Essays on Political Economy and Macroeconomics

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Political Economy and Government

have examined a dissertation entitled

"Essays on Political Economy and Macroeconomics"

presented by Brian Wheaton

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________________

Typed name: Professor Robert Barro, Chair

Signature ______________________________

Typed name: Professor Edward Glaeser

Signature ______________________________

Typed name: Professor Gautam Rao

Date: April 12, 2021

Essays on Political Economy and Macroeconomics

A dissertation presented

by

Brian Wheaton

to

The Committee on Higher Degrees in Political Economy and Government

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Political Economy and Government

Harvard University

Cambridge, Massachusetts

April 2021

Dissertation Advisors:
Professor Robert Barro
Professor Edward Glaeser

Author:
Brian Wheaton

Essays on Political Economy and Macroeconomics

Abstract

This dissertation consists of four essays on a range of topics in political economy and macroeconomics which are united by having current policy relevance.

The first essay studies the effects of social policy laws on beliefs and attitudes held by the public. Do laws move public attitudes in the direction of the law, or do they induce systematic backlash, whereby the attitudes of the public move in the *opposite* direction of the law? I set up a model showing that, in the context of identity utility, systematic backlash is the likely outcome. Using a difference-in-differences identification strategy, I examine every major U.S. social policy law from the 1960s onward, and I find statistically-significant and robust evidence of backlash in each and every case.

The second essay (co-authored with Matthew Lilley) studies whether politicians can actually be rewarded for good performance, as suggested by retrospective voting models, or whether erroneous beliefs can hinder the actualization of such models. Looking through the lens of the coronavirus pandemic, we find evidence that *beliefs* about state death rates – which are often erroneous – are actually more important for politician approval than the *true* death rates.

The third essay studies the effects of the flat tax reforms adopted by most Eastern European countries between the mid-1990s and early 2010s on macroeconomic outcomes including GDP

growth, investment, and inequality. Setting up a simple model of intertemporal investment, I show that tax progressivity should negatively impact investment (even holding constant the average level of taxes). Turning to the data, I find statistically-significant and robust evidence of increased investment and, consequently, GDP growth resulting from the flat tax reforms.

The fourth essay (co-authored with Robbie Minton) studies the influence of minimum wages on monetary policy efficacy. In a model, we show that monetary policy shocks should relax the *real* minimum wage and thereby lead to an expansion in (minimum-wage) employment. Consequently, monetary policy should be more effective where the share of minimum-wage workers is higher. We provide extensive empirical evidence in support of this relationship and the underlying mechanism.

Table of Contents

Acknowledgments

I am tremendously grateful to my doctoral advisors: Alberto Alesina, Robert Barro, Ed Glaeser, and Gautam Rao. They provided enormously helpful guidance in virtually every respect of the word over the course of the past six years. They made me a better researcher while supporting me in my own research decisions. I am also grateful to many other faculty members at Harvard from whom I have sought advice over the years: Isaiah Andrews, Gabriel Chodorow-Reich, Ben Enke, Emmanuel Farhi, Martin Feldstein, Claudia Goldin, Oliver Hart, Larry Katz, Greg Mankiw, Rachel McCleary, Nathan Nunn, Vincent Pons, Ludwig Straub, Marco Tabellini, and Elie Tamer. And I am grateful to my undergraduate advisors at University of California, Berkeley – Stefano DellaVigna and Gerard Roland – without whom I would quite possibly not even be at Harvard in the first place.

I am grateful to my co-authors, Matthew Lilley and Robbie Minton, who have been an absolute joy to work with at all times and who have also taught me a great deal about being a good researcher and a good collaborator myself. I am additionally grateful to a broad array of other fellow PhD students with whom I interacted extensively during the PhD program: Augustin Bergeron, Justin Bloesch, Kevin Connolly, Krishna Dasaratha, Enrico Di Gregorio, Josh Hurwitz, Casey Kearney, Andrew Lilley, Michael-David Mangini, Casey Petroff, Aakaash Rao, Robert Siliciano, and many more.

I am grateful to the Association for Comparative Economic Studies, the Lynde and Harry Bradley Foundation, the Institute for Humane Studies, and the Warburg Fund for generous support which made this research possible.

Chapter 1

Laws, Beliefs, and Backlash

1 Introduction

The literature on law and economics has increasingly distinguished between the *functional* role of laws and the *expressive* role of laws. That is, most laws serve dual purposes: they provide civil or criminal penalties which incentivize compliance (functional), but they also provide a signal of society's goals, norms, and standards for acceptable behavior (expressive). Laws vary quite broadly in the extent to which they exhibit each of these two roles. Deeply-buried legal clauses on the precise conditions under which certain tax credits apply may provide relatively little in terms of signaling norms, but they come with a well-defined incentive (i.e., the threat of audit) not to deviate from the law. On the other hand, a gay marriage law – in addition to legally allowing marriage for gay people – may plausibly influence the attitudes and beliefs of heterosexual individuals who are not otherwise functionally bound by the law. Indeed, a statement such as this can be made for many social policy laws.

But if social policy laws do have an effect on attitudes and beliefs, what effect will they have? A straightforward and sensible conjecture would be that, by legislating better conditions or enhanced treatment for a certain group of individuals, public attitudes toward that group would also become more positive. However, it is also possible that legislating better conditions or enhanced treatment for a group could lead to backlash – that is, to attitudes toward the group becoming more *negative*. In a mechanism not dissimilar from a social version of crowd-out, individuals may push back against the law as they seek to preserve their preferred norms.

Furthermore, if these expressive effects of the law do indeed tend to push in the direction of backlash, then in cases where the functional effects of the law are minimal (in terms of bettering the circumstances of the group in question), the backlash may actually *overwhelm* any direct improvements produced by the law. This is a fundamentally empirical question, and distinguishing between the aforementioned hypotheses is the subject of this paper.

To guide this effort, I begin by constructing a model of the effect of social policy laws on actions and beliefs. In this model, each family has preferences over a continuous political spectrum. Broadly speaking, they may be conservative, moderate, or liberal, and this is represented by their bliss point. They prefer to take actions – which may represent the attitudes they express to others, the votes they cast, or a range of other ideologically-coded activities – as close as possible to their bliss point, and they also prefer their children to express ideological preferences similar to their own. Children's preferences are formed by a weighted average of parental actions, the law, and (optionally) the actions of other families in society. I show that these simple assumptions are sufficient to generate systematic backlash against laws.

Intuitively, a law that clashes with a family's ideological preferences places the persistence of that family's preferences into the next generation under threat. Their children will move away from their ideology and toward the law – unless the family pushes back against it. Consequently, families find it optimal to move in the opposite direction of the law in an attempt to preserve the values which are important to them. For example, a conservative family facing a newly-implemented liberal law will find it optimal to express *more* conservatism than they would under a conservative law in order to "save" their child from the influence of liberalism (and vice versa). And a liberal family facing a newly-implemented liberal law is able to reduce their expressions of liberalism and rely, in part, on the law to inculcate their children. A version of the model that additionally allows the actions of other families in broader society to influence children's preferences yields the additional prediction that backlash will persist most strongly and successfully in ideologically-homogeneous communities.

With these theoretical results in mind, I move to the data – focusing first on the state Equal Rights Amendments of the 1970s, which aimed to legislate equality between men and women along various dimensions. The 1970s featured a very public and often-contentious debate as to whether an Equal Rights Amendment (ERA) should be added to the U.S. Constitution and the constitutions of the individual states. These proposals involved adding language to their respective constitutions declaring men and women to be fundamentally equal and subject to equal rights and treatment. The ERA was highly expressive in nature; that is, even its advocates conceded that the legal consequences of the ERA were not known with certainty, and its symbolism was often touted as amongst its most important functions (Mansbridge 1986). The ERA was one of the most salient and visible issues of the 1970s, with GSS data from the late 1970s/early 1980s revealing that 88.4% of individuals had heard of the ERA and 82.2% understood what it was. While the attempt at adding a Federal Equal Rights Amendment to the U.S. Constitution eventually failed, roughly half of U.S. states eventually managed to successfully pass state-level Equal Rights Amendments by ballot initiative.

I leverage the staggered introduction of these state ERAs using a difference-in-differences strategy to identify the effect of a law declaring men and women equal on views about whether men and women are indeed equal – and a variety of other related outcome variables. Using individual-level survey data from the American National Election Study (ANES), I find evidence of a polarization effect, whereby women in states that pass an ERA become marginally more likely to believe in women's equality but men instead react by becoming sharply and significantly less likely to believe in said equality. The two key threats to identification in this setting – migration and policy endogeneity – are unlikely to play a major role given the sign of the effect, as they would entail men who oppose male/female equality moving disproportionately to ERA states and states on a more socially-conservative trajectory being more likely to adopt the ERA, a socially-liberal law. Still, in order to deal with any potential endogeneity, I perform a variety of robustness checks. In particular, I focus on individuals in border counties: comparing the evolution of views on female equality along one side of the border between two states to those along the other side of the border, before and after one of those two states introduces an ERA. I run specifications including state-specific time trends. I conduct permutation tests and a wild bootstrap-t procedure as alternative robust methods of generating standard errors within-sample. I restrict the sample to the closest ERA referenda. And I present evidence from dynamic difference-in-differences specifications that pre-trends are non-existent and the effects do not fade out over time.

In addition to the primary result of backlash, I also find considerable evidence in support of other testable implications of the model. Backlash is significantly stronger amongst men with children, and backlash is successfully passed on to the next generation, albeit with reduced intensity. Backlash occurs on both sides of the political spectrum. Persistence of backlash into the next generation is stronger in ideologically-homogeneous communities. And laws are found to play a unique role in generating backlash; more bottom-up components of the women's movement – such as female entry into the labor force, which I study using a shift-share design, and female election to political office, which I study using a close-election RD design – do not generate backlash.

Next, I provide evidence against alternative mechanisms. First, I provide evidence – using data on second-order beliefs – that the backlash does not merely represent a re-definition of what gender equality is understood to mean by survey respondents. Second, I show that the backlash is not a consequence of the campaign leading up to the law but rather a consequence of the law itself. Third, I find no evidence that persuasion effects – with ERA opponents ramping up their efforts to convince people – are responsible for the backlash, nor do I find any evidence that the media more broadly contributed to the backlash; if anything, it appears to have mitigated it. Fourth, I discuss why an explanation hinging on policy mood – whereby liberal laws may simply tend to be passed shortly prior to conservative shifts in public-opinion – is inconsistent with the results. Fifth, I find no evidence that the backlash is the result of fears on the part of men about increased labor-market competition from women. Sixth, I find evidence against the hypothesis that the backlash merely represents (potentially-irrational) anger at government on the part of those who disagreed with the ERA. Seventh and last, I provide evidence as to why a desire to merely influence the law – without any role for transmitting one's ideological preferences to one's children – is unlikely to be responsible for the backlash.

Finally, I show that backlash is not merely an idiosyncratic consequence of the Equal Rights Amendments. Using survey data from the ANES, the GSS, and Gallup, I present evidence from dynamic difference-in-differences regressions that virtually every major social policy law of the past half-century has induced sharp and significant backlash with no pre-trends. The Civil Rights Acts of the 1960s, the legalization of abortion in the 1970s, the relaxation of gun control beginning in the 1980s, the Defense-of-Marriage Acts of the 1990s, the legalization of marijuana beginning in the 2000s, the legalization of gay marriage in the 2010s, and more – across various categories of social policy and across the ideological spectrum, backlash has time and time again been the consequence. These findings suggest that an important trade-off exists between the direct, functional consequences of a law and the backlash it induces amongst the public. More succinctly, aggressive pushes for social change through legislation may face a significant cost in terms of countervailing cultural backlash.

2 Literature Review

My work builds on and contributes to a number of related literatures within political economy and public economics. There has been a growing effort in recent years to understand the interplay between institutions and culture. A large body of work that dates back to the foundation of cultural economics studies the effects of culture on insitutions. Alesina and Giuliano (2015) extensively summarize this literature in a survey paper. The converse relationship – the effects of institutions on culture – received less attention at first but has been the subject of a growing literature in recent years.

The theoretical literature on the expressive role of the law and its effect on cultural norms and attitudes began in legal journals, seeded by the seminal work of Sunstein (1996). Kahan (1997), Cooter (1998), and Posner (1998, 2000) followed shortly thereafter. Within economics, much of the theoretical literature on the effects of law/institutions on culture relates heavily to the broader literature on cultural transmission. Bisin and Verdier (2001) model the dynamics of cultural transmission, finding that families which perceive their cultural traits to be in the minority double-down on said traits in order to inculcate their children with them and ensure the traits persist. Tabellini (2008) models how enforcement of laws and the broader legal framework contribute to the choice of which values parents attempt to instill in their offspring and consequently the level of cooperation in society, finding the existence of a rich two-way interplay between values and institutions. Greif and Tadelis (2010) model the evolution and persistence of "crypto-morality" – situations prevalent in history wherein families adhere secretly to one morality while openly practicing another in an attempt to thwart institutional pressure for change.

The theoretical literature on the effects of institutions on culture is not limited solely to studies of cultural persistence, however. Benabou and Tirole (2011) model the interplay between laws and norms, arguing that laws both impose material incentives and signal a society's values/norms – and that optimal incentive-setting can differ in the presence of social norms, with laws crowding-out and undermining social norms in certain cases. Acemoglu and Jackson (2017) also model the interplay between social norms and the enforcement of laws, finding amongst other things that more restrictive laws can reduce the incidence of law-breaking behavior amongst individuals who are primarily law-abiding while increasing the incidence of law-breaking amongst individuals who are primarily law-breaking. Departing slightly from the relationship between legal institutions and social preferences, Bowles and Polania-Reyes (2012) survey the related (broader) literature on the relationship between economic incentives and social preferences, finding that crowding-out of social preferences by economic incentives appears to be more common than crowding-in.

My model builds on – and owes much to – the aforementioned approaches. It also owes homage to the very broad public choice literature generally and the median voter theorem specifically in its setup of a spectrum of ideologically-coded choices faced by each agent. This literature is far too broad to review in great detail but was seeded by Black (1948) and Downs (1957). The work of Acemoglu and Robinson (2008) on the substitutability of de facto and de jure power – with reductions in de jure power of a group being ameliorated by increased investments in de facto power – is also highly relevant.

There also exists an empirical literature on the effects of institutions on culture, beliefs, and norms, chiefly focused on the very long run. An early example is Shiller et al. (1992), which focuses on the former communist-led states of Eastern Europe. Using cross-country survey data, Shiller et al. find little evidence of a so-called Homo Sovieticus unmotivated to work and innovate. Also using cross-country survey data in the post-communist context, Roland (2012) observes that, in most dimensions, attitudes about the role of government and the role of markets in transition economies is not converging with those in Western market economies, potential evidence that these preferences come from much longer-run historical factors than the communist experience. Alesina and Fuchs-Schündeln (2007) take their analysis beyond cross-country correlations and look within Germany, focusing in particular on the treatment effect of the East German communist-led system on East Germans. They find that East Germans remain more interventionist and pro-government than West Germans but that the former appear to be converging to the West German norm.

Becker et al. (2016) exploits a regression discontinuity to examine the effects of institutions on beliefs, looking on either side of what was once the Habsburg (Austrian) Empire border. The Habsburg Empire was marked by a characteristically well-functioning bureaucracy, and Becker et al. explore whether this institutional characteristic induced a persistent increase in trust toward government, of which they find some evidence. With a narrower bandwith of 25 kilometers, Peisakhin (2010) surveyed 1675 people living in villages on either side of the former Habsburg-Russian border, finding large and statistically-significant differences in terms of various cultural outcome variables between the two groups. Lowes et al. (2017) study the persistent effects of the institutions of the highly centralized Kuba Kingdom of Central Africa on modern rule-following, finding evidence that the legacy of the Kuba Kingdom is actually that of reduced rule-following and increased cheating – potentially indicative of substitutability between formal institutions and informal culture/social norms.

A subset of this literature uses lab or field experiments to induce variation. Tyran and Feld (2006), Sutter, Haigner, and Kocher (2010), and Dal Bó, Foster, and Putterman (2010) explore the effect of democratic rules on behavior, the latter finding that cooperation is greater under the same rule when that rule is chosen democratically versus when it is assigned exogenously by a computer. Bursztyn, Egorov, and Fiorin (2017) run an experiment on Amazon mTurk finding that exogenous increases in participants' perceptions of Donald Trump's popularity make individuals more likely to exhibit anti-immigration views and behavior.

Fewer papers examine specific laws or examine a short/medium-run setting wherein the dynamics of change in attitudes, beliefs, or norms can be studied at a higher frequency. Gruber and Hungerman (2008), studying the repeal of the Blue Laws in the United States, is an early exception. Recent examples are Fouka (2020), who studies the German-American forced assimilation laws passed in two U.S. states in the early 1900s, and Abdelgadir and Fouka (2020), who study the 2004 French hijab ban – both of which are found to induce backlash. This backlash, however, is of a somewhat different form than the kind I study. It concerns how groups *targeted* by a social policy law respond to that law, whereas I look beyond this realm and study how the non-targeted majority group responds as well. Ang (2019), who studies the specific case of the 1975 revision to the Voting Rights Act and finds evidence of backlash amongst the white majority, is perhaps the study which relates most closely to mine.

By studying individual laws in a short, medium, and long-run setting where the dynamics and pre-periods of legal change are clearly observable, I am able to tightly relate my empirical results to the theoretical research on the effects of laws on attitudes and norms generally – and to my model in particular. In so doing, I hope to tie together the theoretical and empirical literatures on the effects of institutions on culture, attitudes, and norms. And by extending my empirical analysis to cover the major U.S. social policy laws of the past half-century, I hope to make a substantial contribution to the literature on backlash and reveal that backlash is, in fact, a remarkably general phenomenon occurring across the spectrum of laws.

3 Model

3.1 Baseline Model

Consider a setting where, in each generation t, society is made up of a set of N families. Each family has some most-preferred point, $b_{i,t}$ (i.e., a bliss point), along the real line $(-\infty, \infty)$, which corresponds to the left/right political spectrum on a given issue. In other words, some families may be left-wing, some may be centrists, and others may be right-wing. And amongst left- and right-wing families, some may be more extreme than others. Each family i in generation t takes an action, $x_{i,t}$, along the left/right spectrum. Families prefer to take actions as consistent as possible with their ideological bliss point. Actions may represent virtually anything ideologically-coded. For example, a family which favors traditional gender roles will want to make statements in favor of traditional gender roles, vote for the party that is more likely to ensure traditionalism in gender roles, push for a personal relationship and division-of-labor between spouses that reflects traditional gender roles, etc.

Furthermore, families have preferences not only over their actions but also over the ideological preferences, $b_{i,t+1}$, with which they inculcate their children. This reflects the fact that parents typically care about inculcating their children with ideological preferences similar to their own and that parents typically want their children to behave in ways consistent with the parent's views. A left-wing parent, for example, may recoil at the idea of their child becoming a conservative while a right-wing parent, conversely, may recoil at the idea of his child calling himself a socialist. These preferences can be implemented with the following utility function, $u_{i,t}$,

$$u_{i,t}(x_{i,t}) = -(x_{i,t} - b_{i,t})^2 - \alpha(b_{i,t+1} - b_{i,t})^2,$$

where α denotes the extent to which families care about inculcating their children with preferences close to their own, relative to taking actions close to their own preferences.

While parents have direct control over their own actions, their control over their children's actions is indirect. Children's ideological preferences are formed, in part, by observing the actions taken by their parents. However, parents lack total influence over their children. The law set by society, L, also influences children's preferences. Intuitively, while parents have influence over the preferences their children are inculcated with, they are not the sole role models for their children. Their children also look to the broader world around them, learning about the law (potentially through instruction in school or from the media). In other words, children's preferences are formed according to

$$b_{i,t+1} = \gamma x_{i,t} + (1-\gamma)L,$$

where γ denotes the importance of parental actions in the formation of children's preferences.

Proposition 1: Provided $0 < \gamma < 1$ and $\alpha > 0$, the optimal action of families moves positively with the family's bliss point but inversely with the law (i.e., backlash occurs). That is, $\partial x^*_{i,t} / \partial b_{i,t} > 0,\ \partial x^*_{i,t} / \partial L < 0$.

The proof for Proposition 1 (and the other propositions in this section) is provided in Appendix A.1. As one would expect, a family's optimal action is increasing in its bliss point.

That is, more right-wing families will tend to have more right-wing optimal actions, and more left-wing families will tend to have more left-wing optimal actions. The second comparative static is the more surprising one: backlash against laws. That is, families optimally move in the opposite direction of the law. For example, if the law moves from a right-wing policy to a left-wing policy, families optimally move their actions toward the right. The key reason is that families want their children to behave in a manner consistent with their ideological preferences – and the advent of a law out-of-line with their preferences makes this harder. They must double-down further to counteract the influence of the law.

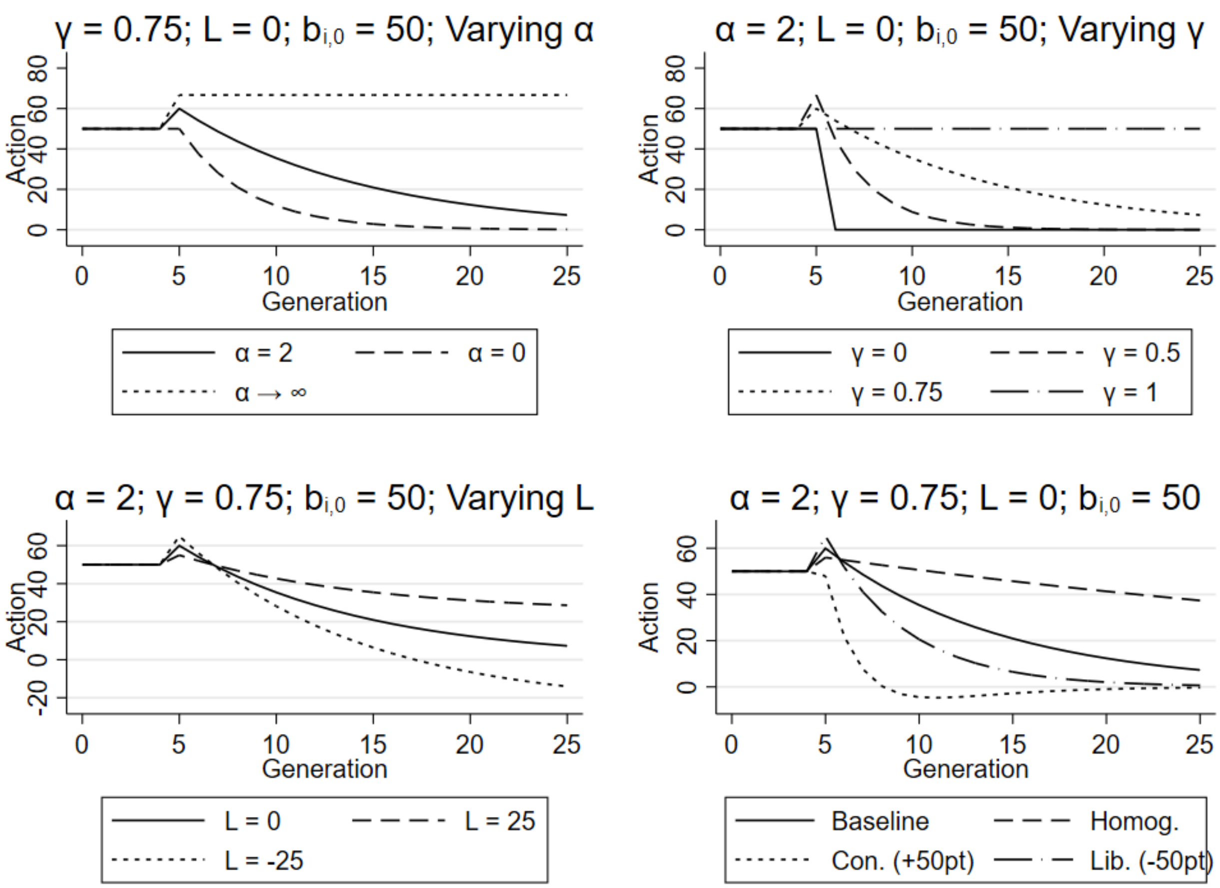
It is worth noting that "backlash" occurs on both sides of the political spectrum. As noted, if the law switches from a right-wing policy to a left-wing policy, the right-wing families double-down to counteract the influence of the law. Meanwhile, the left-wing families *no longer* have to take actions more left-wing than their underlying preferences to counteract the influence of the law, as the law is now in line with said preferences. Thus they can relax somewhat and stop doubling-down; they too can move rightward.

The model also has implications for the dynamic effects of laws – and the persistence of backlash across generations.

