policy makers start from a commonly shared set of evidence and make their choices based on a combination of

evidence and values or beliefs. The common set of facts allows for citizens/voters to learn something about each policy makers' values by observing their policy choice.

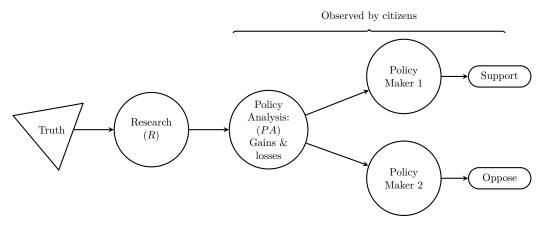


Figure 2.2.2: Policy-making with high TR/credibility in research and policy analysis

Low TR in policy analysis allows for multiple policy reports to compete as the best representation of the same empirical consequences of a policy issue. Or equivalently, even when there is one report that stands out as more objective, policy makers can choose to grant credibility only to parts of the evidence presented in one such report. Representatives of those who lose under the policy issue under study might accept as credible only its negative effects as valid, while winners could give credibility to the gains. This latter phenomena was mentioned in the introduction when quoting a former director of the Congressional Budget Office.

With low TR, the policy analyst has a large set of methodological choices that will be publicly observable. Following the similar nomenclature of Simmons et al (2011), I call this problem the policy analyst degrees of freedom (PADF). As with researcher degrees of freedom, here the analyst has a menu of options to, consciously or unconsciously, tweak the results in order to obtain a desired outcome. This problem is depicted in figure 2.2.3 where, even with high TR in research, now there are three policy reports representing different quantifications of gains and losses for the same policy issue.

The problem of multiple reports due to low TR in policy analysis and PADF, increases geometrically when compounded with low TR in research and researcher degrees of freedom. This scenario is represented in figure 2.2.4, where there are multiple versions of what is the best research and multiple

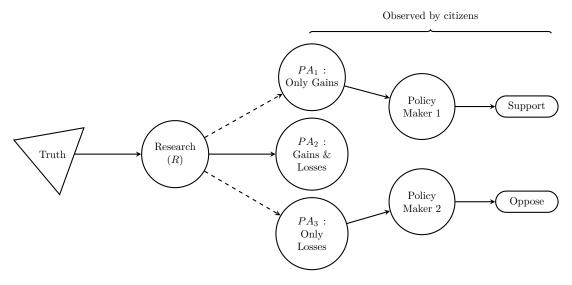


Figure 2.2.3: High Credibility in Research and Low Credibility in Policy Analysis

policy analyses. In this context, different policy makers, instead of weighing the same evidence in accordance with different values, now disagree on what the evidence is, failing to reveal what their differing values are. In Figure 2.2.4 each policy maker now claims as credible a policy estimate that supports either normative position, becoming indistinguishable from each other.

Multiple policy analysis of the same issue, and over the same dimensions, can only prevail if it is not possible to critically appraise its content. If the quality of a report is not observable we should expect policy makers to put a large value on reputation as a proxy of quality. Doberstein (2016) finds evidence that the credibility of the same policy brief varies drastically when only the name of the authoring institution is replaced. This large reputational premium is evidence of how difficult it is to critically examined a given analysis on the basis of its own merits (sound reasoning, methodology, data and execution).

Moreover, this reputational premium can help to perpetuate low levels of TR in policy analysis: reputable institutions are the ones better suited to increase TR in policy analysis, but at the same time benefit from receiving a branding premium if overall low levels in TR prevail.

The response to the reproducibility crisis, described in section 2.1, aims to shed light on the true quality of the empirical work done in research. This enables separation of good from bad research,

and provides complete transparency on how different findings are aggregated⁶. With high TR in research, the link between evidence and policy is strengthened as now all policy estimates are based on the same inputs from science. However, if TR is still low in policy analysis, as in figure 2.2.3, different policy reports can still provide opposing accounts of the empirical consequences regarding a policy issue. In this setting policy makers still disagree on what the empirical evidence is.

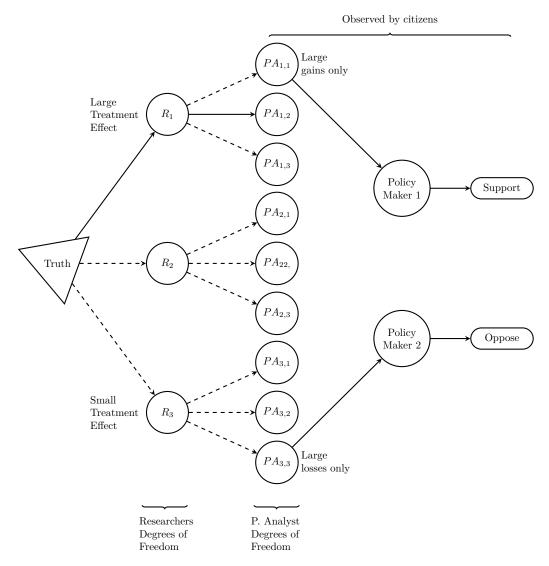


Figure 2.2.4: Policy-making with low TR in research and policy analysis

⁶For a clear example of how different pieces of research can be aggregated in a transparent and reproducible fashion, see Hsiang et al (2013) and their tool to reproduce the findings in their meta-analysis at http://dmas.berkeley.edu/

2.2.1.2 LITTLE KNOWLEDGE EXISTS REGARDING HOW RESEARCH AFFECTS POLICY ESTIMATES

To understand this issue, it is helpful to zoom in on the policy analysis component of figure 2.2.1 and describe its process in further details. Figure 2.2.5 represents a simplified model of the process involved in a policy analysis. Analysts use three primary sources: information from previous research (e.g. elasticities, behavioral parameters), data to contextualize the specific policy issue (e.g. microdata for the specific context where the policy issue is discussed), and guesswork to fill in any missing pieces required to complete the analysis (e.g. extrapolation parameters, take-up rates, distributional effects). All of these sources are used to generate inputs that are used in a model (e.g. microsimulation, cost benefit analysis), and this model produces the policy estimates to be used by policy makers.

When a policy analysis lacks transparency and reproducibility, it is not possible to observe the mechanisms through which research feeds into the model. Moreover, as the final policy estimates depend on multiple sources, it is also impossible to attribute possible changes in the final estimates to any specific source. This prevents us from understanding precisely how research affects the estimations produced by policy analysis.

Vivalt (2015) and Nutley et al (2007) have observed that it is not clear how policy makers use evidence in their decision process. Snilstveit et al (2013) develop a framework to identify gaps of knowledge understood as a relative scarcity of rigorous impact evaluations. But low TR in policy analysis still prevents us from understanding how the emergence of new evidence might affect the final policy estimates. If we were to completely understand how the current research is used in policy analysis (high transparency), we could accurately assess the potential effect of new research on policy estimates. The ability to reproduce the analysis with little effort would allow us simulate the effects of potential new evidence on the current policy estimates. With this information, research resources could be allocated based on where the value of additional information is the highest, and researchers could identify the biggest gaps in knowledge related to specific policies.

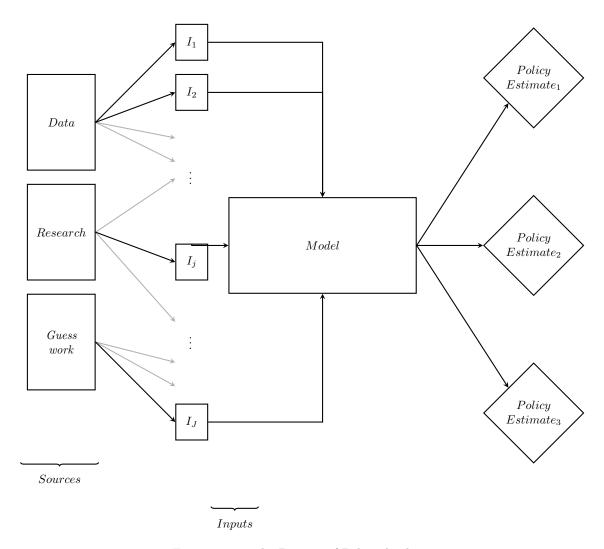


Figure 2.2.5: The Process of Policy Analysis

2.2.1.3 HARD TO AUTOMATE OR SYSTEMATICALLY IMPROVE POLICY REPORTS

The process of policy analysis as described in figure 2.2.5 can be an algorithmic process. Policy analyses are usually recurrent, where the effects of a policy issue are usually re-evaluated over time (e.g. the effect of raising the minimum wage), and similar issues are discussed in different settings (federal and state minimum wage policies, or minimum wage in different countries). A transparent and reproducible policy analysis can be more easily adapted to different settings, and modifications can be closer to well justified upgrades as opposed to somewhat arbitrary changes in methodology.

In addition to the case of the minimum wage, examples of recurrent analyses can be found across different domains: Greenstone (2009) has argued toward a cyclical review of cost and benefits of regulatory policy in the US; in healthcare both the US and the UK review guidelines for appropriateness of care across conditions (see for examples of analyses in Agency for Healthcare Research and Quality (2015) for the US, and Kerr et al (2014) for the UK). For all these cases, key parameters of the analysis can change every month (or few months) as more evidence is becoming available. In this setting, a policy analysis that updates itself automatically might be useful for policymakers.

Policy analysis requires a concrete answer in a heavily constrained environment (time and resources). Regardless of how constrained the answer is, the result of the process can be summarized in the model of figure 2.2.5. Whenever research and data do not contain all the relevant information to execute the analysis (given the constrained environment, this is almost all the time) some elements of the model have to be the result of educated guesses made by the analyst. This guesswork is rarely made explicit in a report, but the need to make arbitrary decisions is an almost inevitable task of the policy analyst, sometimes playing a predominant role as represented in figure 2.2.6.

Low TR in policy analysis affects both the possibility of making incremental improvements to the analysis (aiming towards the algorithmic ideal) and makes it harder to keep track of the guess work performed, making consistency over time a much harder goal to achieve.

2.3 Description of the case study

In 2014, the Congressional Budget Office published a report estimating the effects on employment and income of a potential raise in the federal minimum wage, from \$7.25 to \$10.10 per hour. The

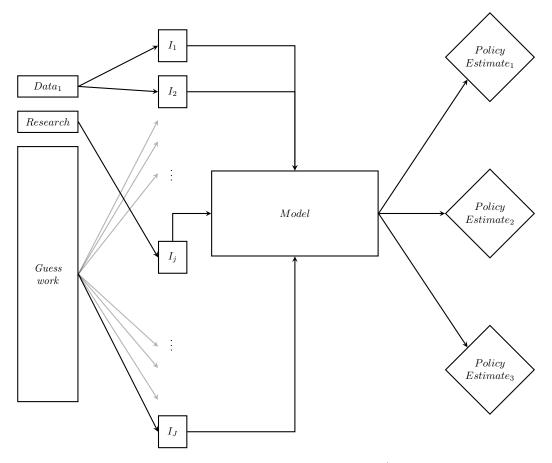


Figure 2.2.6: Elements of policy analysis with heavy time/resource constraints

minimum wage was raised for the last time in 2007 with no adjustments for inflation, so had decreased in real value since. The new proposal involved indexing the new minimum wage to inflation⁷

The report estimated positive effects due to wage gains, and negative effects due to job losses and income losses from those paying for the wage gains (US Congressional Budget Office, 2014). The total wage gain was estimated to be \$ 31 billion for 16.5 million workers. The number of jobs lost was estimated to be around 500,000. The net distributional effects were as follows: \$5 billion net total gains for households below the Federal Poverty Line (FPL); +\$12 billion for households between one and three FPL; +\$2 billion between three and six times the FPL; and net total loss of \$17 billion for household with incomes above six FPL.

 $^{^7{}m The}$ \$9 option described in the previous footnote was not indexed to inflation. All the effects estimated by CBO were to the year 2016 so indexing did not made a substantive difference.

The research on effects of minimum wage on teenage employment is well developed in the US. These effects are measured by estimating the elasticity of labor demand for this population. A intense debate on these effects has driven the research agenda for more than two decades (Card and Krueger, 2015; Neumark and Wascher, 2008; Dube et al, 2010; Clemens, 2015a). The findings can be grouped into two schools: one documents large effects on employment with estimates for elasticity of labor demand for teenagers concentrated around -0.1, and the other reports small effects on employment with estimates concentrated around -0.01 for the same parameter⁸.

The case study on the minimum wage policy analysis was chosen based on four criteria: relevance, generalizability, recurrence and feasibility.

First, the report was clearly relevant in the policy debate. It was cited by proponents and opponents of the raise, and it was featured prominently in the news and editorials of that period. As an illustration, figure 2.3.1 shows how the publication of this report coincides with the highest search intensity in Google for term "minimum wage" in the US.



Figure 2.3.1: Google Search Intensity of "Minimum Wage"

Second, CBO is among the most transparent and rigorous offices of policy analysis. The protocols and tools discussed below should be understood as one additional layer of TR, on top of the best practices already presented in the CBO report. Lessons from TR that apply to the CBO report should apply also to the work of most official agencies and producers of policy analyses. Additionally, the policy issue is widely known, which facilitates extrapolation to other policy analyses.