Proposition 2: Provided $0 < \gamma < 1$ and α is sufficiently large, backlash will persist beyond the initial generation and be successfully passed down to children. That is, $\partial x^*_{i,t+1} / \partial L < 0$.

To help visualize these concepts, Figure 1 displays a few specific cases. It shows what happens to actions over the course of generations for a family with an initial bliss point of $b_{i,0} = 50$ when the law is initially at $L = 50$ as well but changes in generation 5 to $L = 0$ (i.e., the law moves to

Figure 1: Effects of a Law Change, with Varying Parameter Values



Note: Each panel in Figure 1 considers the effects of a law change from L = 50 to L = 0 in generation five on the subsequent actions of a family i which initially has ideological bliss point $b_{i,0}$ = 50. In the top-left panel, the parameter α – governing the extent to which families care about the ideological preferences of the next generation – is permitted to vary, with backlash resulting as long as $\alpha > 0$. In the top-right panel, γ – which governs the extent to which families have ideological influence over their children – is allowed to vary, with backlash resulting as long as parental influence is existent but incomplete ($0 < \gamma < 1$). In the bottom-left panel, the ideological character of the new law is permitted to vary, with backlash occurring in all cases (though of varying magnitudes). In the bottom-right panel, the extension to the model featuring a role for broader society is considered. An ideologically-homogeneous society generates stronger persistence of the initial ideology than a heterogeneous one wherein half of society is 50 points more liberal or half of the society is 50 points more conservative than the family in question.

the left). The top-left panel varies α but holding other parameters fixed. As can be seen, backlash is the result – the family moves its actions in a more right-wing direction. The strength and persistence of this backlash varies in α, the extent to which families care about the actions of the next generation. In extreme case in which families do not care at all about the actions of the next generation ($\alpha = 0$), backlash is non-existent. In the other extreme case in which families care infinitely more about the actions of the next generation relative to their own actions ($\alpha \rightarrow \infty$), backlash is extreme and completely persistent – actions remain permanently more right-wing as a result of the law moving to the left. In all intermediate cases, there is an initial backlash which is weakened over time as future generations converge to the law.

The top-right panel of Figure 1 instead varies γ while holding other parameters fixed. In the two extreme cases – $\gamma = 0$ and $\gamma = 1$ – there is no backlash whatsoever. This is because in the former case parents exert no influence on their children and consequently gain no utility from backlash. In the latter case, parents have total influence over their children and consequently need not backlash in order to pass their preferences onto them unfettered. For intermediate values, the law and parents both have some influence over their children and, consequently, the incentive for backlash exists.

The bottom-left panel of Figure 1 varies the ideological position of the new law while holding other parameters fixed. Here we see that backlash is stronger the more distant the new law is from the family's initial ideological preferences. Intuitively, a more distant law will require even more force to push back against successfully and prevent children from rapidly moving away from the family's preferences – consequently families find it optimal to push even further in terms of their backlash.

3.2 Extension – Norms and Broader Society

The preceding version of the model was purposely kept minimalistic to illustrate how few factors are necessary to generate systematic, rational backlash. Allowing for the actions of *other* families to influence children arguably increases realism, however.

$$u_{i,t}(x_{i,t}) = -(x_{i,t} - b_{i,t})^2 - \alpha(b_{i,t+1} - b_{i,t})^2$$

$$b_{i,t} = \gamma_P x_{i,t} + \gamma_N \bar{x}_t + \gamma_L L$$

In this alternative setup, γ_P denotes the weight of parental actions in the formation of children's preferences, γ_N denotes the weight of the actions of other families in society (social norms), and γ_L denotes the weight of the law, with these three weights summing to 1. As such, a role for broader society now exists. The utility function itself and other parameters are as before.

Proposition 3: Provided α, γ_L, $\gamma_P > 0$, it is once again the case that the optimal action of families moves positively with the family's bliss point but inversely with the law: $\partial x^*_{i,t} \big/ \partial b_{i,t} > 0$, $\partial x^*_{i,t} \big/ \partial L < 0$.

Proposition 4: Consider two different societies with the same law, L. One is homogeneous, with all families sharing identical ideological preferences, $b_{i,t} = L + b$. The other is heterogeneous, with half of families sharing ideology $b_{i,t} = L + b$ and the other half sharing an opposing ideological preference $b_{j,t} = L - b$. Then, for each family i, $|x^{het,*}_{i,t+k} - L| < |x^{hom,*}_{i,t+k} - L|$ for sufficiently high k. That is, actions will converge more rapidly to the law in the heterogeneous society.

In other words, Proposition 4 says that the homogeneous society will be more successful at preserving its ideology than the heterogeneous society. The backlash will persist longer in an ideologically homogeneous society. The bottom-right panel of Figure 1 varies the ideological makeup of the community while holding other parameters fixed; as can be seen, either a community more liberal on average or one more conservative on average than the family of interest will undermine that family's abilities to preserve its ideological preferences. This highlights a subtle but interesting relationship that has much in common with the broad literature on the consequences of ethnic fractionalization (see, for example, Alesina, Baqir, and Easterly 1999 and Alesina and La Ferrara 2005), which is generally found to reduce social capital and reduce a community's ability to organize public goods provision. Here it is *ideological* fractionalization that contributes to a community's inability to retain its values in the face of institutional pressure. Division within the community means that left-wing and right-wing parents are undermining – rather than reinforcing – each other, meaning that the law has relatively more influence than the old norms in heterogeneous communities and consequently families in these communities have little ability to transmit their preferences onward to future generations.

3.3 Extension – Heterogamy

The baseline model treats the family as the decision-making unit. While it is an accurate statement that cross-ideological marriages in the United States are fairly rare, spouses may also differ in other meaningful ways which have implications for backlash. I consider an extension to the baseline model which allows parents to differ in their ideological preferences, the extent to which ideological matters are important to their identity, and the extent of their influence on their child. This is done with the below parental utility function,

$$u_{i,t}(x_{i,t}) = -\omega_i(x_{i,t} - b_{i,t})^2 - \omega_i\alpha(b_{t+1} - b_{i,t})^2 - p(x_{i,t} - L)^2 \qquad \text{for each parent } i,$$

where ω_i represents the extent to which parent i cares about these ideological matters as part of their identity and p represents any penalty – legal, social, or otherwise – for deviating from the law. As can be seen, $\omega_i = 0$ means that the parent gets no utility from taking actions or inculcating their children with preferences close to their bliss point. They do not care about ideological matters. The child's ideological preferences are formed according to

$$b_{t+1} = \gamma_i x_{i,t} + \gamma_j x_{j,t} + \gamma_L L,$$

where i and j represent the two parents – analogous to the baseline model, except separating the two parents into individual units.

Proposition 5: Provided α, $\gamma_L > 0$ and p is sufficiently small, a parent i will exhibit backlash $\partial x^*_{i,t} / \partial L < 0$ if, and only if, ideological matters are important to their identity (i.e., $\omega_i > 0$) *and* they have ideological influence over their child (i.e., $\gamma_i > 0$).

Thus Proposition 5 states that while backlash remains the result once again, it may occur only on the part of one parent if the other parent does not place much importance on the political issue in question or if the other parent has limited influence over his/her children. It is worth noting that while backlash now requires p being not too large, this assumption is quite likely to be satisfied in the context of social policy laws. For example, for anyone who is not a county clerk, it is impossible to "violate" a gay marriage law in any meaningful sense – and certainly not by expressing anti-gay marriage attitudes or voting a certain way.

In Appendix A, I solve additional extensions to the model which endogenize passage of laws. In Appendix A.2, I endogenize the law by allowing families to vote on the law that will be in place in the next period. Given that systematic backlash results from laws, one might wonder whether any laws would actually be passed in equilibrium in the framework of this model. I show that, as long as families are sufficiently forward-looking, they are willing to pass laws and endure the short-/medium-run backlash in order to shift society toward the law in the long-run. In Appendix A.3, I endogenize the law in a different manner – allowing for backlash in the present period to affect laws in the subsequent period. I show that this provides only a limited additional inducement to backlash.

4 Empirical Framework

Does backlash actually exist in practice? In order to test the implications of the model, I first focus on one social policy law in detail: the Equal Rights Amendment (ERA) of the 1970s, which aimed to guarantee equal rights to American citizens regardless of sex and was added to many state constitutions in that era. I examine this law in depth and study a variety of outcomes – attitudes that people express toward male/female equality, voting patterns, labor-market outcomes, the contours of and roles within marital relationships, etc. Then, to show that the ERA is not unique in generating backlash, I broaden the horizon to virtually every major social policy law of the past half-century for which state-level variation exists, studying the attitudinal outcomes corresponding to those laws. For example, with regard to the legalization of abortion, I study the attitudes people express toward abortion; with regard to gun control, I study the attitudes people express toward gun control; etc.

4.1 Data

I draw on survey data from the American National Election Studies (ANES), the General Social Survey (GSS), and Gallup Poll. Since its inception in 1948, the ANES has asked a random sample of Americans questions about political affiliation and intended voting patterns (virtually) every other year. Since the 1960s, the ANES has asked respondents to provide their "feeling thermometer" toward a wide range of groups (various ethnicities, various political groups, etc.) along with a broad array of other questions on political-economic matters. The ANES is publicly available at the individual level, and the restricted-access version contains state and county codes for each respondent from 1952 to the present.

The GSS asks a similarly-broad swathe of socio-political questions and has been running since 1972 – annually from 1972-1994 and every other year since then. It, too, is publicly-available at the individual level, and the restricted-access version contains state codes since 1973 and county codes since 1994. Many questions in the ANES and the GSS have been repeated without modification for decades, allowing for a consistent view of the evolution of public attitudes and positions. Gallup Poll, too, has asked a battery of socio-political questions since the 1930s. Unlike the ANES and the GSS, Gallup is less focused on academic research and hasn't always asked its questions repeatedly and in consistent time intervals, but some popular questions have been asked frequently and fairly consistently, and some of these pre-date the ANES and the GSS, allowing for analysis of specific law changes not possible with the other two datasets.

With regard to my leading example, the ERA, the ANES has asked a question on equality of the sexes since 1972. Individuals are asked to rate, on a seven-point Likert scale, whether their attitude is closer to "Men and women are fundamentally equal" (1) or "A woman's place is the home" (7). I code a response of 1, 2, or 3 as indicating a positive attitude toward equality and also run regressions on a continuous outcome variable generated by converting this scale into a z-score. While the General Social Survey (GSS) also asks a few questions on views of women's roles, it falls short relative to the ANES in this particular context for two reasons. First, the GSS did not begin collecting county codes until the 1990s, long after all of the identifying variation of the 1970s had come and gone. This makes border-county regression specifications impossible. Second, the GSS did not even record state codes in its very first wave (1972) and only asked the questions about women's roles every other wave during those early years. As such, the first GSS wave for which both (i) the questions of interest are present and (ii) state codes are available was 1974. Because of the substantial number of state ERAs passed between late 1972 and late 1974, several crucial years of data are wiped out, reducing by eight the number of states that can be used for identification. Both of these reasons are the key impetus behind choosing the ANES over the GSS.

I additionally obtain data on voting returns from Dave Leip's Election Atlas, data on fertility patterns from the National Fertility Survey (NFS), and data on employment and occupational outcomes from the Current Population Survey's Annual Social and Economic Supplement (CPS-ASEC). As with the ANES, micro data for the CPS-ASEC is publicly available.

Gladstone (2004) lists the states that adopted ERAs and the years in which they were adopted. This information can be used to create a panel dataset indicating whether or not a given state has an ERA in effect in a given year – and the number of years it has already been effective. Such a panel can then be readily merged with the other data sources, yielding a panel dataset containing the ERA indicator, demographic characteristics, and all the outcome variables of interest.

4.2 Regression Strategy

As noted, the ANES, GSS, and Gallup survey data disclose the state of residence of each respondent. Many laws – including most social policy laws – vary sharply at the state level in the United States of America and have been changed over time in a staggered fashion. This allows analysis of outcomes in states where a given law is passed versus states where the law is not passed. To this end, a static state-level difference-in-differences regression approach can be taken.

$$Y_{ijt} = \alpha + \beta \cdot Law_{jt} + \gamma_j + \eta_t + \varepsilon_{ijt} ,$$

where Y_{ijt} denotes the value of outcome variable Y (say, attitudes about male/female equality) of person i in state j during year t, Law_{jt} is an indicator variable denoting whether the law in question was in effect in state j during year t, γ_j denotes state fixed-effects, and η_t denotes year fixed-effects. As the key right-hand-side variable of interest, Law_{jt} indicates whether an individual is in the treatment (1) or control (0) group. Regressions are weighted with the survey weights included in the corresponding dataset. Following Bertrand, Duflo, and Mullainathan, standard errors are clustered at the state level – the level at which treatment is assigned. Note that this yields nearly 50 clusters. While Cameron, Gelbach, and Miller (2008) and Cameron and Miller (2015) have raised concerns about finite-sample, few-cluster inference, they also show that by 50 clusters, these concerns have largely dissipated.

The identification assumption for a standard state-level difference-in-differences specification such as this one is that of parallel trends: the outcome variable of interest would have evolved analogously in the treatment and control if, counterfactually, the treatment group had not received treatment. For example, in the case of the state Equal Rights Amendments, this assumption is that attitudes expressed toward male/female equality in ERA states would have evolved similarly to non-ERA states if the ERAs had not been passed.

There are two key issues with this assumption: migration and policy endogeneity. The migration issue is that, since the ANES data is not longitudinal at the individual level, it could plausibly be the case that individuals are sorting into the states that have the policy they like. As will be seen, this ends up being a non-issue due to the sign of the effects I find. That is, since a negative effect (backlash) is found, any such sorting would only serve to bias the effect toward zero, making the effect I measure in this static specification an *underestimate* of the true backlash. The policy endogeneity issue is that passage of state laws is not randomly-assigned; hence the states which chose to adopt a given law were plausibly on a different political path than those which chose not to adopt the law. Again, the sign of the effect revealed by the regressions will render this a questionable concern as well, unless one believes that states on a more conservative trajectory are more likely to adopt *liberal* laws (and vice versa).

Still, as one way of dealing with the concern of policy endogeneity, I restrict the sample to counties on either side of a border between an (eventual) law-implementing state and a non-law-implementing state and re-run an adapted version of the above specification:

$$Y_{ijkt} = \alpha + \beta \cdot Law_{jt} + \gamma_{jk} + \eta_t + \varepsilon_{ijkt} ,$$

where Y_{ijt} denotes the value of outcome variable Y of person i in state j along border k during year t and γ_{jk} denotes state-by-border fixed effects. So, for example, a different fixed effect is included for the counties along the western side of the Louisiana/Mississippi border versus those along its eastern side, both of which are different from each of the two fixed effects for either side of the Louisiana/Arkansas border. The idea is that, while a state that passes a certain law may plausibly be on a different political trajectory than a state which does not pass that law, communities just along the border of a state are likely to be much more similar – and evolve much more similarly – to the communities right on the other side of that border. And, insofar as they do differ in terms of levels, this will be captured by the highly versatile fixed-effects anyway. In short, the parallel-trends assumption is plausibly more likely to hold in the border-county setting.

Another way I deal with potential concerns of policy endogeneity is by running dynamic difference-in-differences specifications with pre- periods, as follows:

$$Y_{ijt} = \alpha + \sum_{m=A}^{B} \beta_m \cdot Law_{jt}^{m} + \beta_{(B,\infty)} Law_{jt}^{(B,\infty)} + \gamma_j + \eta_t + \varepsilon_{ijt}$$

where Law_{jt}^{m} is an indicator variable denoting whether the law in question was in its m^{th} or $(m + 1)^{\text{th}}$ year in effect in state j during year t. For example, the Connecticut state ERA took effect in 1974. Thus 1975 is its second year in effect, 1976 its third year in effect, etc. The m^{th} and $(m + 1)^{\text{th}}$ years are grouped because some states pass an ERA in an even-numbered year and some states pass one in an odd-numbered year, whereas the ANES (and, recently, the GSS) is collected only every other year.[1] For all dynamic specifications, I set $A < 0$ in order to test for the existence of pre-trends and thereby provide evidence supporting the lack of policy endogeneity, the existence of parallel trends, and the overall cleanliness of the natural experiment. B denotes the point beyond which remaining periods are pooled. For example, if $B = 10$, ERA effects beyond 10 years after ERA passage are all pooled into one coefficient for compactness. This dynamic specification also responds to the concerns raised recently in the applied econometrics literature – such as in Borusyak and Jaravel (2017) – that running static specifications over long time horizons over which treatment effects may plausibly be heterogeneous can bias the static

[1] Consequently, if the *Law* indicators only referred to one specific year m, the treatment group over which the coefficients are estimated would be inconsistent over time. For odd-numbered m, the treatment group would be composed solely of states which passed the ERA in an odd-numbered year; for even-numbered m, the treatment group would be composed solely of states which passed the ERA in an even-numbered year.

regression coefficient. Borusyak and Jaravel also argue that pooling multiple periods into one coefficient may induce bias, so I additionally run a dynamic specification without such pooling.

In order to test the implications of the model and further investigate the mechanism, I run a multitude of specifications wherein I study the heterogeneity of the law's effects across various categories of individuals or communities. These specifications take the form of the above regressions, but with an interaction term between the right-hand-side law variable and the heterogeneity variable of interest. For instance, in the case case of the static specification,

$$Y_{ijt} = \alpha + \beta_1 Law_{jt} + \beta_2 Heterogeneity_{ijt} + \beta_3 \cdot (Law_{jt} * Heterogeneity_{ijt}) + \gamma_j + \eta_t + \varepsilon_{ijt} ,$$

where $Heterogeneity_{ijt}$ is the heterogeneity variable of interest and, consequently, β_3 is the coefficient revealing heterogeneity (or lack thereof) of the law on the heterogeneity variable. For example, if the heterogeneity variable is income, β_3 provides evidence on the extent to which the law in question has a differential effect on high-income versus low-income individuals.

As noted above, while the number of clusters is near 50 for most of the state-level specifications, certain specifications – in particular, the border-county specifications – result in closer to 25 clusters. While simulations performed by Cameron and Miller (2015) suggest that this too is basically high enough to avoid the statistical concerns associated with having too few clusters, to be safe, I alternatively compute p-values using the Wild Bootstrap-t procedure with 2000 repetitions that they propose in order to ensure that the results are robust. For an even further and more transparent robustness check, I compute p-values in-sample by running straightforward permutation tests (i) randomizing both the treatment states and each state's treatment year and, more strictly, (ii) fixing the treatment states but randomizing each state's treatment year for further assurance of robustness.

5 The State Equal Rights Amendments

5.1 Political Economic Context

The idea of an Equal Rights Amendment to the U.S. Constitution was a hotly-debated issue for over six decades, from the 1920s through the 1980s. The amendment sought to end all legal distinctions between men and women in terms of divorce, property, employment, and all other matters. A proposed Equal Rights Amendment was introduced in every session of Congress from 1921 to 1972, failing to secure passage every single time until the last. By the 1970s, individual laws increasingly existed codifying equal treatment in various dimensions, but advocates of the ERA pointed out that they could be overturned by subsequent laws or Supreme Court decisions, whereas an Amendment would have more permanence and be immune to changing composition of the Supreme Court. Perhaps most importantly, the symbolism of the ERA – declaring to society that not only were all men created equal, all women were as well – was viewed as paramount in itself (Mansbridge 1986).[2]

The debate over the Equal Rights Amendment was very public and very salient; it was one of the most major policy debates of the 1970s. Books and documentaries about the 1970s almost invariably include a chapter or episode on the ERA (e.g., Perlstein 2014, Lepore 2018, CNN 2017). Candidates for office were routinely asked for their views on the ERA with greater frequency than almost any other issue of the day. In terms of concrete data, in two waves of the General Social Survey in the late 1970s and early 1980s respondents were asked whether they had heard of the ERA; 88.4% of respondents answered affirmatively. A follow-up question explored whether individuals understood what the ERA meant; an impressive 82.2% did.

[2] Mansbridge, herself an ERA advocate, wrote "One of the most important indirect effects might have been the effect on the public. … To the degree that having an ERA in the Constitution would remind Americans that equality for women ought to be an important goal in their everyday lives, and to the degree that increased commitment to this value would result in changed behavior on practical issues like who takes care of children, the ERA might have reached beyond the law to the social and economic patterns that produced most of the 59-cents [wage] gap" (pp. 43).

While the question of an ERA was a very contentious one indeed, the coalitions that emerged in support and opposition were not formed along strict and predictable partisan lines. The Republican Party included support for the ERA in its platform beginning in 1940, renewing said support at every Republican National Convention through 1976. The Democratic Party followed along beginning in 1944 at that year's Democratic National Convention, renewing this plank every four years through 1984. There were those in both parties who remained skeptical, however, and only in the early 1970s after a strong push by Michigan Democratic congresswoman Martha Griffiths did an Equal Rights Amendment pass both the House of Representatives and the Senate, whereupon it was immediately endorsed and signed by Republican President Richard Nixon in March of 1972. Unfortunately for its supporters, however, due to the constitutional requirement that all amendments be ratified by three-quarters (38) of the 50 state legislatures within 7 years, the Equal Rights Amendment never became law. Despite a three-year extension signed into law by President Jimmy Carter in 1978, the federal ERA fell short by three states.

Opposition to the ERA, rather than splitting cleanly along Democratic/Republican lines, split more along liberal/conservative lines – in an era where there were still large numbers of liberal Republicans and conservative Democrats. Furthermore, it created faultlines between upper-middle-class elites and the working-class populace. Opposition was led by Phyllis Schlafly, who established the STOP ERA coalition after the passage of a state ERA in her home state of Illinois. Schlafly argued passionately that the ERA would directly ameliorate the special protections and privileges women were given in modern American society – and indirectly by undermining the family unit (Schlafly 1972). The ERA, she claimed, threatened to make the American woman a partner expected to support herself financially, due nothing from her husband, even in case of divorce – and it would also be another set of words for the Supreme Court to work with in an era of repeated liberal Supreme Court decisions. Gay marriage, gender-neutral bathrooms, government support for abortion, military drafting of women, and much more would be likely consequences of the ERA, according to Schlafly. Her ideas gathered much support amongst conservatives, and her advocacy is often regarded as a primary factor in the federal ERA's defeat (Mansbridge 1986). Her successful opposition has even been dramatized in the recent Hulu series *Mrs. America* (2020).

That said, through a distinct yet parallel process, Equal Rights Amendments to the constitutions of 20 states had been ratified by the end of the 1980s – with several more approved and ratified decades later. It is these state ERAs passed in the 1970s and 1980s that I utilize for variation. Table 1 lists the state ERAs and their years of passage; Figure 2 displays the states with an ERA on a map of the U.S.

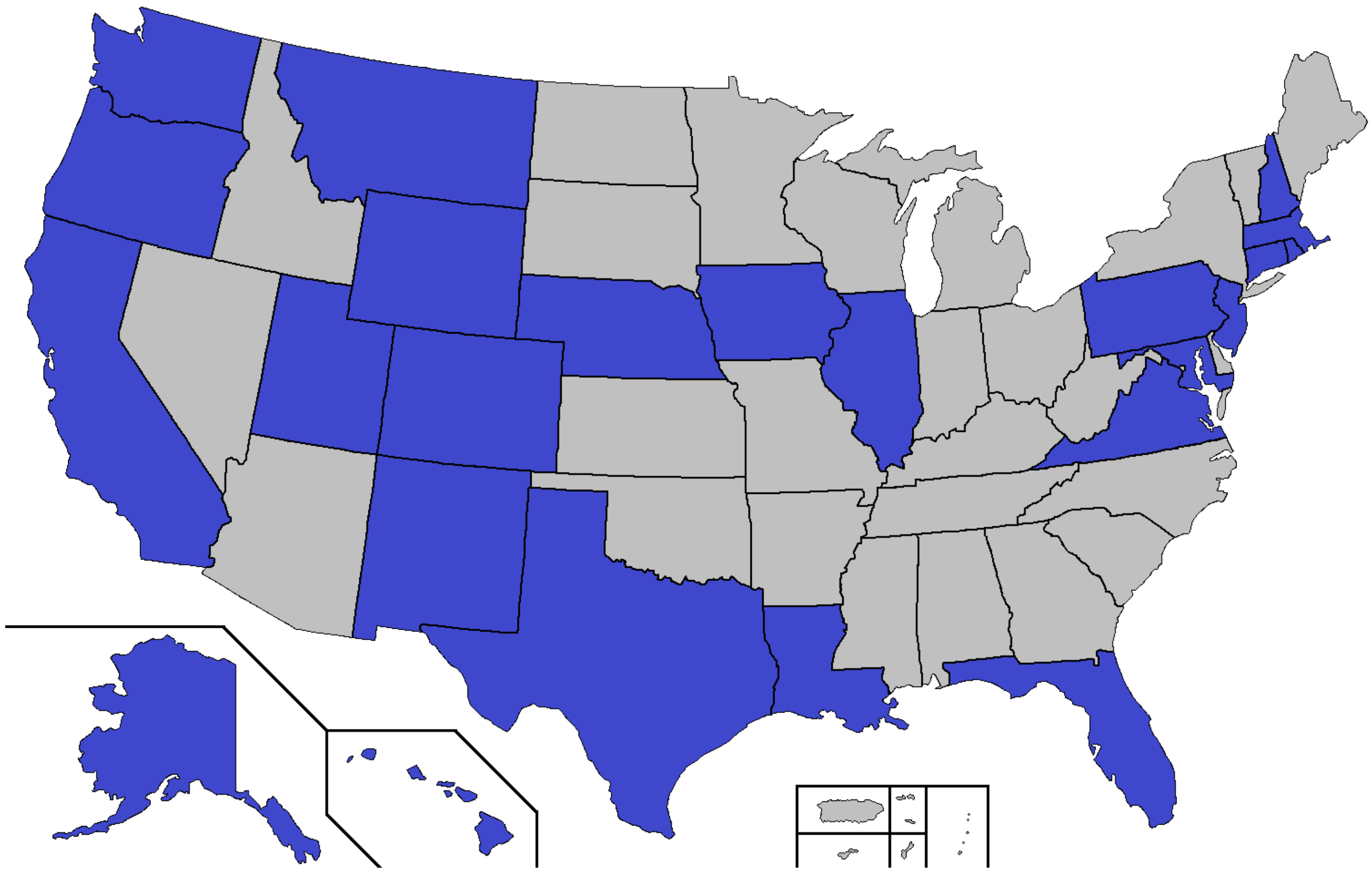
For a number of reasons, the ERA constitutes a desirable natural experiment for studying the effects of laws on attitudes held by the public. One of the reasons is precisely the aforementioned high degree of salience; the ERA was on the mind of the public as an important issue with big implications. Furthermore, because the ERA was initially endorsed by both political parties, the pattern of ERA-adopting states differs from the usual red/blue divide typical of most other laws – and virtually all other social policy laws. There are plenty of states of every political variety and every region within the United States which adopted (and didn't adopt) the ERA. And unlike many laws, the state ERAs were not passed by legislative action but rather by referenda, which allows one to cleanly isolate the effect of the law itself from the campaign leading up to the law. While unanticipated judicially-induced laws (such as the legalization of abortion by *Roe v. Wade*) would avoid entanglement of a campaign effect with a law effect,

Table 1: State ERA Adoption Years

State	Year of Adoption
California	1879
Wyoming	1890
Utah	1896
New Jersey	1947
Illinois	1970
Pennsylvania	1971
Virginia	1971
Alaska	1972
Maryland	1972
Washington	1972
Texas	1972
Colorado	1973
Montana	1973
New Mexico	1973
Connecticut	1974
New Hampshire	1974
Louisiana	1974
Massachusetts	1976
Hawaii	1978
Rhode Island	1986
Florida	1998
Iowa	1998
Nebraska	2008
Oregon	2014
Indiana	2018
Delaware	2019

Note: This table represents the year in which a state Equal Rights Amendment was passed by each of the above states. This information is from Gladstone (2004). My results are identified off of the 16 state ERAs passed in the 1970s and the 1980s, as this is when the big push for the Equal Rights Amendment occurred and when the ERA was a political issue of central importance. Additionally, the main survey outcome of interest is no longer asked by the American National Election Studies in recent years.

Figure 2: State Equal Rights Amendment Map



Note: Blue coloration denotes states with state ERAs (as of 2020).

precisely because these laws came as surprises there was limited public opinion survey data in their pre-period, heavily constraining the statistical techniques and robustness checks one can apply. While I do eventually broaden my focus to study many more social policy laws, these factors render the ERA a natural leading example.

5.2 Main Results

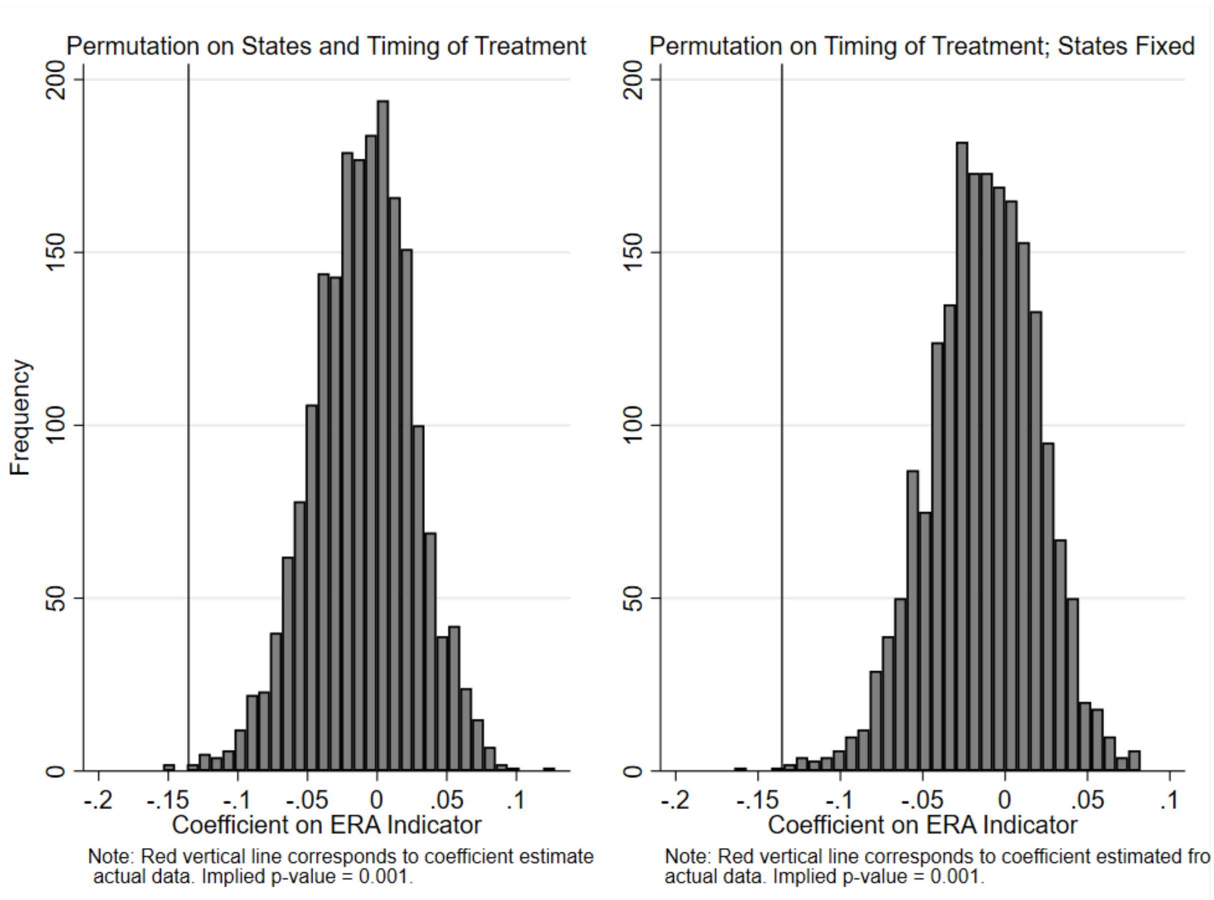
The results of the static specifications discussed in section 4 are displayed in Table 2. The outcome variable in the table is an indicator for whether an individual expresses their attitude as a 1, 2, or 3 on the 7-point male/female equality [1] to inequality [7] scale – i.e., an indicator for positive attitudes toward male/female equality. This results in coefficients that are clean and easy-to-interpret: the percentage-point change in the share of individuals whose position is that men and women are closer to equal than unequal. As column (1) shows, there is an overall backlash effect when both men and women are pooled together in the regression. Columns (2) and (3) make clear the existence of heterogeneous treatment effects: whereas introduction of a state Equal Rights Amendment marginally (but not significantly) increases the proportion of women who believe that men and women are indeed equal, it instead spurs a reaction by men – a *decrease* by nearly 14 percentage points in the share of men who believe in equality of women. Columns (1) through (3) use ANES data from 1972 to 1988 since this corresponds to the first ANES wave in which the aforementioned survey question was asked through the first wave after passage of the final state ERA in my sample. Columns (4) and (5) show that if the end date is instead extended through 1998 (the final year for which ANES geocodes are publicly-available) or 2008 (the final year the aforementioned survey question is asked), the result is nearly identical in magnitude and significance. Columns (6) and (7) turn to the border discontinuity specification.

Table 2: Static Specifications – ERA

Outcome: Indicator for Reporting Positive Attitudes toward Gender Equality	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			State Diff-in-Diff			Border Discontinuity	
Sex:	Both	Male	Female	Male	Male	Male	Female
ERA Indicator	-0.056*	-0.139***	0.032	-0.140***	-0.131***	-0.133***	0.018
	(0.026)	(0.045)	(0.028)	(0.043)	(0.032)	(0.037)	(0.047)
Permutation Test p-values	0.041	0.001	0.334	0.002	<0.001	0.040	0.83
Wild Bootstrap-t p-values	0.048	0.002	0.279	0.002	<0.001	0.021	0.78
Dependent Variable Means	0.541	0.556	0.530	0.613	0.633	0.556	0.539
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-by-Border FEs	No	No	No	No	No	Yes	Yes
Individuals in Sample	All	All	All	All	All	Border Residents	Border Residents
Years of Data	1972-1988	1972-1988	1972-1988	1972-1998	1972-2008	1972-1988	1972-1988
Clustering	State	State	State	State	State	State	State
Observations	15,477	6677	8800	10,448	11,953	2350	3169

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Outcome variable is constructed using the ANES question on attitudes toward male/female equality: "Some people believe that men and women should have an equal role in running business, industry, and government. Others believe a woman's place is the home. Where would you place yourself on this [7-point] scale?" On the scale, 1 indicates total agreement with the former statement; 7 indicates total agreement with the latter. I code a response of 1, 2, or 3 into an indicator variable representing generally positive attitudes toward gender equality. Coefficients in the table can thus be interpreted as changes in the share of individuals expressing positive attitudes toward male/female equality.

Figure 3: Permutation Tests



Note: The permutation test displayed in the left panel randomly selects 16 states to receive a placebo ERA, then re-assigning the year of treatment at random from the list of the 16 actual treatment years of the 1970s/80s-era ERAs. The permutation test displayed in the right panel holds constant the 16 states which receive treatment but re-assigns their treatment years at random.

The backlash effect on the part of males endures with no substantive change in significance.

Figure 3 displays the results of permutation tests run on the main state-level specification for male attitudes (i.e., the specification in column (2) of Table 2). These permutation tests form p-values within-sample rather than relying on standard errors computed from econometric theory to ensure that the results are robust. In particular, the left panel fixes the number of states that adopt ERAs but randomizes *which* states adopt them and randomizes the year in which each state adopts an ERA (by re-assigning the actual treatment years randomly across the placebo states). The right panel fixes specific states which were actually treated with an ERA but randomizes the year in which each state adopted the ERA (again, by re-assigning the actual treatment years randomly across the states). There are minimal differences between the two permutation tests; both yield p-values of 0.001, indicating that the results remain strong. I also run the former test on the other specifications in Table 2, wherein the resulting p-values are reported for each. Also reported are p-values resulting from a Wild Bootstrap-t with 2000 repetitions as another method of generating p-values within-sample, a suggestion of Cameron, Gelbach, and Miller (2008). The results are again robust to this technique.

Figure 4 displays the dynamic difference-in-differences specification with male attitudes toward equality as the outcome. As can be seen, pre-trends do not exist, and the effect is sharp, dramatic, and significant in the near aftermath of ERA passage. Indeed, if one extends the horizon as far as the data permits – 40 years – it can be seen that the backlash effect remains strong and persistent decades later; there is no evidence of fade-out or re-convergence. Figure A-1 presents this longer-horizon dynamic difference-in-differences. Figure A-2 shows the dynamics for female attitudes, which exhibit substantial pre-trends and no significant change on impact. This can be taken as further evidence that, if anything, ERA-adopting states were on a

Figure 4: Dynamic Differences-in-Differences – ERA Effects on Male Attitudes

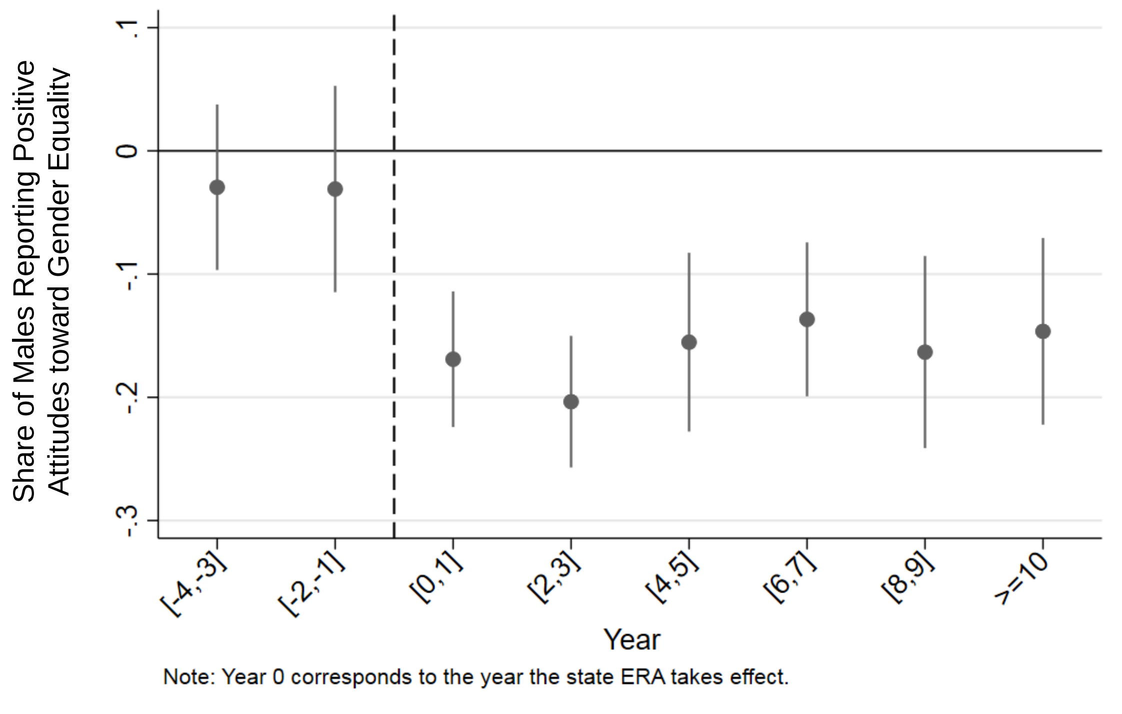


Table 3: ERA Effects on Voting Patterns

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome:	Rep Vote Indicator		Dem Vote Indicator		Rep Minus Dem Vote Share	
Sex:	Both	Both	Both	Both	Both	Both
ERA Indicator	0.050***	0.060***	-0.028	-0.007	0.067**	0.050*
	(0.018)	(0.024)	(0.027)	(0.034)	(0.029)	(0.021)
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes
State-by-Border FEs	No	Yes	No	Yes	No	Yes
Data Source	ANES	ANES	ANES	ANES	Voting Returns	Voting Returns
Individuals in Sample	All	Border Residents	All	Border Residents	All Voters	Border Voters
Clustering	State	State	State	State	State	State
Observations	18,337	3,068	18,337	3068	18,288	1,520

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Regressions in columns (1)-(4) use the individual-level ANES survey data, which asks a survey question about whom the respondent voted for in the most recent presidential election. Coefficients can thus be interpreted as the change in the vote share associated with ERA passage. Regressions in columns (5) and (6) use official voting returns data on county vote shares. All regressions use 1972-1988 data to mirror the main specifications.

more liberal trajectory rather than a more conservative one.

In Table 3, I explore the effect of the state ERAs on voting patterns. Columns (1) through (4) use ANES data and columns (5) and (6) validate these results with official election returns data from Dave Leip's Election Atlas. The result is clear: ERA passage induces a sharp and statistically-significant swing in vote shares toward the Republican Party in the neighborhood of 5-7% -- approximately consistent in both the ANES and official returns data. This is consistent with the anecdotal evidence that the Republican party, as it moved in a more socially-conservative direction in the late 1970s, harnessed the ERA backlash effectively – Phyllis Schlafly, the architect of the STOP ERA coalition, was an important Republican operative and an early supporter of Ronald Reagan in his bid for the presidency. While this is a large swing, it should be noted that the margin of the 1980 Presidential Election was even larger: Ronald Reagan defeated Jimmy Carter by 9.7% of the popular vote. Margins were smaller in certain states than others, so if the aforementioned swing was consistent across states, it would mean that the ERA swung several ERA-adopting states from Carter to Reagan – but fell short of swinging the whole election. Figure 5 shows the dynamics of this effect, revealing no statistically-significant evidence of pre-trends.

Table 4, Panel 1 shows the effect of the state ERAs on a number of placebo outcomes: some of the questions asked most consistently across waves of the ANES. No significant effects are found, apart from one marginally-significant effect that dissipates if one re-runs the regression on border counties. Table 4, Panel 2 shows the effect of the state ERAs on the various "feeling thermometer" questions asked consistently in the ANES. These questions asked individuals how warmly they felt toward various groups on a scale of 0 to 100. Using the full set of such questions that were asked in the early 1970s, I find a significant effect of ERA passage on only

Figure 5: Dynamic Differences-in-Differences –
ERA Effects on Voting Patterns

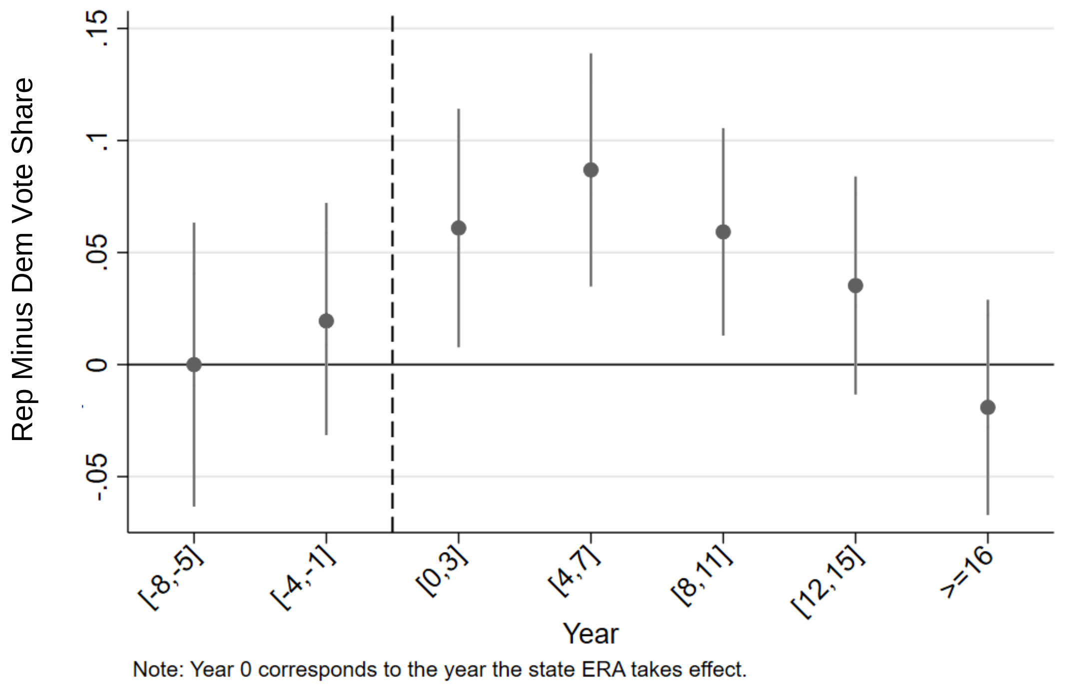


Table 4: Falsification Tests – ERA Effects on Other Outcomes

Panel 1: Important Policy Questions	(1) Approve of Abortion	(2) Support Segregation	(3) Support Busing	(4) Approve of Protesting	(5) Support Protection Minorities	(6) Support Govt Healthcare	(7) Support Govt Job Guar	(8) Govt Wasteful	(9) Trust Govt
Sex:	Male	Male	Male	Male	Male	Male	Male	Male	Male
ERA Indicator	0.041	-0.026	-0.015	-0.085	-0.011	-0.013	-0.123*	-0.010	-0.007
	(0.057)	(0.031)	(0.044)	(0.113)	(0.047)	(0.042)	(0.060)	(0.013)	(0.039)
Observations	3825	3082	3787	2598	2992	3816	7216	6529	6988

Panel 2: Thermometer Questions	(1) Thermom: Blacks	(2) Thermom: Whites	(3) Thermom: Civil Rights	(4) Thermom: Southern-ers	(5) Thermom: Middle Class	(6) Thermom: Police	(7) Thermom: Unions	(8) Thermom: Military	(9) Thermom: Women's Lib
Sex:	Male	Male	Male	Male	Male	Male	Male	Male	Male
ERA Indicator	1.130	0.228	-1.563	0.109	0.987	0.906	-0.348	-0.076	-3.806***
	(0.987)	(1.096)	(1.830)	(0.933)	(1.681)	(1.208)	(1.039)	(1.696)	(1.436)
Observations	13,735	11,741	10,103	5016	8654	10,318	12,133	5509	4575

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. All regressions contain state and year fixed-effects and use ANES survey data from years 1972 to 1988 to mirror the main regressions. Panel 1 studies a selected set of important social and economic questions asked in the ANES. In each column, the outcome is an indicator variable for whether the respondent agrees with the statement in the column title. Panel 2 studies the set of thermometer questions asked as of 1972 in the ANES. Thermometer questions asked individuals how warmly they felt toward various groups on a scale of 0 to 100 (100 being warmest). In each column, the outcome is the corresponding thermometer variable.

one: feelings toward women's liberation activists, which decline markedly. This provides further evidence of backlash.

In Appendix B.1, I further probe these main results. I explore alternative forms of the dependent variable (such as a continuous z-score measure and point-by-point regressions for each of the 7 responses on the 1-to-7 gender equality scale) and conduct robustness checks including the addition of state-specific time trends and the regression approach of Chaisemartin and D'Haultfoeuille (2020). The main result is robust to all of these approaches. In Appendix B.2, I explore a variety of other outcomes, including labor-market outcomes for women, fertility preferences of men and women, and marital happiness. To summarize, I find evidence of worsened labor-market outcomes, more control by men over fertility choices, and worsened happiness for married couples – but not for single men and women. Taken as a whole, these findings may suggest backlashing husbands constraining or otherwise chafing against their wives' choices.

5.3 Testing Other Implications of the Model

Plentiful and fairly robust evidence on the main implication of the model – backlash – was provided in the preceding section. However, the model has other, subtler implications which are also testable. Indeed, if these implications are borne out empirically, the fact that some of them are quite subtle and idiosyncratic to this model should greatly strengthen confidence that the model truly represents the underlying mechanism at work.

First, an obvious implication of the model is that backlash should be stronger amongst those who have children. While the desire to influence society and its future preferences and priorities more broadly than within the confines of one's own family can also motivate some backlash, as shown in Appendix A.3 – the desire to influence one's own children is a powerful channel on its own, and under reasonable parameter values, should account for a large fraction of the total backlash. The ANES, unfortunately, only began asking whether individuals have children of any age later in the 1970s. Earlier – in 1972 – it asked whether individuals had *school-aged* children (specified as 5-18 in the survey questionnaire).[3] Because the ANES began asking the ages of respondents' individual children in 1978, one can construct an indicator for children aged 5-18 from 1978 onward and use this variable to study whether men with children experience a greater backlash to the ERA. This is imperfect, because some individuals who have children (in particular, children aged under 5 or over 18) will be regarded in the regression as not having children. However, this should only bias downward the extent of the heterogeneity I find. Despite the imperfections, column (1) of Table 5 reveals that, indeed, men with children exhibit a significantly stronger backlash.