Third, the discussion around the minimum wage in the US is notably recurrent. This makes it highly likely that a similar policy analysis will be conducted again in the future. The case study can be

⁸This academic debate transcends the minimum wage policy debate in at least two dimensions: in economics has come to represent a challenge the predictive power of the most basic models taught in introductory classes, and in empirical research in general the debate is contemporaneous to a debate between the relative importance of empirical strategies vs theoretical predictions.

directly used in future calculations.

Finally, it was feasible. All of the data were publicly available, the report describes the analysis in detail, and there was only one policy lever to analyze (the minimum wage level).

Chapter 3

METHODS

Contents

3.1 Adapt TOP Guidelines to Policy Analysis	34
3.1.1 Three standards of TR in Policy Analysis	36
3.1.1.1 Workflow	36
3.1.1.2 Data	38
3.1.1.3 Methods	38
3.2 Apply Guidelines to the CBO Report and Reproduce it Using	the the
Highest Standards of TR	40
3.2.1 A Dynamic Document for CBO's report	40
3.2.2 Distributed Version Control (Git)	41
3.2.3 $$ Differences Between Reproducing Research and Policy Analysis	42
3.3 Sensitivity Analysis	42

To address the low credibility critique of policy analysis (Manski, 2013), the general methodological contribution of this dissertation is to develop an approach that aims to increase the transparency and reproducibility (TR) of policy analysis.

The TR approach to policy analysis developed here consists of three steps: (i) translate guidelines from science to the policy analysis setting, (ii) use the guidelines and state of the art tools (git and dynamic documents) to increase TR in policy analysis, demonstrated with the case study, (iii)

conduct sensitivity analysis to identify components of the policy analysis where additional knowledge is most policy relevant.

3.1 Adapt TOP Guidelines to Policy Analysis

Combining the lessons from the reproducibility crisis in science (section 2.1) with the framework used to describe the problems of low transparency and reproducibility in policy analysis (section 2.2), it is possible to explore parallels between low transparency and reproducibility in scientific research and policy analysis.

Table 3.1.1 summarizes the key elements of low TR in research and policy analysis. In both areas, the problems arise from the inability to critically inspect every detail of how the output is produced. In plain English: we do not know how the sausage is made! The scientific community is opening up the kitchen and devising mechanisms for anybody to have access to the cookbook.

Table 3.1.1: Comparison of Low Transparency and Reproducibility (TR) in Research and Policy Analysis

	Research	Policy Analysis		
Output	Peer reviewed publication	Policy report		
Problems of low	Publication Bias.	Low credibility. Unclear		
TR	Specification Search	connection between research		
	(P-Hacking, Garden of	and policy. Hard to improve		
	forking paths). Data	systematically. Data fudging.		
	fudging.			
Common Solutions	Disclosure of key details.			
	Open data and materials.			
Common Tools	Dynamic documentation.			
	Distributed version control.			
Specific Solutions	Test for reproducibility;	Develop reproducibility;		
	Registration of	Systematic and continuous		
	pre-analysis	$\operatorname{updating}$		
Who increases	Researchers, Funders,	Not the policy analysts (Policy		
TR	$_{ m Journals}$	schools? Think tanks? Media?		
		Bloggers?)		

In section 2.1.1.2 I reviewed how the scientific community is addressing the reproducibility crisis. A key part of the response has been the development of guidelines that build explicit standards

from norms that underlie the practice of good research (Miguel et al, 2014; Merton, 1973). The Transparency and Openness Promotion guidelines (TOP) developed by Nosek et al (2015) are a good example of such efforts. Each standard is presented in a way that allows us to qualify different levels of compliance. This allows us to understand the challenges of TR in a continuum from "opaque and not reproducible at all" to "full transparency and reproducibility" (Peng, 2011). In addition to the different standards and levels of compliance, another key feature is that these guidelines are an ever evolving document (currently at version 1.0.1) with dozens of researchers collaborating across different fields.

The central thesis of this dissertation is that the adoption of some of these standards into policy analysis can address the problems described in section 2.2.1. Translating such norms into policy analysis should be seen as one more step in the ongoing improvements towards a more open government (White House, 2013). This step would involve adding an *open methods* component to the growing governmental initiatives of *open data* around the world (Gray, 2014).

Of the three norms described in Miguel et al (2014) to increase TR in science, two can be translated directly to policy analysis: disclosure of key details and open access to all data and materials. The translation of the third norm regarding preregistration and pre-analysis plans (PAPs) is not as direct. PAPs are a key component of increasing research transparency. In policy analysis, a parallel can be drawn but it has to take into account important logistical differences.

When it comes to PAPs, the key difference between the TR practices in research relative to policy analysis is associated with the nature of the output produced in each case. The output for which research is traditionally judged are peer reviewed articles, and its objective is to advance the state of knowledge in some specific domain. To avoid the file drawer problem and publication bias (described in section 2.1.1.1), the norm of preregistration, and use of PAPs, is proposed to encourage the display of all of the scientific analyses, as opposed to only the ones that result in "interesting" results (strong effect size, statistically significant).

The output in policy analysis are policy reports. They represent the best estimation available of a specific policy issue chosen by an official agency or policy research organization, and its objective is to inform a specific policy maker (client). Preregistration, as proposed for academia, is not feasible in this environment. Policy reports are not statements about the truth to be advanced, but the best possible answer to a policy-related question given important resource constraints.

However, this apparent disadvantage can be turned into a strength: policy reports could, in principle, become "living documents" subject to continuous updating and scrutiny from analysts, academics and the general public. As most policy analyses have a cyclical nature, the living document created for the last policy debate around an issue can be understood as the "pre-analysis plan" of the policy analysis for the next time the debate comes to light.

The guidelines drafted here follow the framework developed in the TOP guidelines for scientific TR (Nosek et al, 2015). The TOP guidelines have eight different dimensions or standards to assess transparency and openness (described in section 2.1.1.2). Each standard is scored on a four level scale, from the lowest score at level 0 to the highest score at level 3.

The goal of the adapted guidelines is to identify standards and levels that would address the problems described in section 2.2.1. Three standards are proposed and described here: (i) Workflow, (ii) Data, and (iii) Methods. As with the TOP guidelines, the standards proposed here should be understood as an ever-evolving document. The TOP guidelines list of authors is currently has 40 members, and its latest version is number 1.0.1. The guidelines presented here are version 0.0.1, and will change over time to reflect future collaborations and feedback from key stakeholders.

Currently, the practice of public comment on legislation/regulation in the US represents a step in this direction. However, as described in section 2.2.1 there is still room for systematic improvement and automation of the analytic process.

3.1.1 Three standards of TR in Policy Analysis

3.1.1.1 Workflow

For this section I use the workflow model described in figure 2.2.5 in chapter 2 that represents a stylized version of the workflow of policy analysis. Inputs are built from three primary sources: data, research, and guesswork.

Data is any source of contextual information that can be cited. Examples are surveys, listing of prices and administrative records. Research in this context is understood as the pool of knowledge about the behavioral parameters that can also be cited; examples are effect sizes from a single study or multiple studies aggregated through meta-analysis. Guesswork is any remaining piece

of information needed in the model to compute the final policy estimates that cannot be cited. Examples are extrapolation factors of treatment effects from one population to another, take-up rates, distributional effects, and others.

Each source component then feeds into a model that combines all elements to produce the final policy estimates. To clarify the elements that are derived from each source, I define a intermediate elements as Inputs. Inputs are just an organized list of all the elements described above, such that policy analysts and researchers can be clear of what is data, what is research and what is guesswork.

The model represents a set of procedures applied to inputs in order to quantify the potential effect of the analyzed policy. Examples are micro-simulations and cost-benefit analysis, among others. The model should be characterized by a set of equations and a narrative describing how to apply them. The final policy estimates are the output produced by the model. High TR in the workflow dimension means that all of the components are clearly labeled, together with a clear explanation of how the different pieces feed into each other. The goal of any policy analysis is to generate policy estimates, understood as a set of quantities that reflect the best available information regarding objective/positive losses and gains associated with the implementation of a policy. These policy estimates will serve as an input to be valued normatively by policy makers. TR in workflow requires the set of policy estimates to be presented in a standardized way, achieving consensus among all policy makers on what would be understood to be the commonly agreed-upon evidence.

The policy report should highlight where and why different quantities cannot be compared without a normative assessment. In order to facilitate the normative comparison, all quantities should be reported in the same units.

- Level 1: Identify clearly all of the policy estimates to be used by policy makers. As different policy makers will weigh policy estimates differently (some may focus on the costs, while other on the benefits) all quantities should be presented in the same units (e.g. average increase in per-capita income across quintiles of the income distribution).
- Level 2: Level 1 + Identify clearly all of the inputs to be used in the model and classify their origin from data, research or guesswork.
- Level 3: Level 2 + The complete workflow should fit a diagram similar to that shown in figure
 2.2.5. Users of the analysis should be able to change specific components with minimal effort,

and observe how that change affects the policy estimates.

3.1.1.2 Data

Data used for policy analysis has the purpose of informing decisions in the public sphere. Following new standards of open data (Obama, 2009), these resources should be open by default and a detailed rationale should be made whenever access to the data is restricted. Whenever the raw data cannot be accessed, the policy analysis should provide access to masked or aggregated data. Additionally, detailed instructions on how to go from the raw data to the masked data should be provided, such that other analysts with access to the raw data could reproduce the exact intermediate data.

- Level 1: Policy report states explicitly whether all, some components, or none of the data used in the analysis is available. In the case of differential availability of components, it should be clear which items are fully available, masked, aggregate or not available at all. Clear instructions should be provided to access each of the available components.
- **Level 2:** Policy report is published with the data. The report and data can be accessed in the same place.
- Level 3: Policy report is published with embedded code that calls the data in a repository and changes in the data will produce traceable changes in the report.

3.1.1.3 Methods

Methods in policy analysis should be understood as a detailed set of instructions of all the steps taken to produce the policy estimates. These instructions should describe how each data set, research and guesswork component were used to generate inputs and how those inputs were subsequently used in the model that computes the policy estimates. The intended audience for these instructions should be staffers of policy makers, researchers and other analysts. For this purpose, the material should be presented at different levels of depth with a clear narrative connecting all of the steps involved. This could be achieved by creating links in each section, such that if the user wants to learn more (starting from key assumptions, to narrative explanations, to equations, to code) she can unfold every level of detail needed to reproduce the exact same result.

- Level 1: Methods should be clearly described at different levels of detail. A reader of the policy report should be able to understand all of the components and reproduce qualitatively similar estimates.
- Level 2: Level 1 + all of the code that reproduces the exact policy estimates should be available and running.
- Level 3: Level 2 + the code and narrative should be clearly legible for different audiences (staffers, researchers, other policy analysts). The code should be in the same document that describes each step, and the users should be able to manipulate all of the components of the code and trace its effects on the final policy estimates.

3.2 Apply Guidelines to the CBO Report and Reproduce it Using the Highest Standards of TR

Using the guidelines developed above, I score the CBO report on the effects on minimum wage. The score for each standard is discussed in the results section.

To create a version of the CBO report that meets level 3 (highest) of TR in each of the three standards (workflow, data and methods), it is necessary to describe two tools, borrowed from computer science, that facilitate the achievement of the highest TR: Dynamic Documents, and Git or Distributed Version Control. A brief description of each tool is presented below with an explanation of how it can help to increase TR for each standard.

3.2.1 A Dynamic Document for CBO's report

The goal is to achieve the highest standards of reproducibility and transparency. With a similar purpose in mind, Knuth (1992) developed the philosophy of *literate programming*, where computational code and analytic narrative are deployed in a single environment. The most up-to-date implementations of this methodology are Dynamic Documents and Jupyter Notebooks developed for the programming languages R Xie (2015) and Python, respectively. Both tools achieve the goal of combining code, narrative and mathematical formulae in one place, and both are non-proprietary software, which minimizes barriers to utilization.

In this demonstration I use Dynamic Documents (DD). Each DD consists typically of three parts: a narrative component, a coding component, and an equations or modeling component. All three pieces are combined (or "knit") together in R using the package knitr.

The narrative component is edited using Markdown, an editing language designed for easy adoption and with the option of outputting either to website format (HTML) or printable documents (PDF and .doc files). The coding components are either inside the narrative text or in code chunks implemented in R¹. Code chunks may or may not be displayed in the final output depending the the user's choice. The equations or models are imputed using LaTeX syntax. The final file is a R-Markdown file (.Rmd) to be excecuted in R. When excecuting it, the user can choose among three

¹The package knitr allows to run other languages but all its features are only available in R

outputs: HTML, PDF and a .doc file. This allows the user to put all of the components of the analysis in one place, and update them when new information becomes available in what is known as a one-click reproducible Workflow.

This technology helps to implement the standards described above to their highest level. The definition of a DD implies that the data has to be available, published with the report and set up in a way that reacts to changes in the report (level 3 of data standards). A complete DD should also accomplish the highest standard in the methods dimension as it is possible to describe the methodology and the code in detail, and to run successfully in different machines (dynamic component). Different users should be able to read it and execute it with minimal effort (level 3 methods standard).