Second, the model implies that the backlash should be passed on to subsequent generations, as shown in Proposition 2. In order to test this hypothesis, I run a regression specification analogous to the main dynamic specification -- except with *birth cohorts*, rather than years, as the time variable. In other words, I explore whether children born after the ERA have less favorable attitudes toward male/female equality than children born before. Column (2) of Table 5 shows that this is indeed the case for the male children; men appear to successfully pass their backlash onto their sons, albeit at a reduced intensity, which is further reduced as time goes on – precisely as predicted by the model. Figure 6 reinforces this result with a dynamic difference-in-differences specification, showing that a sharp effect endures, with no statistically-significant

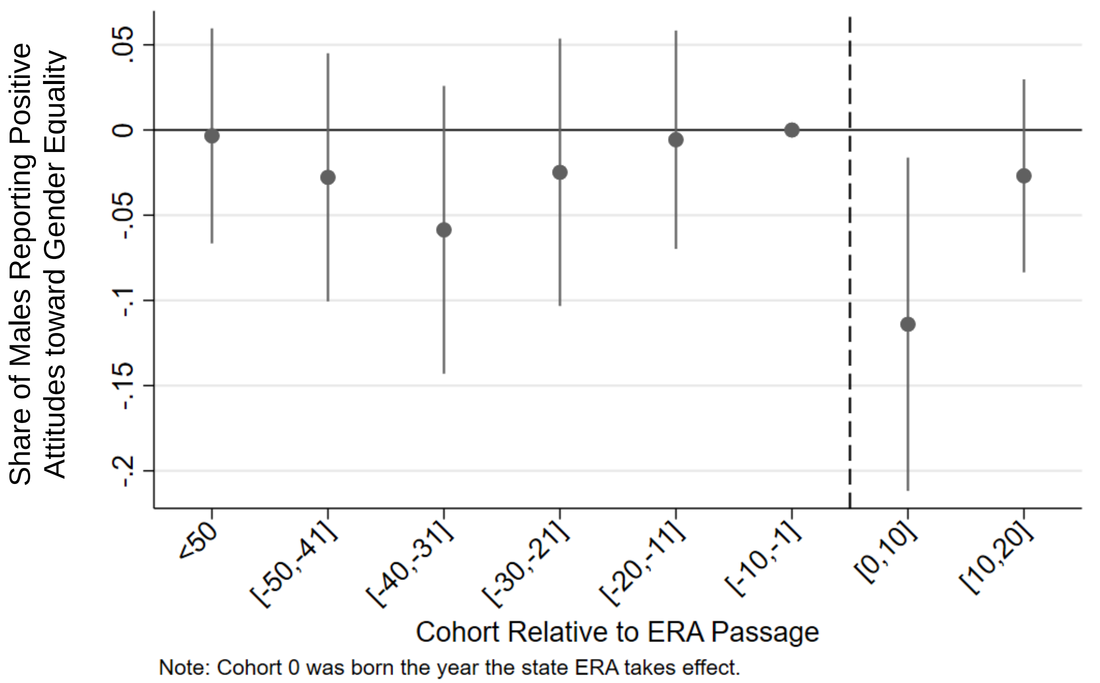
[3] It appears to have asked this question as a flag to determine whether or not the respondent should be asked the immediately following set of questions in the questionnaire, all of which pertain to experiences of parents with school-attending children.

Table 5: Testing Model Implications

	(1)	(2)	(3)	(4)	(5)
Outcome:	Contemp. Attitudes toward Gender Equality	Next Generation Attitudes toward Gender Equality	Contemp. Attitudes toward Gender Equality	Contemp. Attitudes toward Gender Equality	Next Generation Attitudes toward Gender Equality
Characteristic:	Have Children Indicator	N/A	Ideology - Lib/Con Scale	County Republican Vote Share	Ideological Homogeneity Indicator
Sex:	Male	Male	Male	Male	Male
ERA Indicator	-0.078*	-0.092*	-0.154***	-0.149***	-0.062*
	(0.035)	(0.046)	(0.054)	(0.047)	(0.028)
ERA Indicator	-0.043*		0.006	0.072	-0.064**
*Characteristic	(0.019)		(0.030)	(0.212)	(0.026)
Year FEs	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes
Years of Data	1972-1988	1972-2008	1972-1988	1972-1988	1972-2008
Clustering	State	State	State	State	State
Observations	5,249	10,394	4,885	6,513	10,220

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. All columns feature the same outcome variable as in the main regression table: an indicator variable for those expressing positive attitudes toward male/female equality. Coefficients in the table can thus be interpreted as changes in the share of males expressing positive attitudes toward male/female equality. However, columns (1), (3), and (4) result from the standard *year* difference-in-differences specification comparing changes in attitudes in years before versus after ERA adoption. Columns (2) and (5) result from a *cohort* difference-in-differences specification comparing changes in attitudes in birth cohorts born before versus after ERA adoption in order to understand the extent of backlash amongst the subsequent generation. The interacted characteristic in column (3) is a 7-point "Very Conservative" through "Very Liberal" self-reported ideology scale. The interacted characteristic in column (4) is county Republican vote share in 1968. The interacted characteristic in column (5) – the ideological homogeneity indicator – is an indicator for whether the county the individual lives in had a Republican vote share between 40% and 60% in 1968 (this accounts for 50% of counties and is thus an even bifurcation, but changing the threshold does not materially alter the results).

**Figure 6: Dynamic Differences-in-Differences –
ERA Effects on the Next Generation of Men**



pre-trends.[4]

Third, according to the model, backlash should occur on both sides of the ideological spectrum. As seen in Propositions 1 and 3, backlash is not conditional on one's ideological position. As the ANES has asked since the early 1970s whether individuals consider themselves liberals or conservatives (and the intensity of that identification), it is possible to test that implication as well. Columns (3) and (4) of Table 5 reveals that, indeed, both liberals and conservatives exhibit a backlash that does not differ in magnitude. Column (3) uses the ideological self-identification from within the ANES as the interaction variable; column (4) uses 1968 county-level Republican vote share as the interaction variable. The conclusion is the same in both cases.

Fourth, as shown in Proposition 4, persistence of backlash into subsequent generations should be stronger in ideologically homogeneous communities than it heterogeneous ones. This is arguably the most subtle of the implications. However, one can use data on county vote shares in the 1968 Presidential Election – the last one before the advent of the state ERAs – to determine whether individuals live in an ideologically homogeneous or ideologically heterogeneous community.[5] Column (5) interacts the *cohort* static specification with an indicator variable for whether the individual's county of residence had a 1968 Republican vote share between 40% and 60%. This cutoff is chosen because almost exactly 50% of counties fall into that category, allowing for an even bifurcation into "homogeneous" and "heterogeneous" counties. As can be seen in column (5), the persistence of the backlash into the next generation

[4] Note that, because the first of the state ERAs was passed in 1970 and because the question on attitudes toward male/female equality was last asked by the ANES in 2008, no individuals born more than 20 cohorts after ERA passage are available for analysis. This is why the dynamic graph ends at +20.

[5] A measure of the share of liberals and conservatives at the county level would be somewhat more ideal since Democrat:Liberal :: Republican:Conservative was not a perfect correspondence in this era, but such data unfortunately does not exist.

is indeed significantly stronger in ideologically homogeneous communities.[6] In short, this subtlest of implications, too, is borne out in the data.

Fifth, laws should play a unique role in generating backlash, stronger than more bottom-up approaches. In a sense, this is more of an assumption of the model than an implication – it represents the fact that the law, L, is given an special role ($\gamma_L > 0$) in forming children's preferences. While the extension of the model does allow a role for the actions of others in society ($\gamma_N > 0$), every single family in a society rarely moves in concert in the way that a change in legislation does – and thus is unlikely to be capable of inducing strong backlash in the same way as a law. This can be tested by analyzing the other components of the women's movement. While the ERA was one of the movement's primary pillars, it did not stand alone. The entry of women into the labor force, the election of women to political office, and other new laws (such as those pertaining to contraceptive access) were also fundamental to it. In Appendix C, I explore these broader aspects of the women's movement and present evidence that, indeed, laws generated backlash while its more bottom-up aspects did not.

Finally, it is worth discussing the fact that backlash is observed only on the part of males. While this is not a direct implication of the baseline model, it is in fact an implication of the extension of the model which allows parents to differ in their beliefs, their identity, and their influence on their children (Section 3.3). As shown in Proposition 5, in that context, if gender roles are fundamentally important to male identity but of lesser importance to most women (i.e., $\omega_{father} > \omega_{mother} \approx 0$), then backlash to the ERA would indeed be exclusive to men. And if sons primarily look to and are inculcated with their fathers' behavior ($\gamma_{father} > \gamma_{mother} \approx 0$ for male children) while daughters primarily look to and are inculcated with their mothers' behavior ($\gamma_{mother} > \gamma_{father} \approx 0$ for female children), then the backlash would solely be passed on to male

[6] The results are qualitatively the same if the threshold is altered to 33%/67% or 25%/75%.

children. Indeed, there is much evidence from the psychology literature supporting both of these assumptions. The key importance of gender roles to male identity has been studied extensively in the body of literature known as masculinity research, summarized by Levant and Richmond (2007). Meanwhile, classic psychoanalytic theory, dating back to Freud (1909), posits that children increasingly relinquish their attachment to their opposite-sex parent at an early age and begin to identify with their same-sex parent, with boys subsequently emulating their fathers and girls emulating their mothers. More recent research has provided empirical evidence for the importance of the father-son/mother-daughter channel in the transmission of gender role attitudes in particular (Young 1995, Moen et al. 1997). With these well-established results in mind, the findings of the empirics fall directly in line with the model.

6 Alternative Mechanisms

6.1 Redefinition – a Fake Backlash

What if the law caused no change whatsoever in attitudes? What if it merely caused the definition of gender equality to be redefined? For example, recall that the main ANES survey question asks individuals to state their attitudes toward male/female equality along a scale of 1 to 7. Consider an individual who is generally supportive of feminism but indifferent about an ERA. Perhaps prior to the ERA he would have considered himself a "2" – close to total commitment to male/female equality. But the fact that the ERA is now law and he is only indifferent might make it harder for the individual to consider himself near the forefront of male/female equality. So perhaps he now marks himself as a "3" or a "4", which would appear as backlash, despite the fact that his attitudes have gone unchanged.

The first response to this conjecture is quite simply that, if it was the case, material

Table 6: Alternative Mechanisms

	(1) Dem Party's Position	(2) Rep Party's Position	(3) Campaign Effect	(4) Law Effect	(5) ERA and ERA DMA	(6) DMA Within-State	(7) 1973-75 Recession	(8) Marital Status	(9) Gov. Trust Index
Outcome:	Views of Dem Party's Attitude	Views of Rep Party's Attitude	Attitude toward Gender Equality	Attitude toward Gender Equality	Attitude toward Gender Equality	Attitude toward Gender Equality	Attitude toward Gender Equality	Attitude toward Gender Equality	Trust in Govt.
Sex:	Male	Male	Male	Male	Male	Male	Male	Male	Male
ERA Indicator	0.008	-0.044*	-0.030	-0.159***	-0.212***		-0.152***	-0.120**	-0.935
	(0.031)	(0.019)	(0.040)	(0.057)	(0.053)		(0.041)	(0.047)	(1.367)
ERA_DMA					0.067†	0.092**			
Indicator					(0.038)	(0.039)			
ERA Indicator							-0.256	-0.019	
*Characteristic							(0.322)	(0.028)	
Characteristic							ΔUR73-75	Married	
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State	DMA	State	State	State
Observations	7,933	7,933	4,010	2,994	6,570	6,570	6,570	6,570	6,969

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. All regressions use 1972-1988 data for consistency with the main specifications. Columns (1) and (2) test whether the backlash results from the campaign instead of the law itself; in column (1), states where the ERA was on the ballot but never passed are the treatment group and states without the ERA on the ballot are the control group; in column (2), states where the ERA passed are the treatment group and states where the ERA was on the ballot but never passed are the control group. Columns (3) and (4) use data on respondents' views about a political party's attitudes toward gender equality to demonstrate there is little evidence of the main effect representing a mental redefinition of what it means to support gender equality resulting from the ERA. Columns (5) and (6) study whether the backlash is driven by media-market effects, finding no evidence of this using the fact that media markets overlap state borders (such that a media-market primarily serving state A where an ERA is passed may also encompass part of state B, where no ERA is in effect but individuals would nonetheless be exposed to media coverage of the ERA). Columns (7) and (8) explore whether material concerns may be responsible for the backlash, finding no evidence that men in areas harder hit by the 1973-75 recession (who one would expect to be more sensitive to increased labor-market competition) and no evidence that married men (who may experience countervailing positive material effects of the ERA through their wives) exhibit less backlash.

consequences in terms of voting patterns or the relationship patterns between men and women should have gone unchanged – the effects should remain limited to a survey where mental re-indexing of this sort can be done. However, I find evidence of material outcomes in a number of different dimensions.

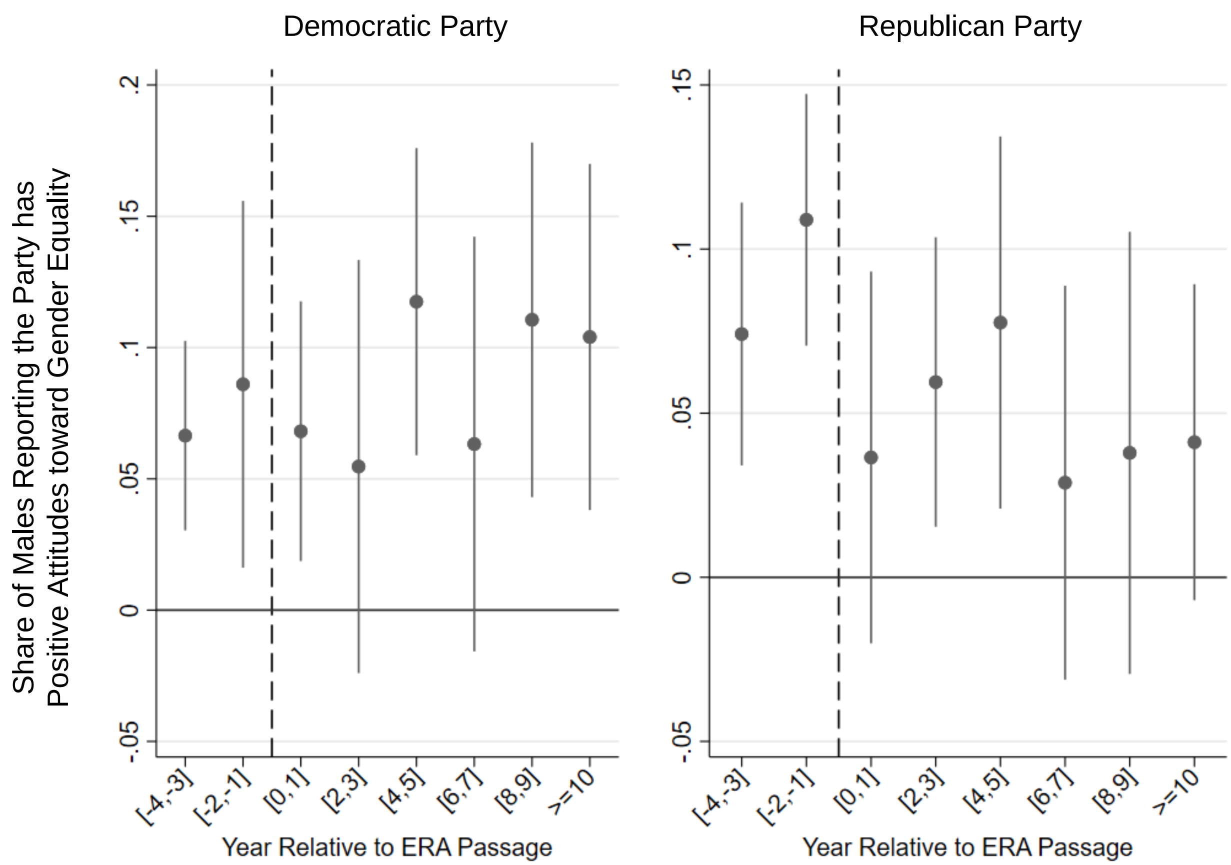
A more direct response relies on the fact that the ANES also asks parallel questions about individual's perceptions of the Democratic Party and Republican Party's positions on the attitude-toward-equality scale. If individuals are mentally modifying the meaning of the index, responses to these two questions should also exhibit a backlash jump after passage of the law. If responses to these questions do not change and the positions of the two parties remain stable while the individual's position changes, this is evidence of a real change in attitudes.

Column (1) of Table 6 reveals that there is no change in individuals' perceptions of Democratic Party attitudes toward male/female equality, but column (2) suggests there may be a change in individuals' perceptions of Republican party attitudes. However, running the corresponding dynamic specifications, represented in Figure 7, reveals the existence of a pre-trend. There is, in fact, no jump in individual's perceptions of either Democratic Party or Republican Party attitudes toward male/female equality resulting from the ERA – just a flat line in the case of the former and a downward trend in the case of the latter (consistent with the Republican party moving in a more socially-conservative direction over the course of the 1970s and 1980s). This suggests that the backlash is not a "fake" one driven by mental re-definition of the survey question.

6.2 Campaign Effects

Was it indeed the law itself which caused the backlash, or was it the *campaign* surrounding

Figure 7: Dynamic Differences-in-Differences –
ERA Effects on Male Perception of Party Attitudes



Note: Year 0 corresponds to the year the state ERA takes effect. The ANES survey questions represented in these graphs is analogous to the main survey question, but instead of asking the respondent's position on the 1-to-7 gender equality scale, they ask where the respondent would place the Democratic party and the Republican party on the very same scale. As in the main specifications, I create an indicator variable representing generally positive attitudes toward gender equality from responses of 1, 2, or 3 on the scale.

the law? That is, could the culprit for the male reaction have actually been seeing confident feminists forcefully voice their views and critiques of society on a regular basis in the months leading up to the state election? This conjecture does not necessarily seem far-fetched. Fortunately, the manner in which the state ERAs were passed allows for a novel way of adjudicating between these two possible mechanisms.

In the case of every single state ERA which was implemented, the ERA was approved by a majority vote through a ballot question in the style of a referendum. The path to such a referendum, however, takes several steps. In order to be approved for the ballot, a proposed ballot initiative must first collect signatures from a fixed (minimum) number of state residents. Typically the number is in the neighborhood of 5 - 10% of the number of votes cast in the most recent gubernatorial election. If the proposal does not receive the requisite number of signatures, it is discarded and does not make it to the ballot. If it does receive sufficient signatures, the proposal will appear on the subsequent state general election ballot, where it will then be subject to a simple Yes vs. No majority vote.

As such, the total effect of a state ERA can be decomposed into the campaign effect and the law effect. To isolate the campaign effect, the treatment group is the group of states where the ERA made it onto the ballot but did not pass. In such states, there would have been broad campaigns in favor of and against the ERA leading up to the general election – but no ERA itself. The control group, then, consists of the states where the ERA didn't make it onto the ballot at all. Meanwhile, to isolate the law effect, the treatment group is the group of states where the ERA passed. In such states there was both a campaign and implementation of a state ERA. The control group is the group of states where the ERA made it onto the ballot but did not pass – which had been the treatment group in the campaign-effect case. With this setup, one holds constant the occurrence of a campaign and identifies purely the effect of the law itself.

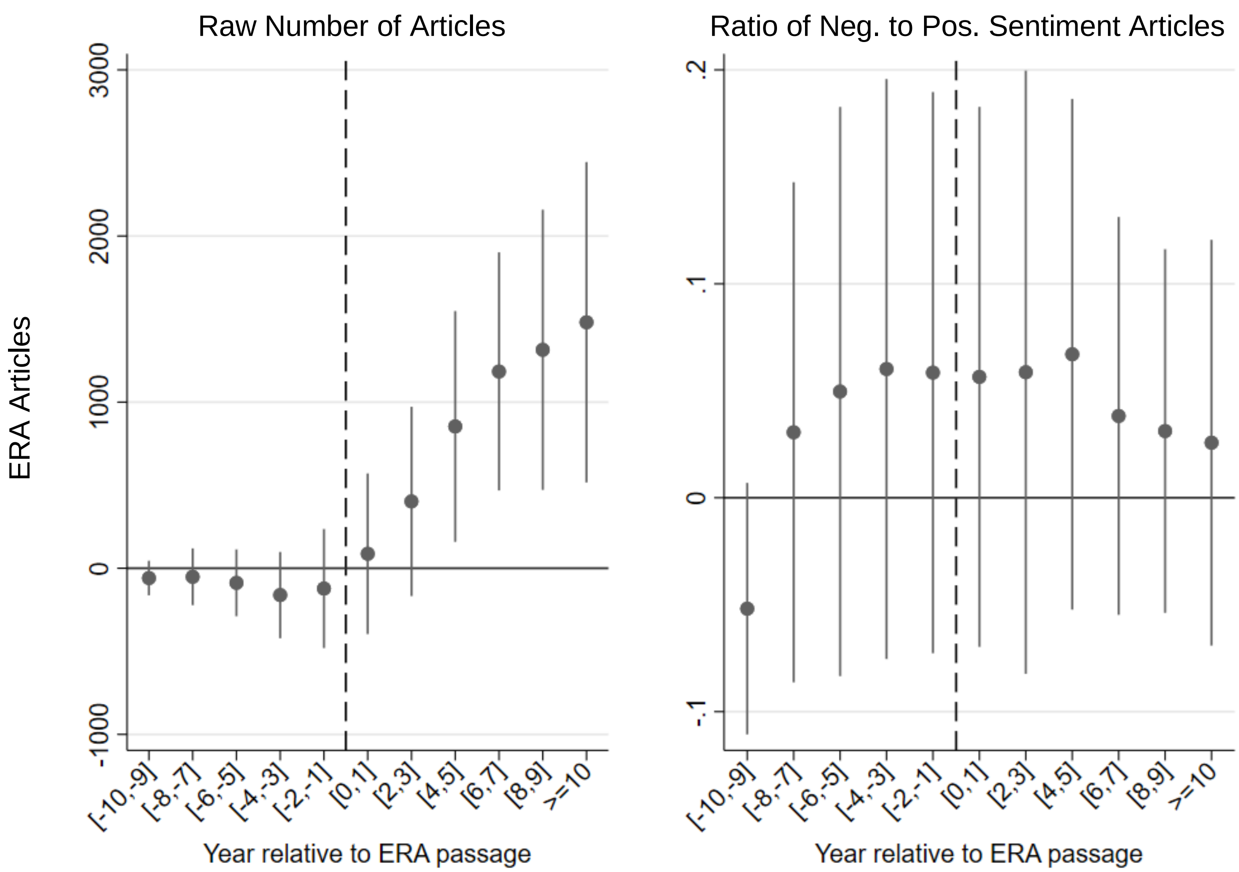
The results of these regressions can be seen in columns (3) and (4) of Table 6. Comparing the two columns, it is apparent that the effect proceeds entirely through the law; the campaign itself has no significant effect whatsoever. While it is possible that there is a difference between successful and unsuccessful campaigns, this difference should be minimal if one compares barely-successful and barely-unsuccessful campaigns. This is done in Table B-4 of Appendix B; as can be seen there, the result does not change qualitatively. Quantitatively, the backlash effect of the law is actually (non-significantly) *larger* when the sample is restricted to these close campaigns.

6.3 Persuasion and Media Effects

Another alternative mechanism is the effect represents ERA opponents ramping up their persuasion efforts in an attempt to convince supporters to turn against the ERA. Since the debate around the Federal ERA was still strongly ongoing after the states had passed their ERAs, ERA opponents would have a very salient reason to continue rallying opposition against the ERA. There is a peculiar facet about this alternative mechanism. If ERA opponents truly possessed such persuasive power, it is a bit odd that they did not make use of it during the campaign and thereby prevent the ERA from being passed in the first place. Still, perhaps it is possible that ERA opponents can speak with a greater, more convincing air of authority once the ERA has been passed and its consequences are beginning to be known to the public.

I present evidence that this does not appear to be the case. Using data from NewspaperARCHIVE, which has amassed a collection of hundreds of millions of local newspaper articles in the United States, I first examine the effects of ERA passage on the number

Figure 8: Dynamic Differences-in-Differences –
ERA Effects on Newspaper Articles about the ERA



Note: Year 0 corresponds to the year the state ERA takes effect. I count as "positive sentiment" any article featuring the words "Equal Rights Amendment" AND "necessary", "good", OR "positive". I count as "negative sentiment" any article featuring the words "Equal Rights Amendment" AND "unnecessary", "bad", OR "negative". Approximately 10% of articles overlap between the two categories. Results remain non-significant if I drop these overlapping articles. Results remain non-significant if I use a broader dictionary of positive and negative synonyms.

of ERA articles appearing in newspapers and then decompose this into the number of negative- and positive-sentiment ERA articles, taking the ratio of the former to the latter.[7] The left panel of Figure 8 demonstates that ERA passage does indeed lead to an increase in the frequency of articles about the ERA. However, as can be seen from the right panel of the same figure, this increase does not occur disproportionately through negative- or positive-sentiment articles. Both increase by approximately equal amounts, and thus the ratio remains roughly constant. Although we cannot know for certain the "convincing power" of a typical negative-sentiment article relative to a typical positive-sentiment article, it is difficult to argue that persuasion is the main channel of the effect given these results, especially when coupled with the finding that persuasion efforts during the campaign didn't do much of anything to attitudes.