The highest TR in the workflow standard is not automatically achieved with the use of DD, but DD does significantly facilitate its achievement. Dynamic documents, as opposed to static printed documents, can have different layers of depth, such that different audiences can choose their desired level of detail. At the simplest level the reader/user should be able to visualize a workflow like the one presented in figure 2.2.5.

3.2.2 DISTRIBUTED VERSION CONTROL (GIT)

Git is a version control system that allows multiple users to edit the same file without losing track of any modifications. Since its development in 2005, it has become a universally required tool for software developers, and in recent years has shown increasing popularity in the research community.

The three main reasons to use git are: (i) It tracks all of the changes done on any file containing code, rendering obsolete the need for multiple versions and names for files. (ii) Allows for multiple users to "clone" a version of the official code and modify it, and (iii) the contributor can request that her modifications be incorporated to the original file ("pulled" back), and everybody following the original can see all of the suggested changes.

Github is a popular website that hosts most of the work using git. The Open Science Framework is another platform that uses git and Github, and is specifically dedicated to researchers.

As described in the previous section, a DD provides all of the elements to potentially achieve the

highest standards of TR in policy analysis. Git provides open access, and allows modifications of the DD, realizing all of the potential for TR policy analysis.

3.2.3 Differences Between Reproducing Research and Policy Analysis

The new reproducibility practices promoted in science were used as a framework to carry out the exercise of reproducing the CBO report into a DD. However, it is necessary to highlight a crucial difference between a reproduction exercise in science and a reproduction in policy analysis.

A scientific report takes the form of a peer-reviewed publication that represents several months or years of research, followed by a review process that can be as lengthy as the research itself. When a scientific publication is subject to computational reproduction, it is expected to succeed (as a precondition of the basic scientific principles discussed in section 2.1.1.2).

A policy analysis report is usually performed under tight deadlines, and it is not unusual to rely on arbitrary assumptions and/or unreproducible calculations. For these reasons the CBO report cannot be reproduced as a way of testing the veracity of the analysis, but as a demonstration of what can be achieved with high TR in policy analysis.

Here reproducibility is used, paired with full transparency, to generate a living document that represents the best policy analysis to date. The expectations are that this living document will serve as a building block to discuss and incorporate incremental improvements on the best available policy analysis for future policy reports on the minimum wage.

3.3 Sensitivity Analysis

When a policy analysis has achieved high levels of TR for each standard, it is possible to conduct a sensitivity analysis for each component used in the policy analysis. For this exercise it is particularly useful to have all of the outputs of the policy analysis (policy estimates) in the same units, and to clearly identify the dimensions to be normatively aggregated by policy makers.

For our case study, the original CBO report presented the benefits and costs in different units: wage gains specific for families that have a net wage increase due to the new minimum wage, wage losses for specific families that have net wage decrease due to job loss, and average income lost for all families that is used to pay for the wage increase, labeled from now on balance losses, and distributed across poverty line bins (less than one FPL, between one and three FPL, etc).

Taking the DD to the highest level of TR in the workflow standard implies that these benefits and costs have to be translated into the same units. For this purpose all of the policy estimates are expressed in terms of average per-capita income gain/loss, across quintiles of income.

With five quintiles and three types of policy estimates (net wage gain, net wage loss, and balance loss), the dimensions of the analysis becomes too large even when looking at a few parameters. As an illustration of how all dimensions could be condensed into a single number, I model the different valuations of hypothetical policy makers using additive weights for each policy estimate and weights to account for different redistributional preferences. The result is a welfare function $(W(\cdot))$ that combines all of the policy estimates and personal valuations of a given policy maker.

Formally, W can be defined as the weighted sum of policy estimates for the wage gain (wg_i) , wage losses (wl_i) and balance losses (bl_i) across all individuals, where each policy estimate receives a weight $\omega_{wg}, \omega_{wl}, \omega_{bl}$, and the distributional preferences are a function of the income quintile of each individual $\omega_i^d(Q_i, \rho)$:

$$W(\rho) = \sum_{i \in N} (\omega_{wg} w g_i + \omega_{wl} w l_i + \omega_{bl} b l_i) \, \omega_i^d(Q_i, \rho)$$
(3.3.1)

with:

$$\omega_i^d(Q_i, \rho) = \frac{(1 - \rho(Q_i - Q_{median}))}{\sum_i \omega_i^d(Q_i)} Q_{max} \quad \text{for } \rho \in \left(-\frac{1}{2}, \frac{1}{2}\right)$$

 Q_i represent the quintile in the income distribution (1 the lowest and $Q_{max} = 5$ the highest), and ρ parametrizes the preferences towards redistribution ($\rho < 0$ dislikes redistribution, $\rho > 0$ likes redistribution). The parameter ρ is restricted to values between $-\frac{1}{2}$ and $\frac{1}{2}$ so all weights are strictly positive. This function was designed ad-hoc only to illustrate a possible set of preferences used by policy makers when observing the policy estimates.

As an illustration, figure 3.3.1 presents values of $W(\cdot)$ for different redistributional preferences (ρ) assuming $\omega_{WG} = \omega_{WL} = \omega_{BL}$. In this example a policy maker that values redistribution positively (with $\rho = 0.1$), will see a value of \$9.7 billion dollars over increasing this minimum wage. Conversely, a policy maker that dislikes redistribution (with a $\rho = -0.1$), would value the proposed policy at -\$5.7 billion dollars. Part of sensitivity analysis is performed over this two sample positions $(W(\rho = 0.1), W(\rho = -0.1))$, so those two values are highlighted in the figure.

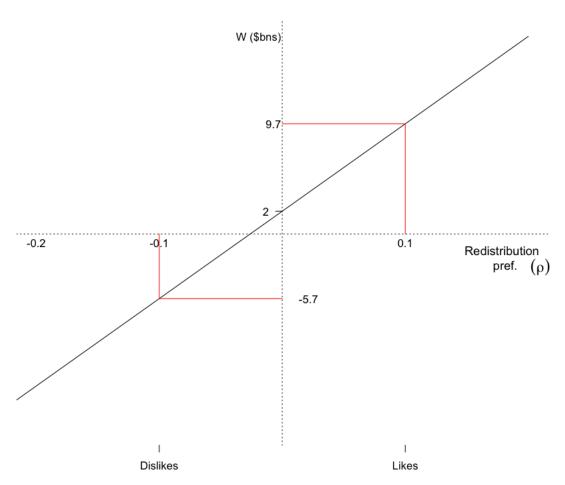


Figure 3.3.1: Normative Policy Estimates for different distributional preferences (ρ) . Red lines represent values of $W(\rho)$ for $\rho = 0.1$ and $\rho = -0.1$, to be used in the sensitivity analysis

Chapter 4

RESULTS

Contents

4.1	Desc	cription of the Dynamic Document for the CBO Report 48	5
4.2	Asse	essing TR in Original CBO Report and in the DD 50	0
	4.2.1	Standard #1: Workflow	0
	4.2.2	Standard #2: Data	8
	4.2.3	Standard #3: Methods	9

In this chapter I describe the Dynamic Document created to achieve the highest standards of TR. Then I use the TR guidelines to compare the DD and the original report from CBO.

4.1 Description of the Dynamic Document for the CBO Report

Two formats of DD were considered for reproducing the CBO report. The first version aimed to display the report verbatim, adding annotations with equations, code, and added narrative to achieve the highest levels of TR. While doing this exercise two difficulties emerged: some methodological procedures were not sequentially connected (for example the data used to compute effects on employment is described as the CPS on page 20, but on page 32 is it possible to infer that it is specifically

referring to the CPS Outgoing Rotation Group), and other policy estimates were found to have no methodological description whatsoever.

In the second format, two major changes were made: first, several discussions in the report that lacked any quantitative analysis were grouped under "other factors" and briefly summarized at the end of each section. Second, the DD was structured around the key policy estimates produced (effects on employment and family income), such that the reader can see the narrative, the equations, the code, and the output in the same sections.

The final version of the dynamic document (DD) for the CBO report can be found at http://rpubs.com/fhoces/dd_cbo_mw. The DD is designed to achieve the best readability in website format (HTML), but a PDF version is provided in the appendix of this dissertation. The reader is invited to explore the DD in its entirety. Here is a brief description of its key features.

Clear description of all the steps. One important benefit of using a DD is that as the analysis is performed in the code it is complemented with a narrative and analytical explanation at every step. For example, the images 4.1.1 - 4.1.3 show that when calling the data, a hyperlink is provided for all of the alternative data sources, and the code can be unfolded in the same document to see how the data is being used.

Dynamic behavior of tables and figures (minimal hard coding). Most of the outputs, including plots and tables, are not hard-coded but are produced every time the source file is executed (in R). This minimizes the risk of human error and facilitates multiple sensitivity tests. For example, images 4.1.4 - 4.1.5 show how the final population is computed analytically, displays the relevant statistics to compute the population, and the code that executes the calculation can be unfolded in the same file.

1 Introduction

- 2 Employment effects
- 2.1 Data, wages, and forecast
- 2.2 Get the N
- 2.3 Get the $\eta \times \Delta w$
- 2.4 Other factors
- 2.5 Computing effects on employment
- 3 Distributional effects
- 3.1 Computing Family income
- 3.2 Imputing policy effects
- 3.3 Computing family income under status quo and minimum wage increase
- 3.4 Other considerations
- 4 Results

Dynamic Documention for "The Effects of a Minimum-Wage Increase on Employment and Family Income"

Download/Modify/Contribute the analysis here Last edit: 2016-12-20

1 Introduction

The role of policy analysis is to connect research with policy. Because of heavy time constrains, policy analyses are typically ambiguous regarding the details of how the analysis was carried out. This creates three problems: (i) its hard to understand the connection between research and policy, (ii) allows policy makers to cherry pick policy reports, and (iii) hinders systematic improvement and/or automation of parts of the analysis. In this document we demonstrate the use of a reproducible workflow to reduce the ambiguity in policy analysis.

Here we attempt to contribute to the policy discussion of the minimum wage. The minimum wage is a contentious policy issue in the US. Increasing it has positive and negative effects that different policymakers value differently. We aim to add clarity on what those effects are, how much do we know about them, and how those effects vary when elements of the analysis change. We select the most up-to-date, non-partisan, policy analysis of the effects of raising the minimum wage, and build an open-source reproducible analysis on top of it.

In 2014 the Congressional Budget Office published the report titled "The Effects of a Minimum-Wage Increase on Employment and Family Income". The report receive wide attention from key stakeholders and has been used extensible as an input in the debate

Figure 4.1.1: Screen shot of DD: Beginning of the document. Brief introduction, floating table of contents, and link to github repository where users can download the source code to reproduce the same document.

Openness and role for Git. The complete analysis presented in the DD is generated using one file. That file can be downloaded and modified by anybody. Modifications to the DD can be suggested by other analysts through GitHub. This is probably the most important feature of TR policy analysis, as it provides a structure to incrementally improve analyses that until now have to be redone every time a policy issue resurfaces.

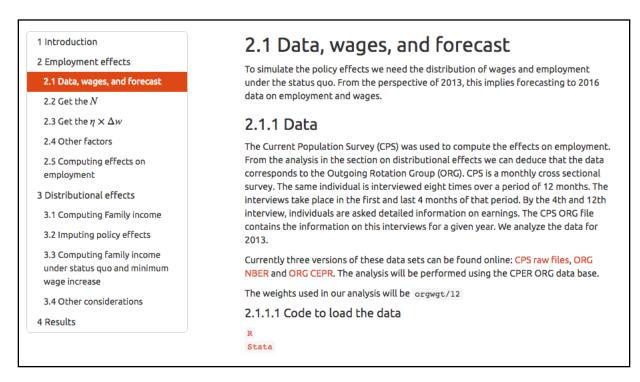


Figure 4.1.2: Screen shot of DD: Everything in one file. Starting from the data. The document lists (with links) the possible data sets to be used, and below a section can be unfold and where it shows the actual code used to call the data from a specific web location

```
The weights used in our analysis will be orgwgt/12

2.1.1.1 Code to load the data

R

call.cps.org.data <- function(){
    data_use <- "CPER_ORG"

# Using CEPR ORG data

if (data_use == "CPER_ORG") {
    # Checking if working directory contains data, download if not.
    if (!("cepr_org_2013.dta" %in% dir())) {
        # create name of file to store data
        tf <- "cepr_org_2013.zip"

# download the CPS repwgts zipped file to the local computer
        download.file(url = "http://ceprdata.org/wp-content/cps/dat
        a/cepr_org_2013.zip", tf , mode = "wb")

# unzip the file's contents and store the file name within t *
```

Figure 4.1.3: Screen shot of DD: Unfolding the code chunk that downloads the data.

```
\widehat{N_{final}^{adult}} = g(201 \hat{6}|2013) \times N_{employed}^{\hat{adult}}(2013) \times P(\hat{w} \leq MW^1|adult) \times (1 - \alpha_1^{\hat{adult}} - \alpha_2^{\hat{adult}})
The table below presents the estimate from 2013 for all each component.
Stata
Characteristics of target population
                                                                                                           Adult
                                                                                                                                             Total
                                                                                                                           Teen
Salary workers (N_{employed}) (millions)
                                                                                                         121.69
                                                                                                                                           125.99
                                                                                                                            4.30
Low wage workers (w \le 11.5p/h) (millions)
                                                                                                           26.85
                                                                                                                           3.73
                                                                                                                                            30.58
% Salary below new MW (P(\hat{w} \leq MW^1))
                                                                                                           14.43
                                                                                                                          74.34
                                                                                                                                            16.48
% of non compliers (\alpha_1)
                                                                                                           14.38
                                                                                                                          21.70
                                                                                                                                            15.16
g(2016)2013)
                                                                                                            1.05
                                                                                                                            1.05
                                                                                                                                              1.05
                                                                                                           15.75
                                                                                                                            2.62
                                                                                                                                            18.45
 N_{final} (millions)
```

Figure 4.1.4: Screen shot of DD: Computing statistics for the relevant population. Every time the data or code changes, the output of this table will change *dynamically*

```
\overline{N_{final}^{adult}} = g(201\hat{6}|2013) \times N_{employed}^{adult}(2013) \times P(\hat{w} \leq MW^1|adult) \times (1-\alpha_1^{\hat{\Delta}oult}-\alpha_2^{\hat{\Delta}dult})
The table below presents the estimate from 2013 for all each component.
 N.final.f <- function(){
   aux.1 <- bind_rows(
     df %>%
     filter(pop_of_int == 1) %>%
     mutate("adult" = ifelse(age>=20, "Adult", "Teen") ) %>%
     group_by(adult) %>%
     summarise( "Salary workers (\$\backslash \{N_{employed}\} \) (millions)" = sum(final\_weights)/le6, \\
                  "Low wage workers ($w \\leq 11.5 p/h$) (millions)" = sum(final_weights * (wages.f
 inal <=11.50))/le6,
                  "% Salary below new MW ($P(\\hat{w} \\leq MW^{1})$)" = wtd.mean( 1*(wages.final <
 =10.10),
                                                na.rm = TRUE, weights = final_weights) * 100),
     df %>%
     filter(pop of int == 1) %>%
      summarise( "Salary workers ($\\hat{ N_{employed} }$) (millions) = sum(final_weights)/le6,
                   "Low wage workers ($w \\leq 11.5 p/h$) (millions)" = sum(final_weights * (wages.
 final <=11.50))/1e6,
                   "% Salary below new MW ($P(\\hat{w} \\leq MW^{1})$)" = wtd.mean( 1*(wages.final
```

Figure 4.1.5: Screen shot of DD: As each element of the table reacts to variations in the code it will the final key statistic used in the calculations of employment.

4.2 Assessing TR in Original CBO Report and in the DD

Using the guidelines developed in section 3.1.1, the two policy analyses (original CBO and reproduction in DD) are compared side-by-side across each standard of TR. The reader can use table 4.2.1 or a checklist provided in appendix 7.2 of this document as a reference.

Table 4.2.1: Summary of Guidelines for Transparent and Reproducible Policy Analysis

Standard	Level 0	Level 1	Level 2	Level 3
Workflow	Policy estimates vaguely described	All of the inputs, and their corresponding sources, used in the calculations are listed	Lvl 1 + Policy estimates are listed, in the same unit if possible	Lvl 2 + all of the components can be modified with little effort
Data	Report says nothing	Clearly stated whether all, some components, or none of the data is available, with instructions for access when possible.	Lvl 1 + report and data are in same place	Lvl 2 + Report has specific lines of code that call the data and changes in the data produce traceable changes in the report
Methods & Code	Key assump- tions are listed	Methods are described in prose. Large amount of work is required to reproduce qualitatively similar estimates	Methods are described in prose, with detailed formulas, and code is provided as supplementary material	Lvl 2 + All is in the same document where changes in the code affect the output automatically

Adapted from TOP guidelines (Nosek et al, 2015) v1.0.1

4.2.1 Standard #1: Workflow

Original: Level 1

It required a large amount of work to identify the components of the analysis in the report. Analytic work was intertwined with narrative justification of other factors where no methodology was provided.

It was also hard to separate methodological information (sources, inputs and model in term of figure

2.2.5) from the policy estimates that are inputs for policy makers. The policy estimates are presented across different tables, visualizations and in different units.

DD: Level 2.5

An important difference between the original report and the DD, in the workflow standard, has to do with how the policy estimates are presented to inform policy makers' decisions. As discussed in section 3.1.1.1, a highly transparent policy analysis should aim to make clear and salient all of the advantages (i.e. benefits, positive effects) and disadvantages (i.e. costs, negative effects) that were produced as part of the empirical analysis. As described in the introduction, this is precisely what happened in response to the CBO publication.

This ideal level of transparency would minimize the opportunities for policy makers and stakeholders to selectively read the results of a report. And when such selective reading happens, this level of transparency would help to properly identify and keep a record of such selection.

Here I suggest that all the information needed by the policy makers should be contained in a single visualization or table. This output should identify all of the positive components produced by the analysis to be weighted differently by different policy makers. It is important that a consensus should be achieved on how to present such an outcome: its format should be invariant to future versions of the report, and when variations occur they should be properly documented and justified.

I begin with a visualization that puts together all the key policy estimates from CBO in their original format, and then argue for a different presentation that is closer to the ideal discussed above.

In a first attempt, figure 4.2.1 presents the policy estimates as discussed in the CBO report. On the x-axis we have all of the population sorted by per-capita income (before wage wage variations), and on the y-axis we have their variation in anual per-capita income. In addition to the two axes 6 red vertical lines were drawn to represent from one to six poverty lines. Each black dot above zero represent net wage gains for individuals in a household, while each black dot below zero represents net wage losses. Both of these changes represent the income variation for specific households that are directly affected by the raise in minimum wage (most do not observe a wage variation and are represented in the black dots clustered over zero). The blue dots represent the reduction in income for all individuals in the economy needed to pay for the raise in the minimum wage. Unlike the wage variations, these changes are applied to all individuals and should be interpreted as average income

variation for individuals in different bins of poverty levels.

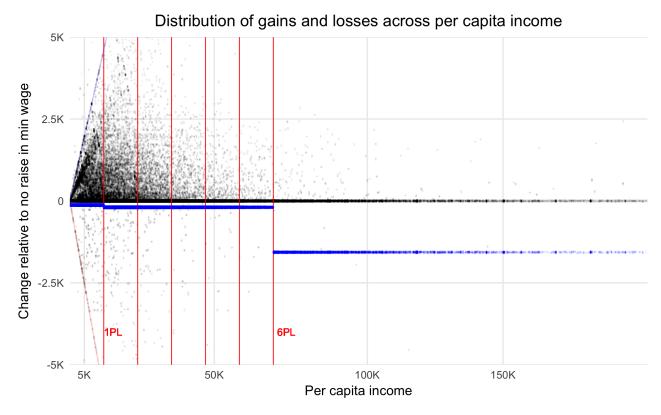


Figure 4.2.1: Gains and losses in different units.

Each black dot represents the variation in household per-capita income due to wage gains or job losses. The vertical red lines represent poverty lines (for a household of one). And the blue dots represent the average loss imputed to each individual in the population, to pay for the gain in minimum wage.

This figures succeeds at bringing all of the policy estimates into one frame, but does not present comparable units (effective wage variations v. average income loss).

In a second attempt, figure 4.2.2 puts the three policy estimates in the same units as average income variation for each source: wage gain, wage, loss, balance loss. Following CBO's original format these averages are take across all populations within each bin of poverty lines.

The problem with this figure is that not all bins contain the same number of people. This is a very good example of how the positive analysis, if presented according to figure 4.2.2, is implicitly suggesting a non-obvious normative component in its analysis: if a policy maker were to compare

Distribution of gains and losses across poverty lines 300 -Wage + -1445 Wage -Balance -200 -\$/year 100 -0 2 3 5 6 1 4

Figure 4.2.2: Gains and losses. Same units, but different denominator. Red bars represent average per-capita income gain due to wage gains. Green bars are the average per-capita income loss due to job loss. And blue bars are the average income loss used to paid for the wage gains. Each average is computed within each poverty line bin, and the number of people in each bin differs.

Poverty Lines

gains and losses in this figure she would be weighing the effect on wealthy people much more than the effects on low income people (for example: there are six times more people below the first poverty line, as there are above six poverty lines).

In the third and final iteration, figure 4.2.3 presents the three policy estimates using the same units (per-capita average within group), and displays the distributional effects, across equally sized groups (quintiles): net wage gains due to raising the minimum wage (wage gain), net wage losses due to raising the minimum wage (wage loss), and income loss required to pay for the increase in the minimum wage (balance loss). Now policy makers and stake holder can look at the same figure and use it to support different positions. This is one of the main goals of increasing TR in policy analysis.

Distribution of gains and losses across quintiles Wage + Wage - Balance - Quintiles of per capita income

Figure 4.2.3: Gains and losses. Same units and denominator. Red bars represent average per-capita income gain due to wage gains. Green bars are the average per-capita income loss due to job loss. And blue bars are the average income loss use to paid for the wage gains. Each average is computed within quintile of distribution of income. By construction the number of people in each quintile is the same.

In addition to transparency in the output, TR in the workflow dimension also involves a clear understanding and display of all the steps involve in the analysis. Following the structure of figure 2.2.5, the DD connects all the sources (data, research and guess work) with the inputs needed to do the analysis. Tables 4.2.2 and 4.2.3 list all of the components of the policy analysis for the case study and how they connect to each other. With all of these components clearly labeled and displayed, now it is easier to reproduce the analysis presented here, to improve it and to critically appraise it.

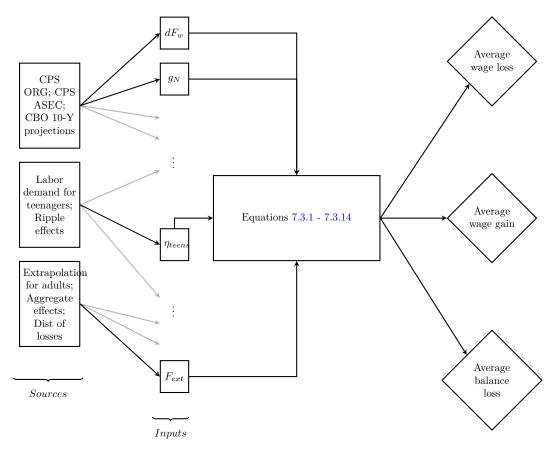


Figure 4.2.4: Simplified workflow diagram for case study on minimum wage

This diagram is a high level representation of how the analysis was carried out. In an upcoming book of case studies on how to achieve computational reproducibility in data-intensive sciences (Kitzes, J., Turek, D., and Deniz, F. (Eds.), 2017), several demonstrations are made of how the workflow representation can be even more transparent. To highlight that the workflow dimension can improve even further, it is scored at level 2.5 instead of 3.

Table 4.2.2: Sources and Inputes Used in Creation of Dynamic Document

Source	Input			
Data				
CPS ORG 2013	Number of salary workers in 2013			
(CEPR version)	$(\widehat{N_{final}^g} g \in \{teen, adult\})$			
	Fraction of workers below the new minimum wage			
	$(P_{\hat{w} \leq MW^1 g})$			
	Average wage variation for those below the new min			
	wage $(\sqrt[8]{\Delta w^g})$			
	Non-compliance rate (α_1^g)			
CPS ASEC 2012	Wages and Non-Wage Income distribution (dF_w, dF_{nw})			
(CEPR version)	Household size (N_h)			
	Hours/weeks worked (\hat{w}, \hat{h})			
State level Min. Wage (DOL)	Trends in state min. wage (MW_t^s)			
10-year economic forecast (CBO)	Predicted worker growth by 2016 (in 2013) $(\hat{g_N})$; Wage			
10-year economic forceast (CDO)	growth in by 2016 $(\hat{g_w})$; Non-wage growth by 2016 $(\hat{g_{nw}})$			
Research				
Elasticity of labor demand for	$\eta_{teen}^{lit} = -0.1$			
teenagers				
Ripple effects	From $R_{lb} = \$8.7$ to $R_{ub} = \$11.5$ with a "ripple" intensity of $R_I = 50\%$			
Guess Work				
Extrapolation factor from teenagers	$F_{ex} = 1/3$			
to adults				
Adjustment for effective wage	$F_{adj} = 4.5$			
variation and population				
Net benefits	$\hat{NB} = \$2billion$			
Aggregate consumption effects on	$\hat{OF} = 40,000 new jobs$			
employment				
Distribution of balance losses	$dBL = (1\%, 29\%, 70\%)$ if income $\in [0, 1PL, 6PL, +)$			
Fract. of wage losses used to pay	$F_{subs} = 1$			
wage gains				
Job killing process: fraction of jobs	Cut wages in half for twice the number of jobs destroyed			

Table 4.2.3: Components of Model and Policy Estimates Used in Creation of Dynamic Document

Model	Policy estimate		
	(per quintile)		
Predicted household income with and	Average gain in per		
without min wage increase.	capita income due		
Depends on:	to net wage		
$\widehat{N_{final}^g}, P_{\hat{w} \leq MW^1 g}, \overline{\% \Delta w^g}, \alpha_1^g,$	increase.		
$dF_w, dF_{nw}, N_h, \hat{w}, \hat{h}, MW_t^s, \hat{g}_N, \hat{g}_w, \hat{g}_{nw},$	$(\overline{WG_q})$		
$\eta_{teen}^{lit}, R_{lb}, R_{ub}, R_I, F_{ex}, F_{adj}, \hat{OF}$			
Predicted household income with and	Average loss in per		
without min wage increase.	capita income due		
Depends on:	to net wage		
$\widehat{N_{final}^g}, P_{\hat{w} \leq MW^1 g}, \overline{\% \Delta w^g}, \alpha_1^g,$	decrease.		
$dF_w, dF_{nw}, N_h, \hat{w}, \hat{h}, MW_t^s, \hat{g_N}, \hat{g_w}, \hat{g_{nw}},$	(WL_q)		
$\eta_{teen}^{lit}, F_{ex}, F_{adj}, \hat{OF}$			
Distribution of balance losses	Average loss in per		
Depends on : $\overline{WG_q}(\cdot), \overline{WL_q}(\cdot), \hat{NB},$	capita income to		
F_{subs}, dBL	balance wage gains.		
	$(\overline{BL_q})$		

4.2.2 Standard #2: Data

Original: Level 1

There are four data sets used in the analysis: The Current Population Survey (CPS) Outgoing Rotation Group (ORG), the CPS Annual Social and Economic Supplement (ASEC), the 10-Year Economic Projections from CBO, and the Department of Labor data on federal and state minimum wage. Although it is possible to infer from reading the report that those were the data used, they are not clearly labeled, no specific information is provided on which version of the data was used, and the data is not directly accessible from the report.