Somewhat more generally, another way of measuring effects which pertain to information rather than the law in itself is to observe that Nielsen media markets often overlap state borders. Consequently, people watching TV news in one state often receive information about their neighboring state. For example, the majority of TV viewers in the West Texas media market live in El Paso, Texas. This market, however, also encompasses parts of Southern New Mexico. Consequently, the local news (and advertising) in those Southern New Mexico counties will be heavily geared toward West Texas. So individuals living in Southern New Mexico will hear much about the Texas ERA during the campaign and after it is passed (given the salience of the ERA issue in that era), but they will not themselves be subject to the law or its provisions. One can thus run a regression specification which includes two indicator variables – an indicator for whether the respondent's state is an ERA state (the standard indicator variable), another for

[7] I count as "positive sentiment" any article featuring the words "Equal Rights Amendment" AND "necessary", "good", OR "positive". I count as "negative sentiment" any article featuring the words "Equal Rights Amendment" AND "unnecessary", "bad", OR "negative". Approximately 10% of articles overlap between the two categories. Results remain non-significant if I drop these overlapping articles.

whether the state containing the majority of the respondent's media market is an ERA state. One can also run a within-state regression with state-by-year fixed effects which relies on comparing counties that are in a non-ERA media market to counties in that are in an ERA media market within the same state. Columns (5) and (6) of Table 6 runs both of these specifications, and they reveal that information effects through the media are not responsible for that backlash. Indeed, if anything, this channel results in a more *positive* view of male/female equality.

6.4 Policy Mood

Some political scientists – beginning with Stimson (1991) – have conjectured and provided evidence that aggregate public opinion in the United States has undergone a series of oscillations between liberal and conservative positions. This suggests it may not be too surprising for liberal laws to be followed by a conservative shift (and vice versa) not as a result of the laws themselves but of pre-existing trends. Such trends, however, are unlikely to be driving the backlash I uncover. First of all, Stimson's analysis pertains to aggregate, national-level public opinion, not state-level public opinion. Because the laws I examine are state laws, which are implemented in a staggered fashion, for policy mood to drive my result it would be necessary for differing public-opinion cycles to exist in different states. And if this were true, it would smooth national-level public opinion and make the very cycles Stimson observes non-existent or at least quite muted. In any case, my dynamic specifications include pre-periods, and as was seen, there was no evidence of differential trends prior to treatment in the ERA-adopting states compared to the non-ERA-adopting states. Finally – and perhaps most crucially – Stimson's public-opinion cycles occur across a broad range of ideologically-coded outcomes simultaneously. The public shifts from being more liberal across a broad range of domains to being more conservative across a broad range of domains (or vice versa). My falsification tests showed that implementation of the ERAs led only to a backlash in the dimension of women's rights, not other domains.

6.5 Labor-Market Issues

What if the backlash to the ERA entirely boils down to material economic causes? Men may be concerned that the ERA will give women an edge over men in the labor market with regard to hiring and promotion – or simply that it would entice more women into the workplace, increase competition, and drive down men's wages. This conjecture yields several testable implications. If it is so, then (i) men for whom worries of competition and job precarity are greater should experience a larger backlash; men who are more comfortable or less worried about job/wage loss should be relatively less concerned. Additionally, (ii) married men should experience a relatively weaker backlash (other things equal), as the benefits obtained by their wives should at least partially offset the losses they experience, meaning the net reduction in household income would be lesser for married men. Finally, (iii) there should be backlash to *actual* female labor-market entry. That is, if the backlash to the ERA is a consequence of greater female involvement in the labor force, then greater female involvement in the labor force – measured directly – had better induce backlash itself.

Testing these first two conjectures is straightforward. For (i), it is possible to leverage the fact that the 1973-1975 recession was beginning and intensifying just as most of the ERAs were being passed. One can interact the severity of the recession (peak county unemployment rate) with the ERA indicator to test for heterogeneity. More simply, one can interact the income quantile variable in the ANES with the ERA indicator to study whether poorer men undergo a greater backlash. In neither case is any significant heterogeneity found, as revealed column (7) of Table 6. Column (8) tests (ii), and there, too, no significant heterogeneity is uncovered.

With regard to conjecture (iii), as shown in Table C-1, the entry of women into the labor force – instrumented for using the previously-described shift-share – did not induce any statistically-significant backlash. If the entry of women into the labor force itself did not generate any backlash, it is hard to argue that the channel through which the ERA generated backlash was entry of women into the labor force. Also, as discussed previously, if anything, the ERA appears to be associated with reduced female labor force participation and reduced female presence in higher-tier occupations.

6.6 Anger

One possible conjecture is that the backlash need not be rational or calculated at all. It may simply be that those who opposed the ERA feel anger toward the government for imposing a law with which they disagree. The immediate implication of such a mechanism, however, is that conservatives should undergo backlash against the Equal Rights Amendment, whereas liberals should not. This implication can be tested on the data, and as we have seen, in Table 5 it already was. Liberals and conservatives both undergo backlash – consistent with the paper's main model but not this alternative. Additionally, it should be noted that another implication of this alternative mechanism is that anger/distaste toward the government actually does increase. Column (9) of Table 6 – which makes use of the trust-in-government index present in the ANES since 1960 – does not even find statistically-significant evidence that this occurs.

6.7 Overturning the Law

A closely-related, more rational version of aforementioned mechanism relates to changing the law. What if individuals backlash against the law because doing so influences what the law will be in the next period? In Appendix A.3, I model why such a mechanism is unlikely to be capable of driving strong backlash. Intuitively, whereas an individual has a uniquely privileged role in inculcating his children with his ideological preferences, any given individual will not have much control over the law. The marginal contribution of one individual to a backlash movement aiming to overturn a law is minimal – a drop in the policy ocean, so to speak. This can offer a very slight additional inducement toward backlash, but not a major one.

7 Beyond the ERA – Other Laws

The state Equal Rights Amendments generated significant and persistent backlash, but is this unique to the ERAs, or does it hold true more generally for other laws as well, as predicted by the model? To answer this question, I investigate some of the most major, most salient social policy laws of the past half-century.

7.1 The Civil Rights and Voting Rights Acts

Racial issues have remained at the forefront of U.S. social policy for virtually the entirety of this country's existence. During the Civil Rights Movement, the federal government passed three landmark laws advancing the rights of Black Americans: The Civil Rights Act of 1964, the Voting Rights Act of 1965, and the Civil Rights Act of 1968. The 1964 Act desegregated public accommodations (such as shops, restaurants, and recreational areas), and consequently it was binding in all the Southern segregated states but not in Northern states where public accommodations were not segregated. The 1965 Act prohibited racial discrimination in voting by outlawing voting requirements that had historically been used to disenfranchise black voters.

Examples included literacy tests and the requirement that another registered voter in good standing with the community be required to vouch for you in order to vote. It was binding in a subset of these Southern states which did not meet the Act's requirements in terms of equality in accessibility to voting – specifically, Texas, Louisiana, Mississippi, Alabama, Georgia, South Carolina, and Virginia.[8] The 1968 Act prohibited discrimination on the basis of race or national origin in housing; individuals and neighborhoods would no-longer be able to deny sale, rental, or financing on these bases. It was binding across the country, as such discrimination had not been limited to the South.

What were the effects of these laws on attitudes toward blacks? Unfortunately, the ANES doesn't start asking relevant questions until the mid-1960s – too late to use for a dynamic specification that allows for observing potential pre-trends. Gallup, fortunately, began asking a relevant question in the 1950s: "If your party nominated a generally well-qualified man for president and he happened to be black, would you vote for him?" This question provides the best information available in this era at reasonably high frequency on attitudes of the general white population toward black people.

It is worth noting that all three acts were, additionally, binding only to the extent that there was any black population in the area. That is, an area that was nearly all-white would scarcely have been affected by these laws; for example, desegregation in public accommodations would not mean having to serve any blacks. Life for the white populace would continue virtually unchanged. Not so in a place that was 40% black. Consequently, it is necessary to interact the law variable with the black share of population in this setting.

Columns (1) through (3) of Table 7 reveal that, indeed, the Civil and Voting Rights Acts of

[8] A handful of counties in other states – principally North Carolina and Florida – were also bound by the 1965 Act. Whether I exclude these from the analysis or simply mark them as untreated does not meaningfully change the results. The Act was later amended in 1975 to encompass additional jurisdictions, as analyzed in Ang (2019).

Table 7: Other Social Policy Laws

	(1) Civil Rights Act 1964	(2) Voting Rights Act 1965	(3) Civil Rights Act 1968	(4) Gay Marriage Bans	(5) Gay Marriage Legal-ization	(6) Gun Control	(7) Marijuana Legal-ization	(8) Interracial Marriage Legal-ization	(9) Death Penalty Legal-ization	(10) Abortion (Roe v. Wade)
Outcome:	Would Vote for Black President	Would Vote for Black President	Would Vote for Black President	Feeling Thermom: Gays	Feeling Thermom: Gays	Support Gun Permits	Support Marijuana Legalization	Support Interracial Marriage	Support Death Penalty	Support Abortion
Sub-population:	Whites	Whites	Whites	All	All	All	All	Whites	All	All
Law Indicator	0.234***	0.258***	0.436***	2.882***	-4.757***	-0.033**	-0.042**	-0.085***	-0.061***	-0.088***
	(0.067)	(0.059)	(0.030)	(1.014)	(1.470)	(0.014)	(0.018)	(0.027)	(0.014)	(0.028)
Law Indicator	-0.892***	-0.989***	-0.511***							
* Black Pop. Share	(0.357)	(0.340)	(0.181)							
Dependent Var. Means	0.368	0.368	0.368	42.267	42.267	0.767	0.303	0.123	0.689	0.247
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Years of Data	1958-1978	1958-1978	1958-1978	1984-2012	2012-2016	1986-2016	1973-2016	1958, 1968	1974-1980	1972, 1976
Clustering	State	State	State	State	State	State	State	State	State	State
Observations	23,100	23,100	23,100	20,332	8,057	25,778	36,099	2,891	8,451	4,296

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Data for Civil Rights Act and interracial marriage regressions is from Gallup. Data for gay marriage and abortion regressions is from the ANES. Data for gun control, marijuana, and death penalty regressions is from the GSS. Civil Rights Acts and interracial marriage regressions are restricted to whites only because whites were initially the only group of individuals asked these questions by Gallup. In the case of the Civil Rights Acts, their provisions would only be binding in places with any black population, so it is necessary to interact the indicator variable for the law with black population share. The outcome variable in columns (1) through (3) is an indicator for whether respondents would be willing to vote for a black candidate for president. The outcome variable in columns (4) and (5) is a 0-to-100 feeling thermometer measuring respondents' attitudes toward gay people. The outcome variables in columns (6) through (10) are indicators for individuals' support of the corresponding issue.

Figure 9: Dynamic Differences-in-Differences –
Effects of Other Major Social Policy Laws on Corresponding Attitudes

Note: Year 0 corresponds to the year the relevant law took effect. In the top-left panel, the space between some coefficients is not to (time) scale because Gallup did not always ask the relevant question at consistent intervals in a way comparable to academic survey datasets such as the ANES or the GSS. Furthermore, in the top-left panel the plotted coefficients are the interaction terms between black population share and an indicator for the 1965 Voting Rights Act. As discussed in more detail in the body text of the paper, this is because – unlike the other laws studied here – the Civil Rights Acts of the 1960s were only binding where black population actually existed.

the 1960s engendered a strong and significant backlash, with attitudes toward blacks becoming more negative. Notably, this occurs only in areas with a black population, which is sensible for the aforementioned reason – in places with no black population, when whites were compelled to desegregate public accommodations or surrender the vote to blacks, they effectively weren't compelled to do anything. They may possibly have gotten to experience the "warm glow" that came with patting themselves on the back for being a part of the new paradigm of racial equality, without having to undergo any real lifestyle changes whatsoever.

The top-left panel of Figure 9, which focuses on the Voting Rights Act, shows that there are no visible pre- trends prior to this effect (the Figure plots the interaction coefficient between the legislation and black population share), and consistent with both the ERA case and the model's implications, the effect constitutes a sharp level shift in the immediate aftermath of the law's implementation. These findings are consistent with the historical record and anecdotal accounts of the era. The South of the 1960s was marked by "massive resistance" to desegregation on the part of white southerners and an increase in the popularity of explicitly racial rhetoric on the part of white southern politicians. Restaurant owner Lester Maddox, for example, won the office of governor in Georgia in 1966 after his public profile was elevated when he brandished an axe handle and chased off black patrons seeking to be served in his restaurant. Apparently – these findings would suggest – such politicians were catering to the hardened preferences of their constituents.

7.2 Gay Marriage Bans and Legalizations

Gay marriage has been another of the biggest and most contentious social policy debates of the past several decades. Beginning in the 1990s and extending into the 2000s, there was a push spearheaded by conservative activists for state Defense-of-Marriage Acts and Defense-of-Marriage Amendments (DOMAs). These laws defined marriage as exclusively between a man and a woman and consequently explicitly proscribed gay marriages. The movement started slowly but gathered strength in the early 2000s – particularly after Massachusetts legalized gay marriage in 2004. In that year alone, 13 states passed such an amendment. At their peak in 2012, 33 states had a DOMA in effect. Unlike the state ERAs, they were almost uniformly successful in referenda, with only two ever failing (Arizona in 2006 and Minnesota in 2012). Even California – often regarded as amongst the most liberal states – passed one in 2008.

California, however, would mark the beginning of the end for the DOMA movement, as it was the first such amendment to be totally held up by courts and not implemented. Challenges to other DOMAs were soon mounted across the states, and many state courts struck down DOMAs and legalized gay marriage in 2013 and 2014. Then, only three years after the number of DOMA states peaked, the Supreme Court struck down all DOMAs and legalized gay marriage nationwide in *Obergefell v Hodges* (2015). Because the DOMAs were rolled out in a staggered fashion and because some states had struck down their own DOMAs and legalized gay marriage before the Supreme Court decision did so nationwide, state variation was generated *in both directions* with regard to gay marriage law.

Unlike the the ERA and the Civil Rights Acts, the DOMAs were fundamentally conservative in nature. The legalization of gay marriage was liberal. This offers a unique opportunity, essentially within-law, to study whether backlash occurs against laws in both ideological directions. Since the late 1980s and early 1990s, the ANES has asked questions about attitudes toward gay people. It has repeatedly asked a question about one’s general “feeling thermometer” toward gays – whereby respondents are asked to rate how warmly they feel toward gay people on a scale of 0 to 100. It has also asked questions about attitudes toward gays serving in the military and adoption of children by gays. I study the effects of the implementation – and then the repeal – of the DOMAs on these attitudes.

Column (4) and column (5) of Table 7 suggest that indeed backlash does occur against both liberal and conservative laws. DOMAs induce warmer attitudes toward gays and more support for gays serving in the military and adopting children, as shown in the former table. The striking down of DOMAs and consequent legalization of gay marriage does the opposite – inducing more negative attitudes toward gays and (marginally) less support for gays adopting children – as shown in the latter table.[9] The top-right panel of Figure 9 shows the dynamic specification in this setting; once again, backlash was not occurring prior to the law's passage. While there is some evidence of differential attitudes prior to the law change in states legalizing gay marriage, this actually goes in the *opposite* direction of backlash.

7.3 Gun (De-)Control

Gun control constitutes another major social policy debate that has played out over the past few decades in U.S. politics. The debate over concealed carry is one of the central policy debates within the issue of gun control. This concerns the ability of individuals to legally carry a concealed firearm on their person. These laws have been relaxed over time. In 1986, only 9 states were either Unrestricted or Shall-Issue states – states where concealed-carry is allowed with minimal regulatory impediment. By 2020, 42 states were. Did relaxation of gun control induce a backlash?[10]

While the ANES did not ask a question about gun control until more recently, the GSS has

[9] The question on military service was discontinued in 2016.
[10] Concealed-carry policy changes have only moved in a less restrictive, rather than more restrictive, direction over the past several decades, preventing analysis of concealed-carry policy changes in the opposite direction.

asked a Yes/No question about supporting gun permits for decades. This is conducive to analyzing the effects of gun control relaxation on attitudes toward gun control. Column (6) of Table 7 reveals that here, too, there is backlash. The relaxation of gun control leads to more support for gun control. The bottom-right panel of Figure 9 shows the non-existence of statistically-significant pre- trends; as in the other cases, the natural experiment appears to be a clean one. It is worth highlighting that, like the DOMAs, gun control relaxation is a policy typically advocated by conservatives. So here again I find evidence of backlash by against a conservative law change – backlash does not appear to be confined to laws that are at certain points along the political spectrum.

7.4 Marijuana Legalizations

Debates over drug policy have been yet another important front in the "culture war" that makes up the U.S. social policy landscape. Liberals typically support decriminalization/ legalization of at least some drugs, while conservatives typically oppose such policies. Since the 1990s, medical marijuana has increasingly been legalized at the state level, and it currently enjoys that status in 33 states.[11] 17 of these legalizations occurred by referendum; 16 occurred through the state legislature, with the legalizations by referenda occurring earlier on average (2005) than those by legislature (2012). It is important to note that, unlike the other laws profiled in this paper, there was a substantial implementation lag on medical marijuana availability after the law changed – in some cases over 4 years. Consequently, I also obtain the implementation dates (when the first marijuana dispensaries began to operate) for all of the aforementioned legalizations from local news reports, and I use these dates in my regressions.

[11] Recreational marijuana, too, has been legalized in a much smaller handful of states, but it had only been rolled out in two by the time of the 2016 wave of the GSS – not conducive to statistical analysis.

The GSS has asked a simple Yes/No question on attitudes toward marijuana legalization since 1973, which lends itself well to analyzing the effects of these legalizations on attitudes. Column (7) of Table 7 reveals that, indeed, here too there exists a backlash. Marijuana legalization reduces support for marijuana legalization. The bottom-left panel of Figure 9 shows that no significant pre-trends exist in this case, either, though the effect is slightly noisier than some previous laws.

7.5 Supreme Court Potpourri: Interracial Marriage, Abortion, and the Death Penalty

One of the reasons ERA opponents were so concerned about the ERA was because it would give the Supreme Court "another set of words to work with" in an era where the court had become known for rapid and often highly unexpected liberal decisions that had striking implications for the social policy in the United States. Amongst these 1960s/early 1970s court decisions was abortion (*Roe v. Wade*, 1973) – another of the most salient and substantial U.S. social policy debates of the past several decades. This was not the only one, though – the Supreme Court also struck down the practice of prayer in public schools in 1962 (*Lee v. Weisman*), struck down bans on interracial marriage in 1967 (*Loving v. Virginia*), and struck down use of the death penalty in 1972 (*Furman v. Georgia*) – only to re-institute it 4 years later (*Gregg v. Georgia*).

Likely because these decisions were fairly unexpected, limited data exists on public opinion about these issues before the decisions were handed down. For example, Gallup never asked a question about support for school prayer – a very common practice across the country – prior to the court's 1962 decision banning it nationwide. Anecdotally, it is known to be a decision that inspired much consternation amongst a still-very-religious U.S. public, but the lack of data prevents difference-in-differences analysis. The other cases are somewhat more opportune. Gallup asked about interracial marriage, which was banned in some states and legal in others, precisely once before the 1967 decision. Abortion had only been legalized at the state level in some states within the 5 years prior to *Roe v. Wade*. Gallup asked about attitudes toward abortion in 1969, but the majority of the state legalizations occurred between that year and 1972, leaving little variation. Fortunately, the ANES asked about attitudes toward abortion right on the eve of *Roe v. Wade* in late 1972 and then repeatedly thereafter. Finally, the GSS began asking questions about attitudes toward the death penalty in 1975 – after the variation induced by its ban but just prior to the variation induced by its re-institution. This yields just enough data for a static difference-in-differences specification in each of these three cases, but does not permit examining any potential pre-trends. Still, the fact that these decisions were handed down to the states by the federal government rather than taken on the states' own initiative should be encouraging with regard to their exogeneity.

Columns (8) through (10) of Table 7 show that each of these law changes generated significant backlash.[12] The legalization of interracial marriage appears to have reduced support for interracial marriage; the legalization of abortion appears to have reduced support for abortion; and the re-institution of the death penalty appears to have reduced support for the death penalty. Backlash truly does seem to be a general phenomenon across the breadth of social policy laws.

7.6 Economic Policy – State Tax Changes and State Minimum Wage Increases

What about economic policy? Does it generate backlash? All the aforementioned variation

[12] Analysis is restricted to whites only for the interracial marriage case because Gallup only asked whites, not minorities, for their attitudes toward interracial marriage the first time the question was asked.

has come from social policy law. Indeed, the extended model of Section 3.3 suggested that backlash should be stronger (i) for laws on issues to which people have deep, emotional or identity-based connections and (ii) for laws where penalties/enforcement are minimal or ill-defined. Both of these would seem to apply most clearly to social policy laws. Most families probably don't have a deep, identity-based connection to a specific tax rate or the level of the minimum wage – and to the extent that the low tax rates do matter a lot to some families, they may still be reluctant to inculcate their children with a preference for tax-evasion because that would run the risk that they (or their children) are heavily penalized for such actions.

Still, there exists plentiful state-level variation over time on income tax rates and minimum wages, and the ANES asks a battery of questions pertaining to taxation and the role of the government in the economy. Using this variation, Panels 1 and 2 of Table 8 show, respectively, that there is no evidence of backlash in terms of *any* of these outcomes for either tax changes or minimum wage increases – regardless of whether I restrict to border counties or use only federally-induced variation in the minimum wage.[13] This provides some suggestive evidence that, indeed, backlash does not survive the leap from social to economic policy.

8 Conclusion

I find substantial and widespread evidence that laws do indeed affect the attitudes held by the public. However, instead of nudging the public in the direction of the law, the effect is one of persistent backlash. I first set up a simple model in which families care about inculcating their children with ideological preferences similar to their own and the ideological preferences of

[13] For each state, the binding minimum wage is the maximum of the state minimum and the federal minimum. Many states have minimum wages above the federal minimum, but not all do, so it is possible to restrict solely to federally-induced minimum wage changes for plausibly greater exogeneity.

Table 8: Economic Policy Laws

Panel 1: Tax Changes	(1) Government Waste Taxes	(2) Government Waste Taxes	(3) Government Healthcare	(4) Government Healthcare	(5) Government Job Guarantee	(6) Government Job Guarantee	(7) Government More Services	(8) Government More Services
Tax Change	-0.002	-0.007	0.002	-0.004	0.004	0.004	0.006**	0.016***
	(0.003)	(0.007)	(0.003)	(0.007)	(0.003)	(0.005)	(0.003)	(0.004)
Individuals in Sample	All	Border Residents	All	Border Residents	All	Border Residents	All	Border Residents
Observations	26,596	9,600	18,412	6,733	25,913	9,475	26,284	9,586

Panel 2: Minimum Wage Changes	(1) Government Waste Taxes	(2) Government Waste Taxes	(3) Government Healthcare	(4) Government Healthcare	(5) Government Job Guarantee	(6) Government Job Guarantee	(7) Government More Services	(8) Government More Services
MW Change	-0.052***	-0.086***	-0.004	0.024	0.007	0.030***	0.005	0.022†
	(0.012)	(0.015)	(0.015)	(0.016)	(0.018)	(0.011)	(0.018)	(0.012)
Set of MW Changes	All	Fed-Induced	All	Fed-Induced	All	Fed-Induced	All	Fed-Induced
Individuals in Sample	All	All	All	All	All	All	All	All
Observations	33,471	26,993	22,700	18,560	32,636	27,907	26,284	21,606

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. All regressions use ANES data and include state fixed-effects and year fixed-effects, mirroring the main specifications. Columns (1) and (2) use a question about whether the government wastes taxes as the outcome. Columns (3) and (4) use a question about whether the government should provide healthcare. Columns (5) and (6) use a question about whether the government should provide a job guarantee for citizens. Columns (7) and (8) use a question about whether the government should provide more (as opposed to fewer) services. In some cases, these questions included multiple possible responses (e.g., "Yes", "No", "Not Sure"). In all such cases, I generated an indicator variable corresponding to the "Yes" response.

children are formed by a weighted average of parental actions and the law. I show that, in this setting, the optimal action in response to a liberal (conservative) law-change is for parents to shift their actions in a more conservative (liberal) direction. There is a trade-off between public and private pressure, which manifests itself in a "social crowd-out"-type mechanism. A law that clashes with a family's ideological preferences places the persistence of that family's preferences into the next generation under threat. Their children will move away from their ideology and toward the law – unless the family pushes back against it. Meanwhile, if the law moves closer to a family's ideological preferences, the family can ease up somewhat in pushing its ideology onto its children and rely on the state to do so. Consequently, across the ideological spectrum, families move in the opposite direction of the law – backlash.