For example, there are three versions publicly available of the CPS: CPS raw files, ORG NBER and ORG CEPR. None is an exact copy of the other, so it was necessary to add an assumption regarding which data to use. The analysis was performed using the CEPR ORG data base.

A similar situation happened with the year of the data, as the report does not mention if the analysis is using data from 2013 or 2012. Given the time of publication (February 2014) and the dates of release of the CPS data, it was assumed that 2013 was available for the ORG file but not for the ASEC files.

In summary, the data is partially identified, not easily accessible, and not ready to apply a quick reproduction.

DD: Level 3

The original version of the report already had achieved level one of TR in the Sources standard. The additions made in the DD aimed to put the data and report together, and to allow for automatic update of the data sources.

For example, the call for the first data set (Current Population Survey - Outgoing Rotation Group 2013) now happens in section 2.1 of the DD where, for the first-time user, the data is downloaded from the web and requires a one line change of code to repeat the same exercise with different versions of the CPS ORG (different year or data repository).

The same methodology was applied for all four data sets used in the analysis (CPS ORG, CPS ASEC, State level minimum wage data base, and 10 year macroeconomic forecast from CBO). If an analyst would like to perform the same analysis but over different time periods, the analyst must

simply modify the year parameters in the data calls.

4.2.3 STANDARD #3: METHODS

Original: Level 1

The overall CBO methodology is a simulation approach. It consists of forecasting the distribution of wages and family income from 2013 to 2016, and imputing wage gains and job losses according to a set of parameters.

However, level one of TR still leaves many elements of the analysis unexplained and requires a large effort to reproduce qualitatively similar results. For example, a heavily debated issue after the publication of CBO's analysis was the effects on employment: 500,000 jobs lost. As discussed in section 2.3, most of the academic debate can be grouped into scholars that support an elasticity of labor demand of -0.1 (large effects on employment) and those who support estimates closer to -0.01 (small effects). A quick read of the report would suggest that the former estimates were the ones finally chosen (US Congressional Budget Office, 2014, page 25), but using this elasticity would produce an effect on employment on the order of 300,000 jobs lost¹. A more detailed read of the report suggested that this parameter was only applied to teenagers, while the elasticity for adults was adjusted by one third (page 28). Incorporating this into the analysis would render a job loss estimate closer to 100,000. A detailed review of the report (pages 26-28) would reveal an adjustment that follows Neumark and Wascher (2008) and Brown (1999) and would increase the elasticity of teenagers and adults by factors of 3.2 and 19.5 respectively, rendering 1.1 million jobs lost. Only after an exhaustive review of the report did it became clear that the factor used for the final adjustment was 4.5 for both populations (page 28), which renders the reported policy estimate of 500 thousand job lost.

The overall replication process described in the previous paragraph required several days of dedicated work, and the final effect on employment depended, in addition to the much debated elasticity, on other components that were largely ignored in the final report. A review of the technical discussion following the publication of CBO's report did not reveal any of the elements discussed above.

 $^{^1} Assuming target population <math display="inline">\approx 22$ million, and average wage increase for that population of $\approx 14\%,$ and non-compliance rate of $\approx 15\%$

DD: Level 2

In the methodological description, CBO leaves a few unexplained components that were either ignored or guessed, but are explicitly mentioned in the DD. Another dimension for improvement in the Methods standard is the structure of the report. Some analysis is quantitative while another fits a more narrative description, and sometimes it is not clear which is being described. Here the benefit of hindsight allowed for a focus on the key methodological components of the analysis that were discussed after the publication of the original report (focusing for example on only one wage increase option instead of two).

The DD also combines the methodological explanations with the code that applies those methods in each step (level 2), and allows the user/reader to see how the result reacts to changes in the parameters used (level 3).

Repeating the example of reproducing the effects on employment is only necessary to find the policy estimate (478,000 jobs lost) and in the same section the equation behind that calculation is presented, together with a table that contains all of the elements needed for such calculation.

Even though the DD increased TR substantially, there are still elements for improvement. Specifically, the coding component of this exercise still can be upgraded in two dimensions. First, in terms of readability, or use of syntax that allows non-R users to understand what is happening, it scores very low. Second, the users/readers can changes parameters, but have to execute the whole dynamic document to observe changes. One important lesson for future work is that wrapping almost all small chunks of code into functions makes the sensitivity analysis stage a much more straight forward process. Good examples that address these two issues can be found in Kitzes, J., Turek, D., and Deniz, F. (Eds.) (2017).

Chapter 5

Sensitivity Analysis

One of the main purposes of the dynamic document (DD) is that, after achieving high levels of transparency and reproducibility (TR), an arbitrarily large number of sensitivity analysis can be performed to assess how variations in *any* component of the analysis affect the final result.

Here I perform sensitivity analysis with two aims: (i) demonstrate a key feature of the DD, and (ii) taking the policy analysis from CBO as given, I explore which components of the analysis would benefit from more research. Readers/users of the DD are encouraged to perform, with minimal effort, different sensitivity analyses reflecting their own interests.

First I vary a few parameters and compare the output for all policy estimates. Second, to explore variations in all the components, I use the welfare function $(W(\rho))$ described in section 3.3, parametrized to represent policy makers with a preference in favor of redistribution $(\rho = 0.1)$ and one against it $(\rho = -0.1)$.

5.1 Sensitivity of a Few Parameters for all Policy Estimates

Figure 5.1.1 show how all of the final policy estimates (panel A) vary when one parameter is changed at a time. To facilitate the comparison the margins of the vertical axis are fixed at the same value across panels.

Given that most of the academic debate on minimum wage has been between the schools of large effects on employment (with an elasticity for teenagers of -0.1) and small effects on employment (elasticity of -0.01), a natural candidate for the first sensitivity analysis is the elasticity of labor demand for teenagers. As the value chosen in the CBO report analysis was -0.1, the sensitivity is to choose -0.01 (a 90% reduction from the original value). The results are presented in panel B. The overall picture is not that different. The wage losses (green) bar moves closer to zero for all quintiles, but its variation is small relative to the other two sources of change in income (wage gains and balance losses). Moreover, this reduction in wage losses by construction affects the other two policy estimates: more workers are receiving a wage increase, and the balance losses also increase to pay for such gain. The increase in balance losses is twice as large as the change in wage gains due to a key assumption in the model: the firms resources use from fired workers are used one-to-one to offset the increase in payroll. This is another example of a debatable assumption that was not clear in the original report.

Panel C shows how the policy estimates change when, relative to panel A, only the non-compliance rate changes from 15% to 22.5% (a 50% increase). An increase in non-compliance implies less people getting a wage increase, while maintaining constant the amount of workers who lose their jobs (again by assumption of the original model). This implies that families that see a net increase in earnings get a smaller gain across quintiles, and families that see a net loss get a bigger (than panel A) net loss. The overall resources used to pay for the wage gains are also reduced, henceforth the balance losses are also smaller. All the changes observed from panel A to panel C are larger in magnitude than the ones observed from panel A to B. The 50% increase in the non-compliance rate might be large, but there is little literature on the value of this parameter and its relationship with different levels of the minimum wage.

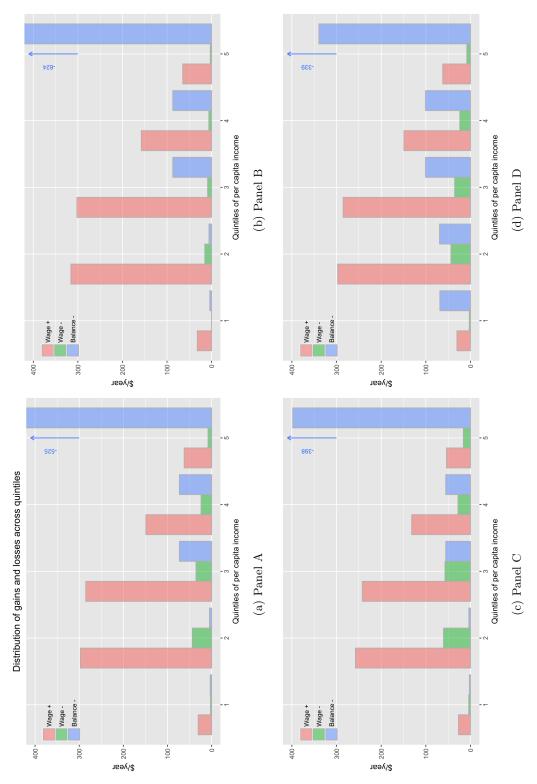


Figure 5.1.1: Panel A: Original policy estimates; Panel B: Change in Elasticity of Labor Demand: from $\eta_{lit}^{teens} = -0.1$ to $\eta_{lit}^{teens} = -0.01(\Delta^-90\%)$; Panel C: Change in non-compliance rate: from 15% to 22.5%; Panel D: Change in Distribution of Balance losses: from $(1PL,6PL) \sim (1\%,29\%,70\%)$ to (20%,40%,40%)

Panel D shows the policy estimates when the distribution of balance losses is changed from $(1\%, 29\%, 70\%)^1$, to (20%, 40%, 40%). In this case nothing changes with the net gains and net losses from wage variations, but the balance losses are now payed more heavily by individuals in quintiles one and two. Now individuals in quintile one are on average worse off with a raise in the minimum wage, and individuals in the second quintile are offsetting about one fifth of their gains in balance losses. This represents the largest variation in income for lowest quintiles across all for panels. This new distribution of losses is completely arbitrary, but is as arbitrary as the originally proposed one. Increasing our knowledge about this distribution of knowledge seems to be of high policy relevance.

The purpose of the brief sensitivity analysis performed in figure 5.1.1 is to illustrate how more transparency in the way in which the policy estimates are produced can increase our understanding of how sensitive the final output is to its underlying assumptions. To perform a more comprehensive sensitivity analysis it is necessary to reduce the dimensionality of the output. For this purpose I use the sample welfare function introduced in section 3.3 and study how hypothetical policy makers would change their valuations when all the underlying parameters of the analysis vary.

5.2 Sensitivity of all Parameters on a Sample Welfare Function

Next, the sensitivity analysis is carried out over a larger set of the inputs identified in the DD (table 4.2.2) and its effect is evaluated over the hypothetical preferences of two types of policy makers: one that favors redistribution and one that does not $(W(\rho = 0.1))$ and $W(\rho = -0.1)$ of equation 3.3.1 respectively).

For each parameter, two types of perturbations were applied: a 10% increase and a 10% decrease from its current level. This methodology was applied to all of the parameters in the table with the exception of those that describe the distribution of balance losses. In the sensitivity analysis, the parameters for the distribution of balances losses were chosen to reflect three scenarios: a higher share of the losses paid by the wealthy (1%, 4%, 95%), a slightly less "progressive" loss scheme (5%, 35%, 60%), a flat distribution of losses (40%, 54%, 6%), and a uniform distribution of balance

 $^{^{1}1\%}$ the total wage gain is paid by individuals with income below the poverty line, 29% by those with income between one and 6 poverty lines, and 70% by those with incomes greater than 6 poverty lines

losses.