Empirically, the leading example I investigate is that of the state Equal Rights Amendments of the 1970s, which aimed to legislate gender equality. Amongst the most hotly-debated issues of its time, the ERA barely failed ratification as an amendment to the U.S. Constitution, but an ERA was successfully added to the constitutions of more than half of all states. Using data on attitudes toward gender equality from the American National Election Studies (ANES) along with a difference-in-differences identification strategy, I find that passage of a state ERA actually leads to sharp reductions in the attitudes men express toward male/female equality. These findings are robust to a border-county identification strategy, state-specific linear time trends, dynamic difference-in-differences, various permutation tests, the wild bootstrap-t procedure, and a restriction to the closest ERA referenda. I also find evidence that this backlash translates into material outcomes – shifting voting patterns toward the Republican party and shifting norms within marital relationships. Beyond this headline result of backlash, the various subtler implications of the model also hold true – for example, that backlash is strongest amongst those with children, that the backlash is transmitted successfully to the next generation, and that backlash occurs amongst both liberals and conservatives. Furthermore, I present evidence against a variety of alternative mechanisms. Neither economic factors, ramped-up persuasion efforts through the media, anger/spite toward government, nor the campaign leading up to the law are found to be responsible for the backlash.

Finally, I expand my focus beyond the ERA. I show that significant backlash has resulted from virtually every major social policy law of the past half-century in the United States, just as the model would predict. The laws I examine include the Civil Rights Acts of the 1960s, the legalization of abortion in the 1970s, the relaxation of gun control beginning in the 1980s, the Defense-of-Marriage Acts of the 1990s, the legalization of marijuana beginning in the 2000s, the legalization of gay marriage in the 2010s, and more.

The fact that backlash has been so systematic – and the fact that it can lead to material consequences – suggests that social policy laws, be they liberal or conservative, may consistently be accompanied by an additional and non-trivial cost that has heretofore been largely overlooked. More precisely, laws come with a functional component – specifying a crime and the punishment that will be enforced for it – and an expressive component – signaling the beliefs and norms of the society that instituted the law. This paper has argued and presented evidence that the expressive component triggers systematic backlash, which suggests that policymakers should consider the extent to which a law will be functional or expressive. Will it, like the Civil Rights Acts, generate a strong backlash that nonetheless pales in comparison to the direct, functional benefit of providing a large portion of the citizenry voting rights and the right to equal public accommodation for the first time? Or will it, like the state ERAs, generate massive backlash that seemingly overwhelms small direct effects? Asking these questions can help shape the efficacy of future social policy.

References

Abdelgadir, A. and Fouka, V. (2020). “Political Secularism and Muslim Integration in the West: Assessing the Effects of the French Headscarf Ban.” *American Political Science Review*, 114(3): 707-723.

Acemoglu, D. and Jackson, M. O. (2017). “Social Norms and the Enforcement of Laws.” *Journal of the European Economic Association*, 15(2): 245-295.

Acemoglu, D. and Robinson, J. A. (2008). “Persistence of Power, Elites, and Institutions.” *American Economic Review*, 98(1): 267-93.

Alesina, A., Baqir, R., and Easterly, W. (1999). “Public Goods and Ethnic Divisions.” *Quarterly Journal of Economics*, 114(4): 1243-1284.

Alesina, A. and Fuchs-Schündeln, N. (2007). “Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences.” *American Economic Review*, 97(4): 1507-1528.

Alesina, A. and Giuliano, P. (2015). “Culture and Institutions.” *Journal of Economic Literature*, 53(4): 898-944.

Alesina, A. and La Ferrara, E. (2005). “Ethnic Diversity and Economic Performance.” *Journal of Economic Literature*, 43(3): 762-800.

Ang, D. (2019). “Do 40-year-old Facts Still Matter? Long-run Effects of Federal Oversight under the Voting Rights Act.” *American Economic Journal: Applied Economics*, 11(3): 1-53.

Becker, S.O. et al. (2016). “The Empire is Dead, Long Live the Empire! Long-run Persistence of Trust and Corruption in the Bureaucracy.” *The Economic Journal*, 126(590): 40-74.

Benabou, R. and Tirole, J. (2011). “Laws and Norms.” NBER Working Paper No. 17579.

Bertrand, M., E. Duflo, and Mullainathan, S. (2004). “How Much Should We Trust Differences-in-Differences Estimates?” *Quarterly Journal of Economics*, 119(1): 249-275.

Bisin, A. and Verdier, T. (2001). “The Economics of Cultural Transmission and the Dynamics of Preferences.” *Journal of Economic Theory*, 97(2): 298-319.

Black, D. (1948). “On the Rationale of Group Decision-Making.” *Journal of Political Economy*, 56(1): 23-34.

Borusyak, K. and Jaravel, X. (2017). “Revisiting Event Study Designs.” Working Paper.

Bowles, S. and Polania-Reyes, S. (2012). “Economic Incentives and Social Preferences: Substitutes or Complements?” *Journal of Economic Literature*, 50(2): 368-425.

Bursztyn, L., Egorov, G., and Fiorin, S. (2017). "From Extreme to Mainstream: How Social Norms Unravel." NBER Working Paper No. 23415.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics*, 90(3): 414-427.

Cameron, A. C. and Miller, D. L. (2015). "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources*, 50(2): 317-372.

Cooter, R. (1998). "Expressive Law and Economics." *The Journal of Legal Studies*, 27(S2): 585-607.

Dal Bó, P., Foster, A., and Putterman, L. (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review*, 100(5): 2205-29.

Downs, A. (1957). "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy*, 65(2): 135-150.

Fouka, V. (2020). "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I." *Review of Economic Studies*, 87(1): 204-239.

Freud, S. (1909). "Analysis of a Phobia in a Five-Year-Old Boy." In *Standard Edition*, 10: 1-149. London: Hogarth Press.

Gladstone, L.W. (2004) "Equal Rights Amendments: State Provisions." *CRS Report for Congress*, Order Code RS20217.

Greif, A. and Tadelis, S. (2010). "A Theory of Moral Persistence: Crypto-morality and Political Legitimacy." *Journal of Comparative Economics*, 38(3): 229-244.

Gruber, J. and Hungerman, D. M. (2008). "The Church versus the Mall: What Happens When Religion Faces Increased Secular Competition?" *Quarterly Journal of Economics*, 123(2): 831-862.

Ferreira, F. and Gyourko, J. (2014). "Does Gender Matter for Political Leadership? The Case of US Mayors." *Journal of Public Economics*, 112: 24-39.

Goldin, C. and Katz, L. F. (2002). "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions." *Journal of Political Economy*, 110(4): 730-770.

Kahan, D. M. (1997). "Between Economics and Sociology: The New Path of Deterrence." *Michigan Law Review*, 95(8): 2477-2497.

Klarner, C., et al. (2013). State Legislative Election Returns Data, 1967-2010. Dataset. Harvard Dataverse.

Levant, R. F. and Richmond, K. (2008). "A Review of Research on Masculinity Ideologies using the Male Role Norms Inventory." *The Journal of Men's Studies*, 15(2): 130-146.

Lowes, S. et al. (2017). "The Evolution of Culture and Institutions: Evidence from the Kuba Kingdom." *Econometrica*, 85(4): 1065-1091.

Mansbridge, J. J. (1986). *Why We Lost the ERA.* Chicago, IL: University of Chicago Press.

Moen, P., Erickson, M. A. and Dempster-McClain, D. (1997). "Their Mother's Daughters? The Intergenerational Transmission of Gender Attitudes in a World of Changing Roles." *Journal of Marriage and the Family* 59(2): 281-293.

Peisakhin, Leonid V. 2010. "Living Historical Legacies: The 'Why' and 'How' of Institutional Persistence – The Case of Ukraine." Working Paper.

Perlstein, R. (2015). *The Invisible Bridge: The fall of Nixon and the rise of Reagan*. New York, NY: Simon and Schuster.

Posner, E. A. (1998). "Symbols, Signals, and Social Norms in Politics and the Law." *The Journal of Legal Studies*, 27(S2): 765-797.

Posner, E. A. (2000). *Laws and Social Norms*. Cambridge, MA: Harvard University Press.

Roland, G. (2012). "The Long-Run Weight of Communism or the Weight of Long-Run History?" In ed., G. Roland, *Economies in Transition*, 153-171.

Rosenberg, R. (2008). *Divided lives: American women in the twentieth century*. New York, NY: Macmillan.

Schlafly, P. (1972). "What's Wrong with 'Equal Rights' for Women?" *The Phyllis Schlafly Report* 5(7): 1-4.

Shiller, R. J. et al. (1992). "Hunting for Homo Sovieticus: Situational versus Attitudinal Factors in Economic Behavior." *Brookings Papers on Economic Activity*, 1992(1): 127-194.

Stimson, J.A. (1991). *Public Opinion in America: Moods, Cycles, and Swings*. Boulder, CO: Westview.

Sunstein, C. R. (1996). "On the Expressive Function of Law." *University of Pennsylvania Law Review*, 144(5): 2021-2053.

Sutter, M., Haigner, S., and Kocher, M. G. (2010). "Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations." *The Review of Economic Studies*, 77(4): 1540-1566.

Tabellini, G. (2008). “The Scope of Cooperation: Values and Incentives.” *Quarterly Journal of Economics*, 123(3): 905-950.

Tyran, J. R. and Feld, L.P. (2006). “Achieving Compliance when Legal Sanctions are Non-Deterrent.” *The Scandinavian Journal of Economics*, 108(1): 135-156.

Young, P. A. (1995). “A Transgenerational Study of Sex Role Norms and Development in Adult Men.” Dissertation, Pepperdine University.

Chapter 2

Erroneous Beliefs and Political Approval: Evidence from the Coronavirus Pandemic[1]

1 Introduction

The question of whether politicians are rewarded for good performance and penalized for bad performance is a matter of paramount political-economic importance. This question – central to models of retrospective voting – is crucial because the existence of such rewards/penalties may incentivize elected leaders to pursue socially beneficial outcomes, helping ensure the accountability of elected government to its constituents and the healthy functioning of democracy. A government that is able to generate perceptions of good performance despite poor *actual* performance may be able to evade responsibility for its actions.

In order for politicians to be rewarded or penalized in this way, however, it is first necessary that public perceptions of performance be at least somewhat accurate. A crucial challenge is that it is often difficult to objectively measure performance. First, there are a multitude of dimensions of both the policies pursued by politicians and the outcomes over which they preside – many of which may be difficult to measure in any objective sense. Second, it can be unclear what role politicians have on each of these dimensions. For example, a growing literature studies the extent to which leaders have actual effects on economic growth, and its findings have been mixed. All of these factors may lead to imperfectly-accurate perceptions of performance.

The precise questions that emerge from these observations are (i) whether voters do actually

[1] Joint with Matthew Lilley, Harvard University.

have accurate beliefs about performance, (ii) whether politicians are rewarded for having good outcomes or merely for *being perceived* as having good outcomes, and (iii) whether inaccurate beliefs yield any cost to society. To answer these questions, we study the Coronavirus Pandemic of 2020-21, which we regard as a setting highly amenable to the investigation of our research questions. During the pandemic, the entire apparatus of state government shifted its priorities toward managing and mitigating coronavirus. Plentiful data on coronavirus cases, testing, and deaths was available at the state level (and finer geographies) on a daily basis. Governors possessed an extraordinarily wide degree of latitude to implement policy responses of their choosing, with comparatively little encumbrance from legislatures. Meanwhile, they also became the highly-visible public faces of their states' efforts, with some – such as New York's Andrew Cuomo and California's Gavin Newsom – holding daily or weekly coronavirus briefings. Furthermore, many opinion polls throughout the period focused specifically on public approval of their governor's handling of the coronavirus pandemic. All this renders the pandemic an ideal setting for studying the accuracy of public perceptions about the performance of their leaders – and the implications of that accuracy.

We conducted an incentivized mTurk survey at the end of July 2020 (during the pandemic's "summer wave"), primarily asking respondents to provide their best guess of how pairs of states performed relative to one another in terms of deaths per-capita. We additionally asked a variety of demographic questions, questions about political identification, and benchmarking questions designed to gauge respondents' perceptions of how well the states *should have* performed, given pre-existing characteristics such as their population density and setting aside factors of leadership/political competence. The survey consisted of approximately 400 mTurk Masters located in the United States, each of whom was compensated a base rate of $1.50 along with a potential incentive bonus for answering the primary questions correctly. We subsequently ran an identical survey three months later, at the end of October 2020 (during the beginning of the pandemic's fall/winter wave).

We find that individuals perform better than random guessing in their pairwise comparisons of state performance – but not substantially better. Respondents only correctly guessed which state performed worse 63.4% of the time. Respondents tended to think that states like Florida and Texas – which received substantial critical media coverage – performed substantially worse than they actually did. We investigated whether there existed any in-group bias in beliefs, finding at most weak evidence of Republican (Democratic) respondents holding biased beliefs about how positively Republican (Democratic) states performed, in relative terms. These results were fairly stable across both the July and October waves of the survey.

Next, we turn to the question of whether politicians are rewarded for good outcomes or merely perceptions of good outcomes. To do this, we regress respondents' guesses about death rates on state fixed-effects in order to provide a measure of how badly people think each state is doing. Next, using opinion-polling data from The COVID States Project on state-level approval of governor handling of the pandemic, we regress these measures of approval on the actual state death rate and these aforementioned fixed effects that capture beliefs. We find that it is not the actual death rate – but rather beliefs about the death rate – which drive governor approval. Controlling for beliefs about the death rate, the effect of a higher actual death rate on approval is actually *positive*, consistent with a potential role for governor media visibility (which tended to be higher in harder-hit states) in boosting approval. As an alternative approach, instead of using the COVID States opinion-polling data on approval of governor coronavirus handling, we use an identical question internal to the survey. This yields the same result – incorrect beliefs strongly affect political approval. All these results, too, are highly stable across both the July and October waves of the survey.

We argue that, in these regressions, making a causal interpretation is reasonable, as reverse causality would entail disliking the politician in question and consequently having negatively-biased beliefs about the state's coronavirus performance. If this was widely the case, we should expect to observe substantial partisan in-group bias in beliefs about states' performance, but as previously noted, we find minimal such bias in the data. Furthermore, the result is robust to the addition of a broad variety of demographic and political control variables, which should net out effects due to pre-existing attitudes to politicians. However, to gain further evidence on causality, we ran an additional survey in December 2020 – this one leveraging experimental variation. Firstly, given that respondents are imperfectly informed about state performance, we elicit governor approval conditional on different counterfactual levels of performance in terms of coronavirus deaths. We find that conditional governor approval is falling sharply in the hypothetical death rate. Second, we shock respondent beliefs about their state's performance (by eliciting their priors and providing them with the true information), and elicit their ex post governor approval. Exogenously inducing higher beliefs about the number of deaths corresponds to lower governor approval. That is, in both experiments, respondents' approval of their governor moves in the expected direction.

Finally, using data from SafeGraph on the median amount of time individuals in various states spent in the home versus outside the home, we show that it is beliefs about state's performance – not actual state performance – which have bearing on social distancing behavior. Individuals engage in less social distancing when their state is erroneously perceived to have performed better in terms of coronavirus deaths. We take this as further evidence that the measured beliefs are real and as evidence that erroneous beliefs may distort behavior in a way potentially harmful to society.

2 Literature Review

Our work relates most directly to the broad literature on retrospective voting, which originated over a half-century ago. Key (1966) seminally argued that "voters are not fools" – that is, that they update their beliefs and actions based on government performance, rewarding or punishing politicians accordingly. Key's informal intuition was subsequently formalized in models by Barro (1973) and Ferejohn (1986). In these models, by re-electing high-performing politicians and voting out poorly-performing ones, voters incentivize good performance by politicians (and thus good outcomes). These theories represented an important divergence from the theretofore standard conception of the voter as mostly lacking in information and voting entirely on the basis of promised future political outcomes rather than past performance. On the empirical front, a large subset of this literature, beginning with Kramer (1971), Fair (1978), and Fiorina (1981), has studied whether voters reward or penalize politicians for economic outcomes, which are taken as objective performance indicators.

Later theoretical frameworks enriched the mechanisms underlying retrospective voting. Persson and Tabellini (2002) and Duch and Stevenson (2008) view retrospective voters as learning about incumbent quality through incumbent performance during his/her period in office. Voters then choose between re-electing an incumbent leader of known quality or voting the incumbent out of office and taking a new draw from the quality distribution. Ashworth (2005) models the effects that such a mechanism have on *politician* decision-making and effort allocation over the course of a career. Recent empirical papers have exploited a variety of natural experiments (e.g., Alt et al. 2011, Gasper and Reeves 2011, Bechtel and Hainmueller 2011, Reeves and Gimpel 2012, Stokes 2016) and controlled experiments (e.g., Malhotra and Kuo 2008, Malhotra and Margalit 2014).

An important subset of this literature has focused on how rational inattention, behavioral biases and cognitive limitations might interact with the concept of retrospective voting. In a world full of rich information, voters may find it suboptimal to process all relevant information and instead choose to rely on heuristics. This strand began with the observation (initially made by Kramer 1971, Fair 1978, and Tufte 1978) that the election-year economy appeared to have larger impacts on voting behavior than conditions in other years of the incumbent's tenure, suggesting a form of availability bias. Hill et al. (2012) and Healy and Lenz (2014) examine this phenomenon in more detail. More generally, it has been argued that voters reward or punish politicians because they are happy or sad for reasons that have nothing to do with incumbent performance, such as foreign economic conditions or football games (Schwarz and Clore 1983, Achen and Bartels 2004, Wolfers 2009, Healy et al. 2010, Campello and Zucco 2016, Busby et al. 2017). These findings are often attributed to a combination of behavioral biases and difficulties in attributing responsibility for outcomes.

Our work relates most closely to this strand of the literature, which probes the behavioral contours of retrospective voting. One notable contrast is that whereas many of these prior studies focus on the effects of irrelevant outcomes (e.g. football games) on voting, our setting enables us to analyse how voters respond to politician actions that they may legitimately (and plausibly) believe are able to substantially impact outcomes. Further, we are able to precisely distinguish between the true and false components of beliefs held by the public about an important outcome (coronavirus mortalities in their home state), thereby enabling us to directly assess the importance of behavioral considerations relative to rational considerations underlying the mechanism of retrospective voting and reward/punishment of politicians. We also contribute by examining a domain (mortalities) distinct from the standard economic outcomes most typically studied.

3 Data and Identification

3.1 mTurk Survey Data

Between July 22nd and August 10th, 2020, we ran a survey on Amazon Mechanical Turk asking participants, for 10 randomly-drawn pairs of states, to guess which state in the pair had fared worse up to that point in terms of coronavirus mortalities per capita.[2] The question about each pair was immediately followed-up by a more precise question asking *how much worse*, as a percentage, they believe their chosen state had fared. Next, for 5 randomly-drawn pairs of states, participants were asked their beliefs about which state *would have* performed worse (and how much worse) due to pre-existing non-political factors such as population density, population age, presence of international travelers, and anything else they deemed relevant. Also included in the survey were demographic questions on sex, age, race, education, income, and state of residence. Political questions – respondents' Presidential election vote in 2016, their party identification, their position on a 7-point ideological spectrum, and the extent of their approval for their governor's handling of the pandemic – were also asked.

The timing of our survey roughly corresponds to the height of the summer wave of coronavirus cases and deaths. Amazon Mechanical Turk is an online platform on which users can opt-in to completing various tasks in exchange for monetary compensation. Our sample was limited to "mTurk Masters," mTurk workers specifically designated by Amazon as top performers due to consistent high-quality answers. Respondents were also required to be US residents. Generally speaking, mTurk workers skew younger than the general population, but this is somewhat less true of mTurk Masters. We compensated respondents with a base rate of $1.50,

[2] The first 5 of the 10 pairs are constrained to include the respondent's home state as one of the states in the pair, since individuals may plausibly have a more accurate picture of the pandemic situation in their home state.

topped up with an incentive bonus of up to $0.50 for accuracy.[3]

To validate our results and check for consistency, we subsequently ran an identical follow-up survey on October 14th and October 15th, 2020. This timing corresponded to the beginning of the fall/winter increase in coronavirus cases and deaths. For this survey, instead of restricting participation to U.S. mTurk Masters, we restricted to U.S. mTurk workers who had completed at least 500 tasks with a success rate of at least 99%.[4] Compensation was again $1.50, with an incentive bonus of up to $0.50. The questionnaire for our survey can be found in full in Appendix B.1.

Finally, on December 21st and December 22nd, 2020, we ran an information-revelation survey experiment on mTurk. In the experiment, we randomly-assigned participants to either a treatment group, a control group, or a hypothetical group. We asked the control group of participants for their guess of coronavirus deaths and pandemic employment declines within their home state, followed by a question on the extent to which they approve of their governor's handling of the pandemic. The treatment group was provided with information on the true figures before being asked about their approval of their governor's handling of the pandemic. The disparity between their priors and the true information induces a shock to the beliefs of respondents in the treatment group, allowing us to discern the effect (if any) of beliefs on governor approval. The hypothetical group was asked a series of hypothetical questions: whether they would approve of their governor's handling of the pandemic *if* they learned that the true coronavirus death rate (or the true decline in employment since the start of the pandemic) was *X*, for a variety of values of *X* (at least one of which is true). As with the second wave of the main survey, we restricted our sample to

[3] This compensation was later increased – ultimately to a base rate of $2.50 and an incentive bonus of up to $0.75 – in order to attract additional respondents.
[4] We had exhausted the supply of U.S. mTurk Masters who were willing to take our survey at the compensation we offered.

U.S. mTurk workers who had completed at least 500 tasks with a success rate of at least 99%. The questionnaire for our survey experiment can be found in Appendix B.2.

3.2 Governor Approval and Other Outcomes

Since March 2020, The Covid States Project, a multi-university group of multi-disciplinary researchers has released a variety of periodic reports on the status of the pandemic and related indicators at the state level. Amongst these reports have been state opinion polling data on approval of governor handling of the pandemic, termed Executive Approval reports by the Project. This data is publicly-available online through the Project's website (covidstates.org), and we utilize it as our key outcome, using opinion-polling data from their July Wave with our July mTurk survey and opinion-polling data from their October Wave with our October mTurk survey.

We obtain data on each state's 2016 Presidential Election victor and margin of victory from Dave Leip's Election Atlas.

3.3 Identification

In order to identify the effects of both actual coronavirus deaths and beliefs about coronavirus deaths, we apply the following procedure. We first note that our data on our main survey question is at the state-pair level. Respondents are asked which of State A and State B they believe has experienced a higher coronavirus death rate and, subsequently, the factor by which they think deaths are higher in their chosen state. From this, we construct the logarithm of the ratio of the coronavirus death rates in the two states. The logarithmic transformation ensures the data is coherently normalised and allows for ease of interpretation of subsequent regression coefficients. Each observation can be represented as either the log of the factor, z, by which State A's coronavirus death rate exceeds State B's coronavirus death rate *or* the log of the factor $1/z$ by which State B's coronavirus death rate exceeds State A's coronavirus death rate.

From this data, we wish to extract an estimate of average beliefs about each state's death rate relative to other states. Since each observation pertains to the relative level of deaths in a state *pair*, (separately) for each survey wave we regress respondents' guesses about relative death rates on state indicator variables as follows, in order to estimate state fixed-effects:

$$\log(X_{isr}) = \gamma_s + \delta_r + u_{isr}\,, \tag{1}$$

where X_{isr} denotes the guess of respondent i about the factor by which the death rate of state s exceeds the death rate of state r, δ_r and γ_s denote state fixed-effects for state r and state s respectively, and u_{isr} is the error term. The estimated fixed effect for each state can be extracted as an estimate of beliefs regarding a state's death rate, as desired.

However, an immediate challenge arises. There is no convincing theoretical reason for any particular *rotation* of any particular observation, namely which state should be considered s and which should be considered r. Further, with a set of fixed effects for s and another for r, this regression generates two separate estimates of beliefs about each state. The two sets of point estimates will not in general be equal, will vary based on arbitrary rotation of datapoints, and it is unclear which (or what combination of them) should be interpreted as beliefs.

Fortunately, there exists a simple and elegant fix that works by negating the arbitrary nature of rotation decisions. That is, we duplicate each observation in the dataset, representing each observation with both rotations and weighting each by half in our analysis. By construction, this yields $\gamma_s = -\delta_s$.[5] The estimated γ_s vector thus provides a measure of how badly, on average, people

[5] To see this, suppose the OLS estimates are $\gamma_s \neq -\delta_s$, and note that $y_{ijk} = -y_{ikj}$. This yields fitted values (hat) y_{ijk} and y_{ikj}, and residuals e_{ijk} and e_{ikj}. (Or uhat). Consider alternate candidate solution vectors (tilde) $\gamma_s = (\gamma_s - \delta_s)/2$, (tilde) $\delta_s = (\delta_s - \gamma_s)/2$ (such that tilde $\gamma_{s\,=\,-}$ (tilde) δ_s). This yields fitted values (hat tilde) $y_{ijk\,=}$ (hat) $(y_{ijk} - y_{ikj})/2$ = (hat tilde) $-y_{ikj}$ and analogously residuals (tilde hat) $e_{ijk\,=}$ (hat) $(e_{ijk} - e_{ikj})/2$ = -(hat tilde) e_{ikj}. For any real scalars $a \neq b$, $a^2 + b^2 >$

think state s is doing in terms of coronavirus death rates, with a higher value corresponding to perceptions of higher deaths.