Table 5.2.1: $\%\Delta W$ for a $\%\Delta$ in inputs. Two sample policy makers: dislikes (W(-0.1) = -\$5.7bn) and likes (W(0.1) = \$9.7bn) redistribution

	Re-distributional Preferences			
	Dislikes $(\rho = -0.1)$		Likes $(\rho = 0.1)$	
Source Input	$10\%\Delta^+$	$10\%\Delta^-$	$10\%\Delta^+$	$10\%\Delta^-$
Data				
Annual wage growth (g_w)	-3%	2%	-2%	1%
Annual growth in N	0.8%	-0.7%	0.5%	-0.4%
Research				
η_{teen}	-4%	4%	-2%	2%
Ripple Scope $(8.7, 11.5)$	37%	-23%	21%	-13%
Ripple Intensity $(50\%\Delta w)$	5%	-5%	3%	-3%
Guess Work				
Extrapolation factor (F_{ex})	-2%	2%	-1%	1%
Non compliance (α_1)	-7%	7%	-4%	4%
Substitution factor (F_{sub})		20%		-8%
Net benefits	-4%	4%	2%	-2%
Distribution of balance losses		'		
Current: $(1\%, 29\%, 70\%)$				
(1%, 4%, 95%)	22%		13%	
(5%, 35%, 60%)	-16%		-9%	
1/N	-127%		-72%	

Table 5.2.1 suggests that, in addition to the elasticity of labor demand, there are many other components of the policy analysis that can have consequential effect on the final decision made by policy makers.

For example, changes in the current values of the ripple effects inputs would play a pivotal role in either of the two hypothetical positions. A 10% variation in the scope parameters (the range of the ripples), from (8.7, 11.5) to (7.8, 12.7) would increase the policy makers' valuations by 37% for those against and by 21% for those in favor of raising the minimum wage.

It is important to acknowledge that the academic debate around the elasticity of labor demand would represent a 90% reduction from the current value used in the analysis. This implies changes in valuations around 36% and 18%, making it a consequential debate for policy purposes. What the results in table 5.2.1 suggest is that other components play an equally important role, and yet much less is known about them.

The ripple effects could have a much larger or narrower scope and intensity, as the literature around

it is scarce. All of the parameters in the guess work category can have a much wider range as almost nothing is known about them.

To explore further the issue of how the welfare valuations would change for different ranges in the parameters I use again as an example the components involved in computing the final elasticity of labor demand. As the DD shows; three components are needed to compute the final elasticity: the elasticity of labor demand for teenagers, the extrapolation factor from that elasticity to the elasticity of adults, and a adjustment factor that aims to correct for the target population (a detailed discussion of all three can be found in section 2.3.2 of the DD).

For these three components I compute the sample welfare function for values in favor of and against redistribution. The ranges of each factor are chosen to reflect plausible values. The extrapolation factor chosen in the CBO report is 1/3 based on the argument that the adults' elasticity is smaller. As no clear justification for that reason could be found, I set a range from 0 to 1.5 to reflect that it most likely is smaller, but could also be equal or slightly greater. The adjustment factor is much harder to justify as the chosen value of the report (4.5) differs drastically from its value in the data (~ 18) . For this reason, the range is allow to vary from 1 (no adjustment) to 20. Finally the range for the elasticity of teenagers is set from 0 to 0.2. Values above 0.2 would imply, holding the other parameters constant, a number of job losses of about 1 million workers, a magnitude equivalent to half a percentage point of the national unemployment rate (from 4.7% to 5.3%). This informal type of bounding can be make much more rigorurous when complemented with some of Manski's prescriptions discussed in the introduction and detailed in Manski (2009).

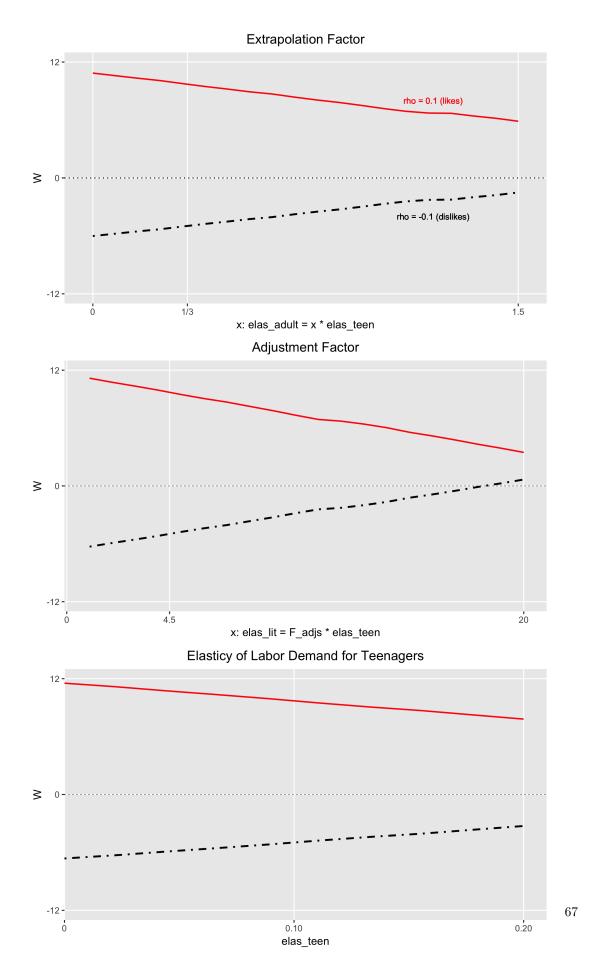


Figure 5.2.1: Value of welafare function for sample position in favor of redistribution $(W(\rho=0.1))$ and agains redistribution $(W(\rho=0.1))$, for three parameters in the analysis: extrapolation factor (top panel), adjustment factor (middle panel), and elasticity of labor demand for teenagers (lower panel).

Figure 5.2.1 shows how the welfare function, for $\rho = 0.1$ (red) and $\rho = -0.1$ (dashed black), varies for all three components. The vertical axis was fixed to compare the slope of each curve across plots. The key takeaway from this figure is that all three components seem to change the welfare function in similar fashion, but the discussion related to dis-employment effects is disproportionately occupied by the value of the elasticity of labor demand.

Figures 5.1.1 and 5.2.1 and table 5.2.1 suggest that if all gains and losses are weighted equally, the academic discussion between an elasticity of -0.1 or -0.01 might not be as policy relevant. Alternatively, the debate around the elasticity of labor demand for teens could be compatible with a very high weight on those specific type of losses.

In the last sensitivity analysis I explore further how the overall welfare, of the two sample policy maker, changes when the distribution of losses changes. Figure 5.2.2 shows for each panel a distribution of balance losses, and the two numbers inside the plots are the welfare function values for each sample position. The left panel is the original distribution of losses as proposed by CBO. A plausible explanation for this distribution is that most of the wage gain is paid by a reduction in profit by wealthy business owners, and a small fraction of the gains is paid either by a small increase in general prices or a reduction in profits of middle income business owners. The first panel on the right (top-right) is a more extreme version of this, where individuals with a income above 6 poverty lines pay almost the entirety of the wage increase.

The second, third and fourth panels from the right (top to bottom) represent a version where the distribution of balance losses is paid either less progressively, uniformly, or slightly regressively. This could happen if the wage increase is finally paid by an increase in prices of goods that are more commonly consumed by population of lower income. There is not much evidence to help us choose the scenario that best represents the effects of raising the minimum wage in the US, and the changes in the sample welfare functions are larger than any of the changes discussed in the previous sensitivity analysis.

If the relative valuations of each type of gain and loss are relatively similar, the policy debate could benefit from additional research in elements like non-compliance rates, adjustment factors, scope and intensity of ripple effects, and the distribution of balance losses used to pay for the increase in wages. It seems that the most consequential parameters are the ones that describe the distribution of balance losses. Very little is known about who finally carries the burden of a raise in the minimum

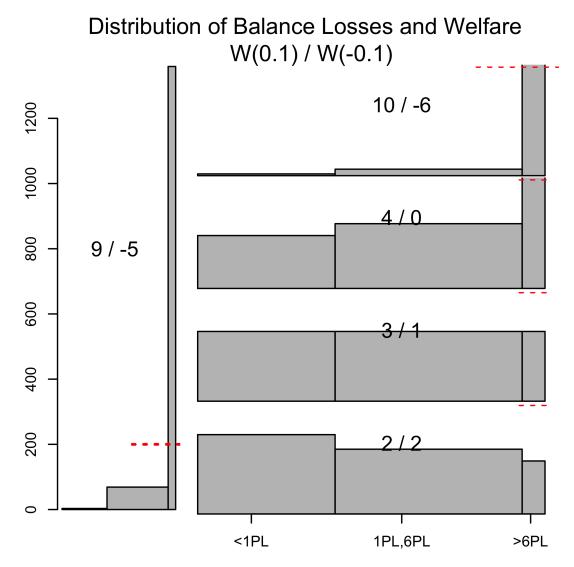


Figure 5.2.2: Different distribution of balance losses. Width of each bar represents population size, height is the average amount paid by each person in each bin. The numbers inside each panel represent the welfare function for values $W(\rho=0.1)/W(-\rho=0.1)$. Panel on the left is the distribution used in the original report. Upper right is a distribution where losses are more progressive (1%, 5%, 94%). The second panel on the right has a less progressive distribution (30%, 50%, 20%). The second-to-last panel has a uniform effect over the population (40%, 54%, 6%). And the bottom right panel has a slightly regressive distribution of losses (45%, 50%, 05%)

wage. Understanding this distribution should be an priority in a evidence-based policy debate around the minimum wage.

MaCurdy (2015) analyses the distributional effects of the minimum wage, using a input-output consumption approach combined with a simulation model. His simulations suggest that the burden of balance losses are paid in a larger share by wealthy households, but in a smaller magnitude than the one assumed in the CBO analysis. Harasztosi et al (2015) estimates the effects of a large minimum wage increase in Hungary and concludes that most of the raise is passed to consumers through prices. To estimate the effect of raising the minimum wage on the final distribution of balance losses the simulation approach (MaCurdy) could be combined with the effects estimated on workers and firms (Harasztosi and Lindert)

Finally, the sensitivity analysis performed here represents only a small fraction of all the variation that can be studied once, particularly if combinations of variations are considered. The goal is to motivate readers to clone their own version of the DD and perform multiple analyses of interest. This is one of the main benefits of the open-source feature of the dynamic document developed for this dissertation.

Chapter 6

Conclusions and Extensions

6.1 Conclusions

This dissertation translates and implements guidelines and tools for transparency and reproducibility (TR) in science into policy analysis. The goal of this methodological innovation was to propose an additional solution in response to the critique of low credibility of policy analysis (Manski, 2013). Increasing TR in policy analysis increases its credibility, helps to provide a clear connection between research and policy analysis, and allows for systematic improvement and automation in specific and recurrent policy analyses. Drawing a parallel between the reproducibility/credibility crisis in science made it possible to identify similar solutions. After translating standards of TR into the policy analysis setting, as a case study I implemented the highest level of TR into a report from the Congressional Budget Office on the effects of raising the minimum wage.

The results from the case study show how a dynamic document in open source format, the highest standard of TR, can draw a clear connection between research inputs and the output of a policy analysis. The transparency component helped to identify potential weaknesses in the original policy analysis, and the reproducibility component allowed for a comprehensive sensitivity analysis that shed light on where new knowledge could be the most valuable. As an example, the sensitivity analysis suggests that, from the perspective of policy relevancy, the research agenda on the effects of minimum wage on employment is over-studied relative to other areas like the distribution of losses

used to pay for the increase in wages.

The open source nature of the dynamic document aims to provide the foundation for a constructive debate around the technical issues of a recurrent policy issue. It allows for updates in short term and long term dimensions. The vast majority of policy analysis is to "put down fires", i.e. to address pressing issues that need some type of immediate solution. For this reason the technical contributions in the short term category should accept the proposed model as the correct one and oversee its correct implementation. In parallel, contributors to the DD can propose structural modifications to the analysis, to be addressed in the next cycle of the policy debate.

If this approach to policy analysis were to become the status quo, benefits would be observed in at least four dimensions. First, the cost of producing the next (or marginal) report would be reduced substantially. Second, any modification to the original report would be incremental, as opposed to arbitrary, as the policy analysis debate discussion would now have a framework to systematically improve upon previous work. Third, as multiple analyses adopt this approach, and different reports become "living documents", updating a shared parameter (for example: estimates regarding the deadweight loss of a tax/subsidy or estimates about the statistical value of a year of life) could easily be implemented across reports. And fourth, as clarity is added to the positive elements of a policy discussion, it would be easier to have a normative policy debate that clarifies the positions of different policy makers.

Finally, a key issue that is not addressed in this dissertation is who should be responsible for the implementation of this approach. Policy analysts face strong time and resource constraints, and adding a set of protocols and techniques seems difficult. Those best suited to this task should have a less stringent resource constrain, but be closely related to the policy issue and analysis. Possible candidates are banks of knowledge as proposed by Clemens and Kremer (2016), public policy schools, think tanks or expert commissions.