We exploit this same procedure to extract several other pertinent measures from the survey data. First, we again take respondents' guesses about death rates and construct the extent of erroneous beliefs held by individual i about how much higher death rates are in state s relative to r. This involves dividing the guessed ratio by the true ratio, or in logarithms, $\log(\tilde{X}_{isr}) = \log(X_{isr}) - \log\left(\frac{d_s}{d_r}\right)$, where d_s and d_r are the respective per capita death rates. Second, we use our survey question on what respondents expected the relative death rate B_{isr} *should* be in a given state pair, taking into account factors like population density and age, while putting aside factors of political competence. In these two cases cases, we re-estimate Equation (1) replacing X_{isr} with $\tilde{X}_{isr}$ and B_{isr} in turn, and extract the resulting state fixed effect estimates. This yields state-level measures of average erroneous beliefs about deaths rates, and benchmark expected death rates abstracting from political competence, for each state.

Next, with these state level measures in hand, we regress our outcomes of interest – most notably, political approval – on these estimated fixed-effects and on the natural logarithm of the *actual* state death rate from coronavirus. That is,

$$Y_s = \alpha + \beta_1 \cdot \log(DeathsPerMil_s) + \beta_2 \cdot \hat{\gamma}_s + \varepsilon_s, \qquad (2)$$

where Y_s is a state-level outcome of interest (such as governor approval rate for handling of the pandemic), $DeathsPerMil_s$ is the actual coronavirus death rate per million population, $\hat{\gamma}_s$ are the fixed-effects estimated in the preceding regression, and ε_s is the error term. Thus the effect of a

---

$2*[(a+b)/2]^2$, so this constitutes an improvement under the OLS objective function, a contradiction. Note that an analogous argument holds when $y_{ijk} = y_{ikj}$ (in which case, $\gamma_s = \delta_s$) or with an additive constant in either case ($y_{ijk} = a \pm y_{ikj}$).

1% increase in *actual deaths* on the outcome Y is given by $\beta_1/100$. The effect of a 1% increase in *believed deaths* on the outcome Y is given by $\beta_2/100$. As noted, we weight each observation by one-half, and we also use robust standard errors.[6] In specifications where we pool the first and second wave of our survey, we add a fixed-effect for the wave and cluster by state. For robustness, we run additional specifications with an assortment of demographic and political control variables added to the above regression equation. This is done to partial out any correlation of beliefs with these controls, which themselves may plausibly drive political approval.

We also run a specification where we focus on the effect of erroneous beliefs by replacing $\hat{\gamma}_s$ with our measure of erroneous beliefs about state j. This yields a regression with a slightly modified interpretation of the coefficients: β_1 now corresponds to the effect of the actual deaths, holding the error in beliefs constant (rather than holding beliefs themselves constant), while β_2 corresponds to the effect of the erroneous component of beliefs.[7] Since political approval may effectively handicap each governor based on how exposed people deem their state as being due to pre-existing factors (largely) outside of government control, we also add the measure of expected death rates abstracting from political competence as a control variable to Equation (2). We then run an alternative version of all of the above specifications wherein we use respondents' binary guess about which state has a higher death rate (as opposed to the precise continuous factor by which the state has more deaths per capita).

Finally, for the governor approval outcome, we run an alternative one-step regression leveraging our internal mTurk survey data on respondent approval of governor handling of the

[6] In practice, we obtain a separate set of belief estimates (and governor approval data) for each state-survey wave pair. Accordingly, we cluster standard errors by state.
[7] Since the true log death rate ratio $\log(d_s/d_r)$ is a state-pair (or state-pair-wave specific constant), the calculated erroneous beliefs are identical (up to an additive constant) to $\hat{\gamma}_s - log(d_s)$. As a result, in these subsequent regressions, the coefficients on this measure of erroneous beliefs, controlling for the log of actual death rates, are identical to the coefficients on beliefs controlling for the log of actual death rates.

coronavirus pandemic. Because we have individual-level data on this outcome, it is not necessary to generate state fixed-effects for use in a second-stage regression. We can instead retain the first observation for each respondent and run a version of regression equation (2) which, in place of $\hat{\gamma}_s$, directly includes the log of the respondent's guess X_{is} ($\equiv X_{isr}$) of how much worse his home state, *s*, is doing relative to some randomly-selected state *r*. That is,

$$Y_{is} = \alpha + \beta_1 \cdot \log(DeathsPerMil_s) + \beta_2 \cdot \log(X_{is}) + \varepsilon_{is} \tag{3}$$

4 Results

In Table 1, we first report some simple descriptive statistics about the characteristics of our sample. Average age, share male, and median household income of the sample are consistent with the U.S. general population. The sample, however, has a somewhat higher education, share of non-Hispanic whites, and share of liberals/Democrats than the U.S. general population. In certain specifications, we control for these variables in order to ensure that the deviations from representativeness have no effect on our results.

4.1 Accuracy/Bias in Beliefs

Survey respondents correctly guess which state had performed worse (through the date of the survey) in terms of coronavirus death rates 63.4% of the time.[8] Restricting only to state pairs involving the respondent's home leaves this figure almost exactly unchanged. Figure 1 displays a scatterplot of the relative frequency with which survey respondents guess each state had a higher

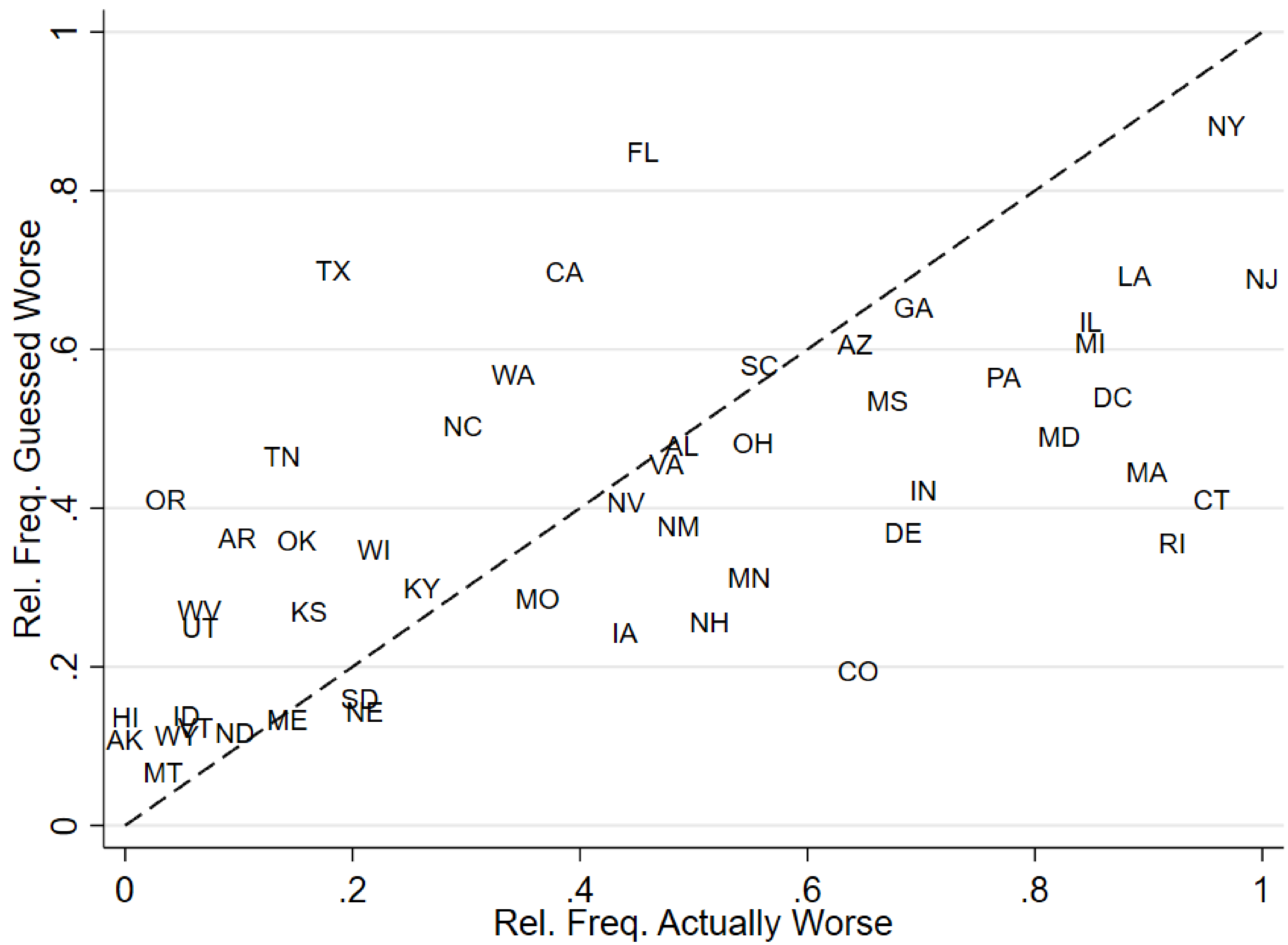
[8] Since completely uninformed random guessing would yield a 50% correct rate, this is consistent, for example, with respondents only truly knowing the answer in 26.8% of cases.

Table 1: Descriptive Statistics of mTurk Sample

	mTurk Sample	U.S. Population
Share Male	0.485	0.492
Median Age	40	38.4
Share White, non-Hisp.	0.726	0.600
Share w/ B.A. or Greater	0.491	0.331
Median HH Income	65,885	65,712
Share Clinton Voters	0.423	0.268
Share Trump Voters	0.274	0.257
Share Liberals	0.483	0.279
Share Democrats	0.426	0.354
Observations	613	-

Note: U.S. Population data is from the 2019 American Community Survey (demographic variables), election returns and turnout data (voting variables), and 2020 American National Election Study (ideology variables).

Figure 1: Scatterplot of Believed and Actual Coronavirus Death Rates



death rate (than states with which it is being compared in the pairwise questions) against the relative frequency it *actually* had a higher death rate. This reveals which states actually performed better than respondents believe (those above the 45-degree reference line) and which states performed worse. As can be seen from the scatterplot, the states with the largest positive gap between actual and perceived performance (i.e., those most erroneously perceived as performing poorly) are Texas and Florida – two states which received particularly negative media coverage despite having moderate death rates. States with the largest negative gap between actual and perceived performance (i.e., those most erroneously perceived as performing well) are Rhode Island, Connecticut, and Massachusetts – New England states which were quite intensely impacted by the first wave of the pandemic but received limited attention in the media relative to New York, which only performed slightly worse but was front-and-center in terms of media coverage in early months of the pandemic.

Given the politically-charged nature of discussions surrounding state performance during the pandemic, one might wonder whether there exists any partisan bias in perceptions of death rates. That is, do Democrats have unjustifiably positive views of the performance of Democratic states while Republicans have unjustifiably positive views of the performance of Republican states? To study this question, we regress the natural logarithm of individual respondents' *excess* believed deaths (believed minus actual) on an indicator variable for the governor's partisan alignment and an "cross-party" indicator variable for whether the governor is of the opposite political party to the respondent. The regression analysis in Table 2 follows this approach. As seen in column (1), there is strong evidence that beliefs about Republican states' death rates are excessively pessimistic relative to beliefs about Democratic states' death rates. Turning to the coefficient on the cross-party indicator variable, there is at most weak evidence of modest partisan in-group bias.

Table 2: Investigation of Partisan In-Group Bias

	(1)	(2)	(3)	(4)
Outcome: Excess Believed Deaths (Log)	Baseline	Baseline + State FEs	Baseline + Actual Deaths	Baseline + Actual Deaths + State FEs
Republican Governor	0.250***		0.009	
	(0.028)		(0.021)	
Cross-Party Governor	0.058*	0.048*	0.038	0.043*
	(0.032)	(0.025)	(0.025)	(0.023)
Deaths per Capita (Log)			-0.714***	-0.917***
			(0.015)	(0.049)
State FEs	No	Yes	No	Yes
Clustering	State	State	State	State
Observations	12,024	12,240	12,024	12,240

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

Respondents, when considering a state whose governor is of the opposite party affiliation, believe that the state's deaths per-capita are 5.8% higher relative to respondents who share a party affiliation with the governor. This coefficient is small in magnitude and, furthermore, it is only statistically-significant at the 10% level. Column (2) adds state fixed-effects, such that in-group bias is identified only from within-state variation in beliefs; namely for each state, the partisan difference in beliefs about that state held by respondents. Neither the magnitude or significance of the estimated partisan in-group bias is meaningfully altered. Columns (3) and (4) repeat the analysis of columns (1) and (2), albeit with actual log deaths per capita added as a control variable, again yielding no meaningful change in the estimates of partisan in-group bias. However, adding the control for log deaths causes the coefficient on governor party to become a tightly estimated zero. In other words the negative bias in beliefs about states with Republican governors in Column (1) is an artefact of respondents being largely unaware which states had done better, combined with the average Republican-led state having fewer per-capita deaths at the time.

4.2 Effects on Political Approval (Observational)

Having constructed aggregate beliefs about the relative death rate for each state, we turn to the key question of how beliefs about state performance affect political approval. Table 3 displays versions of the regression specification described in Section 3.3, with state-level average approval ratings of governor handling of the pandemic as the outcome variable. Column (1), however, begins with a univariate regression of governor approval on the log of the death rate. There is evidence of a *positive* (albeit slightly weak) association between the death rate and governor coronavirus approval. A 10% increase in deaths roughly translates into a 0.29% increase in governor approval. This regression, however, masks a more complex relationship. Column (2) is

Table 3: Effects of Actual and Believed COVID Deaths on Governor Approval (Continuous Beliefs)

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Governor COVID Handling Approval (Strongly Approve + Approve)	Deaths	Deaths + Beliefs	Deaths + Excess Beliefs	Deaths + Bench. + Beliefs	Deaths + Bench. + Excess Beliefs	Deaths + Bench. + Beliefs + Controls
Deaths per Capita (Log)	2.93**	6.33***	-4.95	5.94***	-15.24**	5.38***
	(1.43)	(1.32)	(3.17)	(1.23)	(7.48)	(1.38)
Believed Relative Deaths (Log)		-11.29***		-21.18***		-13.06**
		(3.40)		(7.18)		(4.94)
Excess Believed Relative Deaths (Log)			-11.29***		-21.18***	
			(3.40)		(7.18)	
Benchmark Relative Deaths (Log)				11.84*	11.84*	8.60*
				(6.65)	(6.65)	(4.45)
Political Controls	No	No	No	No	No	Yes
Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State	State
Observations	100	100	100	100	100	96

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

the specification directly corresponding to Equation (2) in Section 3.3. It reveals that higher *believed* deaths are associated with lower approval, whereas higher *actual* deaths are associated with higher approval. A 10% increase in actual deaths translates into a 0.63% increase in governor approval; a 10% increase in believed deaths translates into a 1.13% decrease in governor approval. In other words, the intuitive relationship whereby voters punish politicians for bad outcomes (here, deaths) is entirely driven by *perceptions* of the outcome, not the outcome itself. This suggests potential challenges to ensuring politicians are properly incentivized through public opinion and voting.

Column (3) contains the results of the analogous specification in which we transform beliefs into "excess beliefs" by subtracting the truth from beliefs. This re-frames the regression specification to yield a slightly different interpretation. Holding the error in beliefs constant, a 10% increase in the death rate is associated with a (non-significant) 0.43% decrease in governor approval. A 10% increase in the *false* component of beliefs about the death rate is associated with a (significant) 1.13% decrease in governor approval.

Columns (4) and (5) repeat the exercises of columns (2) through (3), with an added control capturing beliefs about benchmark death rates. These benchmarks measure how high a death rate respondents would have expected in each state given its pre-existing characteristics (e.g., population density, population age, exposure to international travelers, etc.), putting aside factors of political competence. The addition of this benchmark deaths control increases the coefficient on beliefs approximately two-fold. The positive coefficient on benchmark deaths is consistent with governors being graded on a curve based on their state's perceived inherent exposure to the pandemic. The overall conclusion is substantively unchanged. Finally, column (6) adds a variety of state-level control variables to the specification in column (5), including governor pre-pandemic

Table 4: Effects of Actual and Believed COVID Deaths on Governor Approval (Binary Beliefs)

	(1)	(2)	(3)	(4)	(5)	(6)
Outcome: Governor COVID Handling Approval (Strongly Approve + Approve)	Deaths	Deaths + Beliefs	Deaths + Excess Beliefs	Deaths + Bench. + Beliefs	Deaths + Bench. + Excess Beliefs	Deaths + Bench. + Beliefs + Controls
Deaths per Capita (Log)	2.93**	5.63***	0.68	5.49***	1.23	4.64***
	(1.43)	(1.45)	(1.93)	(1.48)	(2.46)	(1.45)
Probability Believed Worse		-20.36**		-33.38**		-11.19
		(8.36)		(12.72)		(9.13)
Probability Believed Worse than Actual			-12.05**		-10.84*	
			(5.83)		(6.34)	
Probability Benchmark Worse				13.96	-2.45	4.14
				(11.18)	(8.46)	(8.12)
Political Controls	No	No	No	No	No	Yes
Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State	State
Observations	100	100	100	100	100	96

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

Table 5: Effects of Actual and Believed COVID Deaths on Governor Approval (mTurk Governor Approval Data)

Outcome: Governor COVID Handling Approval (Strongly Approve + Approve)	(1) Deaths	(2) Deaths + Beliefs	(3) Deaths + Excess Beliefs	(4) Deaths + Bench. + Beliefs	(5) Deaths + Bench. + Beliefs + Controls 1	(6) Deaths + Bench. + Beliefs + Controls 2
Deaths per Capita (Log)	9.76***	15.64***	-0.74	16.02***	10.65***	10.40***
	(3.37)	(3.92)	(3.20)	(4.01)	(2.57)	(2.33)
Believed Relative Deaths (Log)		-18.07***		-15.99***	-15.21***	-14.43***
		(4.23)		(4.36)	(3.56)	(3.69)
Excess Believed Relative Deaths (Log)			-14.67***			
			(3.62)			
Benchmark Relative Deaths (Log)				-3.83	-5.24**	-4.58**
				(2.34)	(2.42)	(2.13)
Political Controls	No	No	No	No	Yes	Yes
Demographic Controls	No	No	No	No	No	Yes
Wave FEs	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State	State
Observations	612	612	612	612	589	589

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

approval rating, governor party indicator variables, 2016 Presidential election margin (Trump minus Clinton) interacted with governor party, and the natural logarithm of the state's past-seven-day average of new coronavirus cases and deaths (in case recent outcomes correlate with beliefs and are also responsible for driving governor approval). The key results are robust to adding these controls, with the estimated effect of beliefs on approval becoming modestly smaller, but more tightly estimated.

In Table 4, this same analysis is repeated, except with the binary version of the variable containing individuals' guesses about which states experienced higher death rates from coronavirus. We provide these results to generate general robustness to functional form changes, but we emphasize that this discrete measure throws away variation regarding the relative extent to which respondents believe states have performed differently. While the magnitude of the coefficients changes slightly, the overall conclusions are largely unchanged. A parallel analysis is conducted in Table 5, now using the specification in Equation (3) along with the individual-level data on governor approval that we collected in our mTurk survey. Again, the magnitudes of the coefficients differ slightly from the main specifications – notably, the positive coefficients on the actual deaths variable appear larger. The conclusions, however, are unchanged and the effects remain strongly significant under this alternative approach.

4.3 Effects on Political Approval (Experimental)

We next turn to the results of our survey experiment. Two approaches in our survey experiment yield information about the responsiveness of governor approval to beliefs. The first approach entails simply asking respondents whether they would approve of their governor if, *hypothetically*, they learned that deaths per capita were *X*, for a variety of values of *X*. Specifically,

Table 6: Experimental Effects of Beliefs on Governor Approval (Hypothetical Approach)

Outcome: Governor COVID Handling Approval (Strongly Approve + Approve)	(1) Deaths Hypo-thetical	(2) Deaths Hypo-thetical + Person FEs	(3) Emp. Hypo-thetical	(4) Emp. Hypo-thetical + Person FEs
Deaths per Capita (Log)	-31.11***	-33.87***		
	(3.23)	(2.89)		
Employment Decline			-2.95***	-3.76***
			(0.63)	(0.66)
Person FEs	No	Yes	No	Yes
Clustering	State	State	State	State
Observations	615	615	615	615

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

we elicit conditional approval for each of three different deaths rates, one of which is the true death rate for the state and the other two of which are randomly drawn from the set of other states' death rates. To the extent that approval is varying in X, individuals are admitting that their approval of their governor's handling of the pandemic is indeed responsive to their beliefs about how well the pandemic was handled in their state. Columns (1) and (2) of Table 6 regress an indicator for individual-level approval on the natural logarithm of the hypothetical value of deaths per capita the individual is presented with. Column (1) is a simple univariate regression, whereas column (2) adds person fixed-effects, identifying solely off within-person variation. In both of these specifications, the result that approval is indeed responsive to beliefs emerges – with very strong levels of statistical significance. A 10% increase in hypothetical deaths per-capita is associated with an approximately 3 percentage-point decline in approval of governor coronavirus handling. In columns (3) and (4), we provide additional evidence using hypothetical questions pertaining to a different domain – the percent decline in employment since the start of the pandemic. In this domain, too, approval is responsive to beliefs.

The second approach to identifying effects of beliefs on governor approval is a more standard information-revelation experiment. We randomize respondents into a control group or a treatment group. In the control group, they are asked to guess their state's performance in terms of deaths per capita and then about the extent to which they approve of their governor's handling of the coronavirus pandemic. In the treatment group, they are asked to guess their state's performance – and then told their state's *true* performance – before being asked about approval of their governor's coronavirus handling. The treatment group thus receives a shock to their beliefs, allowing us to measure the effect of a shift in these beliefs on governor approval. We do precisely this in Table 7. Column (1) displays the results of the simplest version of such a specification,

Table 7: Experimental Effects of Beliefs on Governor Approval
(Information Revelation Approach)

Outcome: Governor COVID Handling Approval (Strongly Approve + Approve)	(1) Deaths Info Treatment	(2) Deaths Info Treatment + Controls	(3) Emp. Info Treatment	(4) Emp. Info Treatment + Controls
Deaths per Capita (Log)	-25.05**	-29.78**	-4.44**	-4.60*
* Treatment Group	(11.74)	(13.07)	(2.20)	(2.40)
Believed Deaths per Capita (Log)		-0.99		-0.27
		(1.94)		(0.37)
Benchmark Deaths per Capita (Log)		2.10		0.84**
		(2.83)		(0.42)
State FEs	No	Yes	No	Yes
Clustering	State	State	State	State
Observations	346	346	351	351

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

regressing approval on a treatment group indicator, state deaths per-capita, and the interaction term thereof. The interaction term isolates how the information revealed affects governor approval for treatment group members (relative to control group individuals). This is accordingly the key variable of interest. A 10% shock to believed deaths per capita is estimated to lead to approximately a 2.5 percentage-point decline in approval of governor coronavirus handling. Column (2) adds controls for (prior) believed deaths per-capita and benchmark deaths per-capita, yielding little change in significance or magnitude of the estimates. Columns (3) and (4) repeat the exercise for employment instead of deaths, again finding an analogous effect.

4.4 Effects on Social Distancing Behavior

We next examine whether these erroneous beliefs translate into behavioral differences. During the coronavirus pandemic, efforts to "flatten the curve" of coronavirus cases by encouraging individuals to spend as much time as possible quarantining at home – as opposed to outside – were central to the public health response. Erroneous beliefs about the intensity of the pandemic might lead to distortions in behavior, potentially inflicting costs upon society. To test this, we again run a version of the regression specification described in Section 3.3 – in this case, with SafeGraph's measure of median time (in hours) spent at home per day as the outcome variable.[9]

Table 8 cycles through the same regression specifications used in the preceding section, now with the median time at home outcome. As can be seen, the weak initial correlation between higher log deaths per capita and more social distancing in column (1) is a result of omitted variable bias

[9] We use contemporaneous SafeGraph data. That is, we merge observations from the July wave of our survey with July SafeGraph data on social distancing behavior; we merge observations from the October wave of our survey with October SafeGraph data.