6.2 Extensions within the current case study (minimum wage)

- Further disaggregate the CBO results: The results presented in the original CBO report, and reproduced here, are presented at the national level, and aggregated across all demographics. As the dynamic document (DD) allows for one click reproducibility and the open source format allows for easy modifications, it is possible to expand the analysis to subgroups. This way, estimates of the three effects (wage gains, wage losses, and balance losses) could be displayed at a state or county level, separating also by gender and/or age groups.
- Expand margins of interest: Currently the analysis estimates effects on wages (directly and through ripple effects), employment, and the overall effect on the economy. However, there are multiple other dimensions, or margins, where raising the minimum wage could have relevant policy implications. Clemens and Wither (2014) present a non-exhaustive list of possible margins of interest, in addition to the ones studied here, which include; income trajectories of low-skilled workers; firm offerings of benefits including health insurance; firm spending on the quality of workplace conditions; firm substitution between low-skilled labor, high-skilled labor, and capital; firm utilization of inputs with which low-skilled labor is complementary; incomes of firm owners; and prices of goods produced by firms that employ minimum wage workers. Long term revisions of the current DD could add such dimensions into the analysis.
- Interactive policy tool: Open source DDs allows analysts to download the code and modify it according to their interests. However, this requires knowledge in computer programming and a clear understanding of how the analysis works. To address this limitation, the R package shiny complements the DD with a reactive web-based tool that allows users to modify different components of the analysis and see immediately how those changes affect the final output. For a proof of concept click here¹ to see a toy example with simulated data. This tool can be targeted to four types of audiences: researchers, policy analysts, decision makers, and the general public. With this tool, different stakeholders could make explicit their subjective choice of some of the inputs in the analysis.

¹https://fhoces.shinyapps.io/example_min_wage/

6.3 General extensions to TR in policy analysis

• Connecting TR in policy analysis and research: An immediate connection can be made between TR in policy analysis and the transparency and openness framework in science: adding TR to the way parameters from research are selected into the policy analysis framework (figure 2.2.5). In the case study presented here, that would involve a meta-analysis that combines the different studies of the minimum wage, each with specific weights, and a prior distribution (Gelman et al, 2014) to produce the value of -0.1.

Performing meta-analyses also requires a significant effort in identifying the studies, standarizing effects and modeling the average effect appropriately. For the field of development economics, Vivalt (2015) developed a website (aidgrade.org) that aims to collect standarized effects from all the literature, and Rising and Hsiang (2014) developed a web-based tool (dmas.berkeley.edu) that allows researchers to make explicit and reproducible all of the modeling decisions described in the above meta-analysis. Combining high TR in meta-analysis and high TR in policy analysis would allow for simulation of these effects of potential new research (effect size and precision) on the final policy estimates. These two components combined would help to guide the allocation of future research resources based on the potential effect on policy estimates.

- Combining methodologies to increase credibility in policy analysis: As described in the introduction, Manski (2013) outlines the credibility critique on policy analysis, but also provides a set of prescriptions to address the problem. Achieving high TR would facilitate the display of standard errors and performing analysis of the boundaries of the policy estimates (Manski, 2009).
- Formalize the argument of sub-optimal provision of TR in policy analysis. Write a more formal model that characterizes the optimal level of transparency from an individual and societal perspective (as ambiguity increases, policy analysts and policy makers have a higher incentive to focus on non-productive cues-prestige, gender, race, etc- to approximate the quality of the information).
- Track policy maker's choices of "what the facts are": Once an interactive tool is available, this could allow for better tracking of how policy makers are making choices over time. In an ideal scenario, it is possible that different policy makers involved in an issue

could be asked not only to submit their vote towards a specific alternative, but also the set of parameters that they took as "the facts". Doing this over time would reveal information regarding policy maker's choices and valuations (for example: a policy maker might choose high levels of informality/non-compliance when voting for the minimum wage, but low levels of informality when voting for another policy in the future)

6.4 Extensions to other possible policy analyses

Other opportunities for increasing TR in policy analysis are plentiful. Here I focus on examples from organizations/authors that have pioneered the promotion of transparent analysis to make the point that if it can improve among those who are already in the frontier of TR, it can also benefit analyses that are less transparent and reproducible.

• Cost Benefit Analyses for interventions in the developing world: Two organizations, Givewell (a charity curator) and the Jamal Poverty Action Lab (a network of researchers), have made publicly available the spreadsheets used to perform their respective cost benefit analyses. Their spreadsheets are remarkably well annotated and most of the analysis in each cell is in formulas ("code") as opposed to hard coded. This allows for third parties to track how the CBAs are performed and modify key components.

However, as demonstrated with the case study here, we can use the guidelines and tools to increase TR such that all of the analysis is in one single document (as opposed to multiple spreadsheets), with a clear narrative explanation joined with equations that are implemented in code in the same file (as oppose to having to click in each cell to trace the role of each formula when available).

• Health care and appropriateness of care: The U.S. Preventive Services Task Force (USP-STF) publishes detailed micro-simulation studies to assess the effects of different type of medical treatments recomendations (see for example Agency for Healthcare Research and Quality (2015) for colorectal cancer screening). These micro-simulations are described in great detail, and are recurrently updated. However, each micro-simulation study still requires a number of arbitrary decisions that are not always documented, the code is either absent or not designed for reproducibility.

• Regulatory look-back approach from Michael Greenstone. Greenstone (2009) advocates for comprehensive reform of how regulations are assessed, emphasizing the role of evidence and periodic cost benefit analyses. This approach had a strong influence in the Office of Information and Regulatory Affairs (Sunstein, 2014), where cost benefit analyses are now applied to upcoming and incumbent regulations. But instead of reassessing the effects of regulations every few years, a fully TR approach would allow for an almost instantaneous update of the effects of regulations as multiple components of the policy analysis evolve over time.

Finally, it is worth mentioning that practices similar to the ones outlined in this dissertation are increasingly popular in the data science departments of many private companies. Google for example famously hosts all of its code in one big repository that can be accessed by any data scientist in the company. Airbnb more recently announced that in addition to internal requirements of TR, it would publish some of their internal analyses and tool for public use (Sharma and Overgoor, 2017). All of these companies for competitive reasons cannot disclose all of their analyses, but here government agencies within and across countries do not suffer from such limitations.

Chapter 7

APPENDICES

7.1 SUMMARY OF TOP GUIDELINES

	Level 0	Level 1	Level 2	Level 3
Citation Standards	Journal encourages citation of data, code, and materials, or says nothing	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used consistent with journal's author guidelines.	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
Data Transparency	Journal encourages data sharing, or says nothing	Article states whether data are available, and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Analytic Methods (Code) Transparency	Journal encourages code sharing, or says nothing	Article states whether code is available, and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Research Materials Transparency	Journal encourages materials sharing, or says nothing	Article states whether materials are available, and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a frusted repository, and reported analyses will be reproduced independently prior to publication.
Design and Analysis Transparency	Journal encourages design and analysis transparency, or says nothing	Journal articulates design transparency standards	Journal requires adherence to design transparency standards for review and publication	Journal requires and enforces adherence to design transparency standards for review and publication
Preregistration of studies	Journal says nothing	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration of study exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Preregistration of analysis plans	Journal says nothing	Article states whether preregistration with analysis plan exists, and, if so, where to access it.	Article states whether preregistration with analysis plan exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies, or says nothing	Journal encourages submission of replication studies	Journal encourages submission of replication studies and conducts results blind review	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

Figure 7.1.1: Reproduced from: https://cos.io/our-services/top-guidelines/

7.2 Checklist for TR in Policy Analysis

Standard #1: Workflow
□ Level 0
☐ Policy estimates vaguely described
□ Level 1
\square All the inputs, and their corresponding sources (data, research and gues-work) used in the calculations are listed
☐ Level 2
\square Lvl 1 + Policy estimates are listed, in same unit if possible
☐ Level 3
\square Lvl 2 + All the components can be modified with little effort
Standard #2: Data
□ Level 0
☐ Report says nothing
□ Level 1
\square Clearly stated whether all, some components, or none of the data is available, with instructions for access when possible.
\square Level 2
\square Report and data are in same place
\square Level 3
☐ Report has specific lines of code that call the data and changes in the data produce traceable changes in the report

Standard #3: Methods
\square Level 0
\square Key assumption are listed
☐ Level 1
\square Methods are described in prose. Large amount of work is required to reproduce qualitatively similar estimates
☐ Level 2
\square Methods and described in prose, with detailed formulas, and code is provided as supplementary material
□ Level 3
\square Lvl 2 + All is in the same document where changes in the code affect the output automatically

7.3 Equations that define Model in CBO Report

This appendix describes all the equations behind the model that produces the three policy estimates: average wage gain, average wage losses, and average balance losses per quintile. Most of these equations are already in the Dynamic Document, but are all put in the same place here for ease of reference. For the values used for the different components, the reader can consult the DD or table 4.2.2.

First I describe how the effects on employment were computed, then the effects on wages and family income, and the distribution of balance losses (used to pay for the raise in wages). For all estimates, the first step is to predict the distribution of wages using data from a year t into a year in the future t' (t' is always 2016, but t varies for employment and income). I use $\hat{w'}$ to denote the predicted wages. CBO predicted different wage growth rates across deciles of income, and whenever a wage was below the predicted state minimum wage (I use the effective minimum wage of 2016), that wage was replaced by its state minimum.

7.3.1 Employment

The number of jobs lost due to the increase in minimum wage $(\widehat{\Delta E})$ at general level has the following form:

$$\widehat{\Delta E} = N \times \eta \times \% \Delta w + \text{Other factors}$$
(7.3.1)

CBO performs the analysis over two sub-groups (indexed with the letter s): teenagers (ages 19 or less) and adults (ages 20 or more). The final N used for each group is defined as the interaction of the following components: working population in year t = 2013 (N_t^s), predicted growth of the population from year t = 2013 into year t' = 2016 ($g_N(\hat{t}'|t)$), fraction of the population with a hourly wage below the new minimum wage ($P(\hat{w} \leq MW^{new}|s)$), non-compliance rate (α_1^s), and fraction of workers non subject to the Fair Labor Standards Act (α_2^s). Thus, for each subgroup the final population is

$$\widehat{N_{final}^s} = g_N(\hat{t}'|t) \times \hat{N_t^s} \times P(\hat{w'} \le MW^{new}|s) \times (1 - \hat{\alpha_1^s} - \hat{\alpha_2^s})$$
 $s = \{teens, adults\}$ (7.3.2)

The elasticity for teenager (η_{lit}^{teen}) is described in the report as obtained from the literature (η_{lit}^{0}) and adjusting for publication bias, and a larger variation in the minimum wage $(F_{pub.bias}, F_{large.variation})$

$$\eta_{lit}^{teen} = \eta_{lit}^{0} \times F_{pub.bias} \times F_{large.variation} \tag{7.3.3}$$

The elasticity for adults from the literature is define as the one for teenagers with an extrapolation factor.

$$\eta_{lit}^{adults} = \eta_{lit}^{teens} \times F_{extrapolation}$$
(7.3.4)

Then the CBO suggests an adjustment following Neumark and Wascher (2008); Brown (1999). First separate elasticities for workers above and below the new minimum wage.

$$\eta_{lit}^s = p_{w \leq MW}^s \eta_{w \leq MW}^s + (1 - p_{w \leq MW}^s) \eta_{w > MW}^s \qquad \qquad s = \{teens, \, adults\}$$

Second, assume $\eta_{w \leq MW}^s = 0$:

$$\eta^s_{w \leq MW} = \frac{\eta^s_{lit}}{p^s_{w \leq MW}} \qquad \qquad s = \{teens, \, adults\}$$

And third, adjust for the effective average wage variation for each group $(\overline{\%\Delta w^s})$:

$$\widetilde{\eta_{w \leq MW}^s} = \frac{\eta_{lit}^s}{p_{w < MW}^s} \times \frac{\% \Delta MW}{\% \Delta w^s} = \eta_{lit}^s \times F_{adjs}^s \qquad s = \{teens, adults\}$$
 (7.3.5)

Combining equations 7.3.5, 7.3.1, with other factors discussed in the report through which there is a positive effect on employment through an increase in aggregate demand (\widehat{OF}) , we get that the estimated final effect on employment is:

$$\widehat{\Delta E} = \sum_{g \in \{A,T\}} \left(\widehat{N_g^{final}} \times \widehat{N_w^{g}} \times \widehat{N_w^{g}} \right) - \widehat{OF}$$
(7.3.6)

7.3.2 Wages

Now all the elements are predicted and should have a " $(\hat{\cdot})$ " symbol, so we can simplify that notation, and the reader should remember that all quantities below are estimates.

The wages in to the future under the status quo are defined as w'. The wages under the new minimum wage are labeled as w'', and simulated in the following nested way:

First, some wages remain the same for the fraction of non-compliers $(alpha_1)$ implemented as follows, and for the fraction of compliers there is a new wage (w^{new}) :

$$w'' = \begin{cases} w' & if \quad w \in U[0,1] < \alpha_1 \\ w^{new} & o/w \end{cases}$$
 (7.3.7)

Where U[0,1] represents a random draw, for each individual, from a uniform [0,1] distribution.

Then we implement the wage reduction due to job loss. Following CBO's methodology I reduce wages in half for twice the number of jobs destroyed $(2\Delta E)$. The job destruction procedure was implemented in similar fashion as for equation 7.3.8. I define the fraction $alpha_{aux} = 2\Delta E/N_{w < MW'}$

and repeat:

$$w^{new} = \begin{cases} w'/2 & if \quad w \in U[0,1] < \alpha_{aux} \\ \widetilde{w^{new}} & o/w \end{cases}$$
 (7.3.8)

Finally the wages of the remaining workers are modified according to the following rules: if wages are below the of the ripple effects scope parameter (R_{lb}^s) they are replaced by the new minimum wage (MW'); if wages are between the scope of ripple effects (between R_{lb}^s and R_{ub}^s) they are replace by the new minimum wage plus a ripple effect of R^I ; and if the wages are above the scope of ripple effects (R_{ub}^s) they are not changed.

$$\widetilde{w^{new}} = \begin{cases}
MW' & if \quad w' < R_{lb} \\
MW' + R^{I}(w' - R_{lb}^{s}) & if \quad w' \in [R_{lb}, MW') \\
w' + R^{I}(R_{ub}^{s} - w') & if \quad w' \in [MW', R_{ub}) \\
w' & o/w
\end{cases}$$
(7.3.9)

7.3.3 INCOME

With a fully simulated wage under the status quo (w') and under the new policy (w''), we no construct the household income and compare them.

The household income is defined as the sum of all wages and non-wages income for all members in a family. The non-wages earnings are the ones reported in year t, and with a predicted growth (different from wage growth) of $g_{nw}(t'|t)$. For each individual i in a household h, the per-capita income under the status quo (y'_h) and under the new policy $(y''_{i,h})$ are:

$$y'_{i,h} = \sum_{i \in N_h} (g_{nw}(t'|t)nw_i + w'_i)/N_h$$

$$y''_{i,h} = \sum_{i \in N_h} (g_{nw}(t'|t)nw_i + w''_i)/N_h$$
(7.3.10)

Now we say that a individual has recieves a net wage gain if the per-capita household income is greater with the new minimum wage policy. Conversely net wage loss occurs when the net per-capita household income is lower

$$WG_i = (y_i'' - y_i') \mathbf{I}(y_i'' > y_i')$$
(7.3.11)

$$WL_i = (y_i' - y_i'') \mathbf{I}(y_i'' < y_i')$$
(7.3.12)

Finally we define the total balance losses (BL) as all the resources needed to pay for the wage increase. Assuming that a fraction F_{subs} of the resources from laid-off workers is used to pay for the balance losses, and a distribution of balance losses across the population (dBL)

$$BL = \sum_{i} WG_{i} - F_{sub} \sum_{i} WL_{i}$$

$$BL_{i} = BL \times dBL$$
(7.3.13)

With all these elements we now can compute each policy estimate averages across quintiles (Q) of income as:

$$\overline{WG_Q} = \frac{\sum_{i \in Q} WG_i}{N_{pop}/5}$$

$$\overline{WL_Q} = \frac{\sum_{i \in Q} WL_i}{N_{pop}/5}$$

$$\overline{BL_Q} = \frac{\sum_{i \in Q} BL_i}{N_{pop}/5}$$
(7.3.14)

These are the quantities displayed in the final visualization 4.2.3.

7.4 KEY DEFINITIONS

Box 1: Key Definitions

Key definitions:

- Research Estimate: Answer to a specific research question (what is the effect on employment in the fast food industry of an increase in the minimum wage from \$4 to \$ 4.25 dollars per hour in New Jersey in 1992). In the public policy context, this estimate is used as an input of the policy analysis.
- Policy Estimate: Output of the policy analysis. Typically a cost and/or benefit estimate of
 one alternative associated with a specific policy (i.e. according to US Congressional Budget
 Office (2014) an increase of of the federal minimum wage to \$10.10 per hour would destroy
 500,000 jobs and increase the wage of 16.5 million workers)
- Policy Choice: Specific policy plan (i.e. raise the federal minimum wage to \$10 per hour in all states by 2015)
- Policy relevant information: policy estimate that increases the relative merits of a particular policy alternative, that passes basic reality checks, and that appears to have good predictive power. The reported information will also vary in the degrees of ambiguity and documented uncertainty.
- Ambiguity: lack of testable or quantifiable adjectives in a statements (i.e. "There is a good chance that it will rain tomorrow")
- Uncertainty: probabilistic statement of a state of the world (i.e. "There is a 80% chance that it will rain tomorrow")

7.5 Printed version of the dynamic document

Will be attached in final version of the dissertation. It is currently available online and printed is about 50 pages long, so will omit until archiving the dissertation.

BIBLIOGRAPHY

- Abbasi K (2014) The missing data that cost \$20 bn. BmJ 348:g2695
- Agency for Healthcare Research and Quality (2015) Evaluating the benefits and harms of colorectal cancer screening strategies: A collaborative modeling approach
- Anderson MS, Martinson BC, De Vries R (2007) Normative dissonance in science: Results from a national survey of us scientists. Journal of Empirical Research on Human Research Ethics 2(4):3–14
- Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature 533(7604):452-454
- Brodeur A, Lé M, Sangnier M, Zylberberg Y (2016) Star wars: The empirics strike back. American Economic Journal: Applied Economics 8(1):1–32
- Brown C (1999) Minimum wages, employment, and the distribution of income. Handbook of labor economics 3:2101–2163
- Buckheit JB, Donoho DL (1995) Wavelab and reproducible research. Springer
- Camerer CF, Dreber A, Forsell E, Ho TH, Huber J, Johannesson M, Kirchler M, Almenberg J, Altmejd A, Chan T, et al (2016) Evaluating replicability of laboratory experiments in economics. Science 351(6280):1433–1436
- Card D, Krueger AB (2015) Myth and measurement. Princeton University Pres
- Chang AC, Li P (2015) Is economics research replicable? sixty published papers from thirteen journals say'usually not'

- Christensen G, Soderberg C (2015) Manual of best practices in transparent social science research.

 Berkeley, CA: University of California Retrieved April 23:2016
- Clemens J (2015a) The minimum wage and the great recession: Evidence from the current population survey. Tech. rep., National Bureau of Economic Research
- Clemens J, Wither M (2014) The minimum wage and the great recession: Evidence of effects on the wage distributions, employment, earnings, and class mobility of low-skilled workers. NBER Working paper (w20724)
- Clemens MA (2015b) The meaning of failed replications: A review and proposal. Journal of Economic Surveys
- Clemens MA, Kremer M (2016) The new role for the world bank. The Journal of Economic Perspectives 30(1):53–76
- Collaboration OS, et al (2015) Estimating the reproducibility of psychological science. Science 349(6251):aac4716
- Doberstein C (2016) Whom do bureaucrats believe? a randomized controlled experiment testing perceptions of credibility of policy research. Policy Studies Journal
- Dube A, Lester TW, Reich M (2010) Minimum wage effects across state borders: Estimates using contiguous counties. The review of economics and statistics 92(4):945–964
- Dunn WN (2015) Public policy analysis. Routledge
- Franco A, Malhotra N, Simonovits G (2014) Publication bias in the social sciences: Unlocking the file drawer. Science 345(6203):1502–1505
- Gelman A, Carlin JB, Stern HS, Rubin DB (2014) Bayesian data analysis, vol 2. Taylor & Francis
- Gerber A, Malhotra N, et al (2008) Do statistical reporting standards affect what is published? publication bias in two leading political science journals. Quarterly Journal of Political Science 3(3):313–326
- GiveWell (2015) Givewell's cost-effectiveness analyses | givewell. http://www.givewell.org/international/technical/criteria/cost-effectiveness/cost-effectiveness-models, (Accessed on 04/14/2016)

- Goodman SN, Fanelli D, Ioannidis JP (2016) What does research reproducibility mean? Science translational medicine 8(341):341ps12–341ps12
- Gray J (2014) Towards a genealogy of open data. In: The paper was given at the General Conference of the European Consortium for Political Research in Glasgow
- Greenstone M (2009) Toward a culture of persistent regulatory experimentation and evaluation. New perspectives on regulation 113:114
- Harasztosi P, Lindner A, Bank MN, Berkeley HU (2015) Who pays for the minimum wage? UC Berkeley
- Harvard Magazine (2016) Douglas elmendorf on progressive policy and economics | harvard magazine. http://harvardmagazine.com/2016/09/a-moral-conscience-for-economics, (Accessed on 04/19/2017)
- Hewlett RG, Anderson OE (1962) The new world, 1939/1946: A history of the united states atomic energy commission, vol. 1. university park
- Holtz-Eakin D (2010) The real arithmetic of health care reform the new york times. http://www.nytimes.com/2010/03/21/opinion/21holtz-eakin.html, (Accessed on 05/08/2017)
- Hsiang SM, Burke M, Miguel E (2013) Quantifying the influence of climate on human conflict. Science 341(6151):1235,367
- Ioannidis JP (2005) Why most published research findings are false. PLoS Med 2(8):e124
- Kerr M, Bedford M, Matthews B, O'Donoghue D (2014) The economic impact of acute kidney injury in england. Nephrology Dialysis Transplantation 29(7):1362, doi: 10.1093/ndt/gfu016, URL +http://dx.doi.org/10.1093/ndt/gfu016, /oup/backfile/Content_public/Journal/ndt/29/7/10.1093_ndt_gfu016/3/gfu016.pdf
- Kitzes, J, Turek, D, and Deniz, F (Eds) (2017) The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences. University of California Press.
- Knuth DE (1992) Literate programming. CSLI Lecture Notes, Stanford, CA: Center for the Study of Language and Information (CSLI), 1992 1
- Leamer EE (1983) Let's take the con out of econometrics. The American Economic Review 73(1):31–43

- MaCurdy T (2015) How effective is the minimum wage at supporting the poor? Journal of Political Economy 123(2):497–545
- Manski CF (2009) Identification for prediction and decision. Harvard University Press
- Manski CF (2013) Public policy in an uncertain world: analysis and decisions. Harvard University Press
- Mas-Colell A, Whinston MD, Green JR, et al (1995) Microeconomic theory, vol 1. Oxford university press New York
- Merton RK (1973) The sociology of science: Theoretical and empirical investigations. University of Chicago press
- Miguel E, Camerer C, Casey K, Cohen J, Esterling KM, Gerber A, Glennerster R, Green DP, Humphreys M, Imbens G, et al (2014) Promoting transparency in social science research. Science 343(6166):30–31
- Neumark D, Wascher WL (2008) Minimum wages. MIT Press
- Nosek B, Alter G, Banks G, Borsboom D, Bowman S, Breckler S, Buck S, Chambers C, Chin G, Christensen G, et al (2015) Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. Science (New York, NY) 348(6242):1422
- Nutley SM, Walter I, Davies HT (2007) Using evidence: How research can inform public services. Policy press
- Obama B (2009) Transparency and open government. Memorandum for the heads of executive departments and agencies
- Peng RD (2011) Reproducible research in computational science. Science (New York, Ny) 334(6060):1226
- Rising JA, Hsiang SM (2014) A tool for distributed meta analysis
- Rosenthal R (1979) The file drawer problem and tolerance for null results. Psychological bulletin 86(3):638

- Rothstein J (2015) Teacher quality policy when supply matters. American Economic Review 105(1):100-130
- Sharma C, Overgoor J (2017) Scaling knowledge at airbnb âĂŞ airbnb engineering & data science âĂŞ medium. https://tinyurl.com/know6v7, (Accessed on 03/22/2017)
- Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science 22(11):1359–1366
- Smith V, Prescott E, Fama E, 497+ (2014) Economist letter against raising the minimum wage. http://tinyurl.com/kx3xphr, (Accessed on 10/31/2016)
- Snilstveit B, Vojtkova M, Bhavsar A, Gaarder M (2013) Evidence gap maps—a tool for promoting evidence-informed policy and prioritizing future research. World Bank Policy Research Working Paper (6725)
- Stodden V (2014) What computational scientists need to know about intellectual property law: A primer. Implementing reproducible research CRC Press, Boca Raton, FL pp 325–339
- Stodden V, McNutt M, Bailey DH, Deelman E, Gil Y, Hanson B, Heroux MA, Ioannidis JP, Taufer M (2016) Enhancing reproducibility for computational methods. Science 354(6317):1240–1241
- Sunstein CR (2014) Simpler: The future of government. Simon and Schuster
- Tetlock P, Gardner D (2015) Superforecasting: The art and science of prediction. Random House
- US Congressional Budget Office (2010) Letter to the honorable nancy pelosi, speaker, u.s. house of representatives. h.r. 4872. https://www.cbo.gov/sites/default/files/hr4872_0.pdf, (Accessed on 05/08/2017)
- US Congressional Budget Office (2014) The effects of a minimum-wage increase on employment and family income
- Vivalt E (2015) How much can we generalize from impact evaluations. Unpublished manuscript
- Weimer DL, Vining AR (2015) Policy analysis: Concepts and practice. Routledge
- Wesselink A, Buchanan KS, Georgiadou Y, Turnhout E (2013) Technical knowledge, discursive spaces and politics at the science–policy interface. Environmental Science & Policy 30:1–9

White House (2013) Executive order—making open and machine readable the new default for government information. White paper

White House (2014) Congressional budget office report finds minimum wage lifts wages for 16.5 million workers | whitehouse.gov. http://tinyurl.com/kkd8fcu, (Accessed on 10/31/2016)

Williams W (1971) Social Policy Research and Analysis: The Experience in the Federal Social Agencies. American Elsevier Publishing Company

Xie Y (2015) Dynamic Documents with R and knitr, vol 29. CRC Press