Table 8: Effects of Actual and Believed COVID Deaths on Social Distancing Behavior

Outcome: Median Time (per Day) Spent at Home (Contemporaneous)	(1) Deaths	(2) Deaths + Beliefs	(3) Deaths + Excess Beliefs	(4) Deaths + Bench. + Beliefs	(5) Deaths + Bench. + Beliefs + Controls
Deaths per Capita (Log)	20.26*	-19.85*	116.31***	-19.83*	-11.80
	(11.11)	(10.28)	(19.64)	(10.54)	(8.52)
Believed Relative Deaths (Log)		136.16***		136.71***	123.72***
		(20.09)		(4.36)	(28.98)
Excess Believed Relative Deaths (Log)			136.16***		
			(20.09)		
Benchmark Relative Deaths (Log)				-0.67	4.35
				(32.02)	(28.37)
Political Controls	No	No	No	No	Yes
Wave FEs	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State
Observations	100	100	100	100	96

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 5% level; * Denotes significance at the 10% level

masking a (weakly and scarcely significant) *negative* relationship between actual deaths and social distancing and a very strongly positive relationship between *believed* deaths and social distancing. Column (2) suggests that a 10% increase in believed deaths translates into (roughly) 13.6 extra minutes of time spent at home. Column (3) runs the version of the specification where beliefs are transformed into excess beliefs – the error relative to the truth. Thus the coefficient on actual deaths can be interpreted as the effect of the *true* component of beliefs, while the coefficient on excess beliefs can be interpreted as the effect of the *false* component of beliefs. Whether true or false, beliefs appear to have a nearly identical effect. Column (4) adds the benchmark difficulty as a control variable, and column (5) adds the additional political controls. These additions have minimal effect on the conclusions or even the magnitude of the coefficients.

5 Conclusion

In order to shed light on whether the public rewards (or penalizes) politicians for their performance in office and thereby contribute further to the literature on retrospective voting, we study public perceptions of coronavirus death rates and governor approval ratings during the 2020-21 coronavirus (COVID-19) pandemic. We note that, in order for the public to reward or penalize politicians for their performance, it is necessary for the public to have an accurate understanding of that performance. Errors or biases may ameliorate this ability – and thus undermine the incentive structure for politicians to continue performing well. We ran an incentivized survey on Amazon Mechanical Turk in July (Wave 1) and October (Wave 2) of 2020 asking respondents to provide their best guesses, for 10 randomly-drawn pairs of states, which state had the higher death rate (and by how much). We find that respondents choose the correct state 63.4% of the time. We find little to no evidence of partisan in-group bias, though respondents systematically overestimate death rates in Texas and Florida, states which received substantial media attention despite moderate death rates.

Turning to the question of how these partially-erroneous beliefs translate into governor approval, we find that governor approval is driven by beliefs about death rates, not actual death rates. This remains true if one controls for individuals' perceptions of how well the states should have performed, setting aside factors of leadership/political competence. Using data on social distancing behavior from SafeGraph, we additionally show that these erroneous beliefs about state performance translate into altered social-distancing behavior. We thus conclude that considerations related to imperfect information on the part of the public may generate frictions in the operation of retrospective voting models and ability of voters to reward (penalize) good (bad) performance on the part of politicians.

References

Achen, C. H., and Bartels, L. M. (2004). "Blind retrospection: Electoral responses to drought, flu, and shark attacks." Working Paper.

Alt, J., Bueno de Mesquita, E., and Rose, S. (2011). "Disentangling accountability and competence in elections: evidence from U.S. term limits." *The Journal of Politics,* 73(1): 171-186.

Ashworth, S. (2005). "Reputational dynamics and political careers." *Journal of Law, Economics, and Organization,* 21(2): 441-466.

Barro, R. J. (1973). "The control of politicians: an economic model." *Public Choice*, 14: 19-42.

Bechtel, M. M., and Hainmueller, J. (2011). "How lasting is voter gratitude? An analysis of the short- and long-term electoral returns to beneficial policy." *American Journal of Political Science*, 55(4): 852-868.

Busby, E. C., Druckman, J. N., and Fredendall, A. (2017). "The political relevance of irrelevant events." *The Journal of Politics*, 79(1): 346-350.

Campello, D., and Zucco Jr, C. (2016). "Presidential success and the world economy." *The Journal of Politics*, 78(2): 589-602.

Duch, R. M., and Stevenson, R. T. (2008). *The Economic Vote: How Political and Economic Institutions Condition Election Results*. New York: Cambridge University Press.

Fair, R. C. (1978). "The effect of economic events on votes for president." *The Review of Economics and Statistics*, 60(2): 159-173.

Ferejohn, J. (1986). "Incumbent performance and electoral control." *Public Choice*, 50(1): 5-25.

Fiorina, M. P. (1981). *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.

Gasper, J. T., & Reeves, A. (2011). "Make it rain? Retrospection and the attentive electorate in the context of natural disasters." *American Journal of Political Science,* 55(2): 340-355.

Healy, A., and Lenz, G. S. (2014). "Substituting the end for the whole: Why voters respond primarily to the election-year economy." *American Journal of Political Science,* 58(1): 31-47.

Healy, A., and Malhotra, N. (2010). "Random events, economic losses, and retrospective voting: Implications for democratic competence." *Quarterly Journal of Political Science,* 5(2): 193-208.

Hill, S.J., Huber, G. A., and Lenz, G. S. (2012). “Sources of bias in retrospective decision making: Experimental evidence on voters' limitations in controlling incumbents.” *American Political Science Review* 106(4): 720-741.

Key, V. O. (1966). *The Responsible Electorate*. Cambridge, MA: Harvard University Press.

Kramer, G. H. (1971). “Short-term fluctuations in US voting behavior, 1896-1964.” *The American Political Science Review*, 65(1): 131-143.

Malhotra, N., and Kuo, A. G. (2008). “Attributing blame: The public's response to Hurricane Katrina.” *The Journal of Politics*, 70(1): 120-135.

Malhotra, N., and Margalit, Y. (2014). “Expectation setting and retrospective voting.” *The Journal of Politics*, 76(4): 1000-1016.

Persson, T., and Tabellini, G. (2002). *Political Economics*. Cambridge, MA: MIT Press.

Reeves, A., and Gimpel, J. G. (2012). “Ecologies of unease: Geographic context and national economic evaluations.” *Political Behavior*, 34(3): 507-534.

Schwarz, N., and Clore, G. L. (1983). “Mood, misattribution, and judgments of well-being: informative and directive functions of affective states.” *Journal of Personality and Social Psychology*, 45(3): 513-523.

Stokes, L. C. (2016). “Electoral backlash against climate policy: A natural experiment on retrospective voting and local resistance to public policy.” *American Journal of Political Science*, 60(4): 958-974.

Tufte, E.R. (1978). *Political Control of the Economy*. Princeton, NJ: Princeton University Press.

Wolfers, J. (2009). “Are voters rational? Evidence from gubernatorial elections.” Wharton School Working Paper.

Chapter 3

The Macroeconomic Effects of Flat Taxation: Evidence from a Panel of Transition Economies

1 Introduction

The debate over tax progressivity is as old as taxation itself. Since the passage of the Revenue Act of 1913, the United States has never had a flat tax on incomes, nor have any Western European countries in recent history. However, flat taxes have frequently been proposed and debated by economists, politicians, and political parties in these countries. Advocates have suggested that flat taxes incentivize work, investment, and innovation – and potentially even boosts economic growth over the long-run. Detractors have argued that flat taxes lead to budget deficits, boost inequality, and have no effects on economic growth.

In order to provide a framework for assessing these claims, I set up a simple two-period model of consumption and saving decisions individuals face under flat and progressive tax codes. I show that that decreased progressivity has an effect on investment above and beyond the effect of a downward shift in the tax schedule alone (i.e., decreased average rates). Intuitively, individuals take into account not only their taxes in the current period but also the taxes they will have to pay in the future if, for example, they make a higher income in the subsequent period. Thus it is not only individuals' contemporaneous marginal tax rates that matters for their decision-making; the entirety of the tax schedule plays a role. I relate these findings to a basic Solow model, which – like more complex models of economic growth – implies that an increase in investment should produce a transitionary increase in growth over the short- and medium-run.

In the post-communist countries of Eastern Europe and Central Asia, flat taxes have gone beyond the realm of political and intellectual discussion to become reality. Between 1994 and 2011, twenty post-communist countries introduced such a tax at varying—but typically quite low—rates as a percentage of income. At their peak, nearly all Eastern European and Central Asian countries had a flat tax in effect. Since 2011, on the other hand, some of these countries have repealed their flat taxes and reverted to a progressive system of income taxation. These policy changes represent an ideal natural experiment through which to test the multitude of claims pertaining to flat taxation.

Using quarterly GDP data on this panel of flat-tax adopters and a difference-in-differences identification strategy, I find that the adoption of a flat tax structure has a strongly significant positive effect of 1.36 percentage points on GDP growth. This result is robust, remaining statistically significant under a variety of alternative specifications. Using annual data from accounting firm *Ernst & Young* on the tax schedules of the countries in my panel, I construct a variable measuring the fiscal size of each flat-tax reform – in the vein of Romer and Romer (2010) – and I control separately for this and its lags to factor out direct Keynesian stimulus effects. I control for other (potentially-correlated) aspects of the business environment aside from the tax code, as measured by various components of the *Ease-of-Doing-Business Index*. I restrict my analysis to the subset of flat-tax reforms passed after a close electoral victory for the party advocating the reform. I vary lag length, control for convergence, run annual versions of the regression, and use various sets of fixed-effects. This multitude of specifications slightly changes the magnitude of the effect, but it remains significant in all cases.

The finding of increased GDP growth is also robust to a dynamic difference-in-differences specification, which reveals that the effect is transitionary and persists for approximately one decade, consistent with the implications of the model. Indeed, as implied by the model, it is not only decreased average marginal tax rates which are responsible for the growth effect – decreased progressivity is also responsible for it. If anything, the latter factor is more strongly significant. Also consistent with the model is the fact that the effect on growth appears to be realized through increased investment (and labor supply). Other potential competing channels that could be driving the result – such as repeated stimulatory budget deficits, reductions in structural economic distortions, and reductions in the size of the shadow economy – are found to have no significant effect. Indeed, not every effect of flat taxation suggested by advocates and detractors appears to be realized. I find no evidence of increases in inequality or budget deficits, nor do I find evidence of explosions in innovation (as measured by patenting activity) or FDI.

Finally, while the effects are sizeable, I argue that they are also sensible. The effect sizes I measure have direct implications for the elasticity of investment with respect to tax rates and the elasticity of output with respect to investment. I point out that the latter implied elasticity is well within the range uncovered by the existing literature. In the former case of the elasticity of investment with respect to tax rates, insofar as my implied elasticity is on the high side, I argue this is because the existing literature has focused primarily on the direct channel of the impact of an individual's marginal tax rate on that individual's decisions. One of this paper's key contributions is to shine light on the importance of the extent of progressivity throughout the tax schedule, and accounting for this channel should increase the elasticity of investment with respect to changes in the tax schedule.

The structure of the remainder of this paper is as follows. In Section 2, I review some background information related to flat taxation and its macroeconomic effects – both the relevant economics literature and the political-economic context of the post-communist flat tax reforms.

In Section 3, I set up and solve a simple two-period model of the consumption/saving decision under flat and progressive taxation regimes, and I consider its macroeconomic implications through a standard Solow model of economic growth. In Section 4, I discuss my various data sources and the empirical framework – along with its assumptions – that I use to investigate the macroeconomic effects of flat taxation. In Section 5, I extensively cover my main empirical results, a multitude of robustness checks, and an investigation of the mechanism and its consistency with the model. In Section 6, I discuss these findings and argue their sensibility and consistency with known elasticities from the existing literature. In Section 7, I conclude.

2 Political Economic Context

2.1 Post-Communist Context

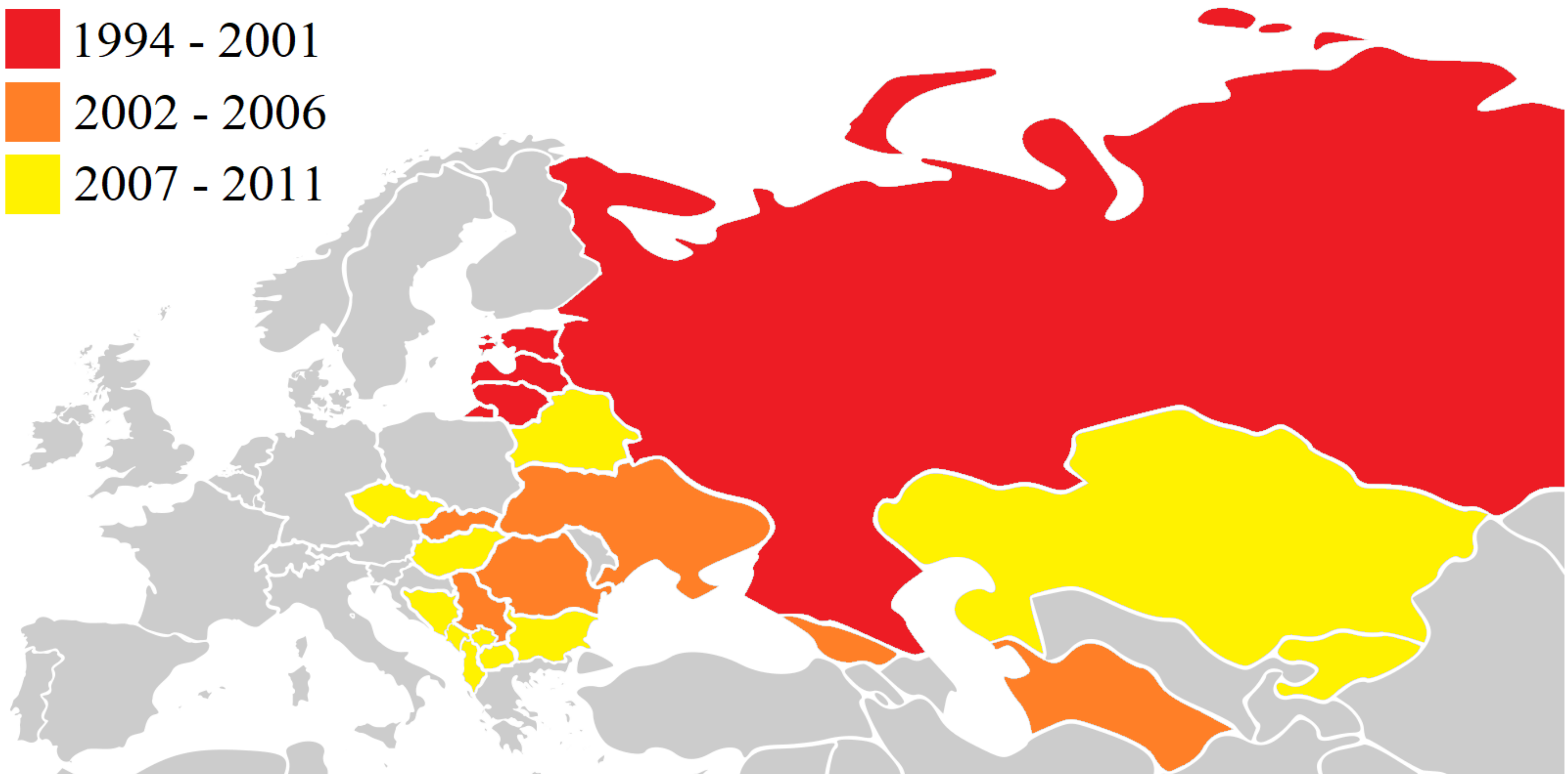
For most of the latter half of the 20th century, the economies of the Eastern European and Central Asian nations were centrally-planned in the Soviet design: fully state-owned and managed by bureaucratic commissions that mandated wages, prices, investment, and output through the auspices of Five-Year Plans. Following the disintegration of the Eastern Bloc and then the USSR itself in 1991, many of these countries thoroughly embraced market-based reforms, which brought the economic systems of the region into greater consonance with the Western European norm. Indeed, many of these countries have gone even further than that in terms of reducing the role of government in economic life.

One key example of this fact is the introduction of flat income taxation. Between 1994 and 2011, twenty countries in Eastern Europe introduced such a tax, at varying—but typically quite low—rates as a percentage of income. Since 2011, on the other hand, five of these countries have repealed their flat taxes and reverted to a progressive system of income taxation. Table 1

Table 1: Flat Tax Timing and Rates

Country	Year of Introduction (Repeal)	Flat-Tax Rate
Estonia	1994	26%
Lithuania	1994	33%
Latvia	1997	25%
Russia	2001	13%
Serbia	2003	12%
Slovakia	2004 (2013)	19%
Ukraine	2004 (2011)	13%
Georgia	2005	20%
Romania	2005	16%
Turkmenistan	2005	10%
Kazakhstan	2007	10%
Macedonia	2007	10%
Montenegro	2007 (2013)	15%
Albania	2008 (2014)	10%
Bulgaria	2008	10%
Czech Republic	2008 (2013)	15%
Belarus	2009	12%
Bosnia & Herzegovina	2009	10%
Kyrgyzstan	2009	10%
Hungary	2011	15%

Figure 1: Flat Tax Reform Map



lists the countries and the associated year of introduction/repeal of a flat income tax, along with the flat rate itself (upon introduction). Figure 1 shows these countries on a map. All of the introductions and repeals were made effective on January 1st of the stated year[1].

To some extent, the blanket term "flat tax" hides the richness in variation amongst the reforms that occurred in these countries. For example, in some cases, only the income tax schedule was modified, with all other taxes in the economy remaining unchanged; in other cases, the entire system of taxation (including corporate taxes, payroll taxes, VAT, etc.) was overhauled. In some cases, the tax reform was a reduction in the general level of taxation; in other cases, it was budget-balanced—a "tilting" of the income tax schedule—or even constituted an increase in the general level of taxation. In some cases, the standard deduction (i.e., the minimum level of income subject to taxation) was increased; in other cases, it was reduced or abolished entirely.

In all cases, the advocates for the reform indicated that they expected it to attract more foreign direct investment and reduce tax evasion. Many were also influenced by the conjecture that such reforms would stimulate economic growth and thus more than pay for themselves in short order. Tax competition was another substantial motive, with Ukraine, for example, choosing its 13% rate to match that of its neighbor Russia in order to avoid being undercut and Belarus, a few years later, choosing a 12% rate in order to undercut them both. Macedonia, in 2007, chose a 10% rate in order to be the lowest in the region. The next year, its immediate neighbors Albania and Bulgaria followed at 10%. Somewhat further east, Turkmenistan introduced a 10% flat tax in 2005. Neighbors Kyrgyzstan and Kazakhstan followed at 10% in the next several years.

Fundamentally, however, the decision to introduce flat taxation in these countries was an

[1] The sole exception is Montenegro, which made its newly-introduced flat tax effective on July 1st, 2007.

ideological one, often implemented after the victory of center-right coalitions. One potential motive of tax/expenditure changes is to use a well-defined fiscal instrument—such as a tax cut on high earners or an increase in government expenditure on infrastructure—in order to offset expected business-cycle fluctuations on the horizon. As argued by Romer and Romer (2010) and much of the subsequent literature on the macroeconomic effects of tax shocks, such endogenous tax changes are unsuitable for studying the effect of tax changes on output. But the very fact that the flat tax was implemented in so many countries with such bewildering rapidity meant that the full extent of its effects was unknown and could not yet have been satisfactorily studied. As such, the proponents of the flat tax had expectations as to what its effects would be, but it would have made a blunt and unlikely fiscal policy instrument. Rather, it was considered an end in itself.

Furthermore, just as the implementation of such flat taxes was typically undertaken by a center-right coalition shortly after an electoral victory, their repeal typically occurred after the victory of center-left coalitions. Again, parties and individuals advocating for repeal did not make arguments based on offsetting expected forthcoming economic fluctuations. Rather, the emphasis was ideological: concerns about fairness and disproportionate burdens on the working-class.

2.2 Literature Review

Arguably the most influential case for flat taxation was made by Hall and Rabushka (1983), who advocated for a broad-based reform to the US tax code. Their basic proposal centered on eliminating exclusions, deductions, or credits to any individual or organization and using the revenue gained to reduce marginal tax rates to a 19% flat tax on wages and business income.

Notably, Hall and Rabushka propose not an income tax but a wage tax. Critics suggested that, because the Hall-Rabushka proposal was a wage tax and the vast majority of income from top earners is capital income, functionally the reform would constitute a transfer from middle- and working-class individuals to wealthier individuals, with those making under $50,000 per year at 1983 prices (equivalent to approximately $125,000 per year in 2017) experiencing an increase in taxation, according to Pechman (1984). Additionally, Auerbach and Kotlikoff (1987), using tax simulation models, find that a shift from an income tax to a wage tax would actually reduce economic efficiency, suggesting that the touted efficiency gains of a flat tax on wages may not be met.

Perhaps as a result of such considerations or perhaps as a result of the anticipated political difficulties of implementing a wage tax, the Hall-Rabushka flat wage tax proposal is not precisely what has been implemented in any of the Eastern European countries. They instead feature more traditional income taxation, except at flat rates, with a standard personal deduction. As such, the above critiques do not directly apply.

There are a number of theoretical benefits of flat income taxation at low rates. First, there is a large body of evidence suggesting that there are indeed behavioral responses to income taxation, with higher rates inducing lower labor supply. The general consensus is that, for the majority of prime-age males in the United States, the earnings elasticity is rather low (in the neighborhood of 0 to 0.1), and that it has declined substantially over time for prime-age females as well (now in the neighborhood of 0.2) as they have become more attached to the labor force (Pencavel 1986, Pencavel 2002, Blau and Kahn 2007). Higher earners, however, from whom the majority of tax revenue in most systems originates, tend to be more sensitive to changes in marginal tax rates, and estimates of their earnings elasticity tend to be in the neighborhood of 0.5

to 0.8 (Saez, Slemrod, and Giertz 2012). This suggests significant benefits vis-à-vis labor supply and economic output in response to a cut in top tax rates, and it also suggests that making up some or most of the lost revenue by increasing the tax burden on low income individuals would not significantly offset those gains .

As a cautionary note, though, Romer and Romer (2014) find an elasticity of 0.2 for the top 1% of earners in the interwar-era U.S., suggesting that—across time and space—high earnings elasticities amongst high-earners are not inevitable. Furthermore, Rebelo and Stokey (1995) explictly investigate the theoretical growth effects of flat-taxation by calibrating an endogenous growth model to the U.S. data. They find that flat-tax reform would have little or no effect on the U.S. growth rate. On the other hand, they do note that factor shares, depreciation rates, the elasticity of intertemporal substitution, and the elasticity of labor supply are crucial parameters to which this result is sensitive, suggesting it is possible that countries with different parameter values may indeed enjoy economic growth effects as a result of a flat-tax reform.

Another potential benefit of a flat tax with low rates—the one cited most often by flat-tax proponents in Eastern Europe—is a reduction in tax evasion, a prediction borne out by many models of tax evasion. Although it may seem strange to think that anyone would increase tax payments as a result of reduced rate, if one models evasion as a costly activity (perhaps consuming time and requiring payments to a team of "creative" accountants), then the reasoning becomes straightforward. Having said that, the effects of a flat tax on evasion are not theoretically unambiguous. Low-income individuals actually experience an increase in rates under most flat-tax proposals; as such, if the majority of evasion comes from low-income individuals, there is no reason to believe that a flat-tax system would ameliorate this issue. If instead most evasion comes from individuals with high incomes, a flat-tax system would indeed be expected to reduce tax evasion, and in their micro-level study of the Russian flat-tax reform, Gorodnichenko, Martinez-Vasquez, and Peter (2009) find exactly such a result.

On the whole, though, there has been relatively little research on the macroeconomic effects of the Eastern European flat tax reforms, a surprising fact given how politically-charged the surrounding debate can be. Right-wing and left-wing commentators alike have made known their strong, even fiery, opinions on the matter[2]. However, this debate has been remarkably unquantified, excepting a spattering of reports and white papers that, for example, attempt to identify the effect of flat taxation on growth rates by comparing the mean growth rate in one given flat-tax-adopting country to the mean growth rates in a set of (often questionably-selected) other countries. As a result of this shortage of well-identified evidence, even a review paper on flat taxation by Keen, Kim, and Varsano (2008) is light on empirical evidence, instead focusing primarily on theoretical implications.

There are a handful of exceptions. Ivanova, Keen, and Klemm (2005) examine micro-level labor-supply responses to the Russian flat tax reform using panel data from the Russian Longitudinal Monitoring Survey, finding little to no evidence of enhanced labor-supply but substantial reductions in evasion (as measured by the gap between household expenditure and reported income). They note, however, that changes in tax enforcement accompanied the flat-tax reform, and it is difficult to decompose how much of the reduced evasion is due to this versus the flat-tax reform.

Mentioned previously, Gorodnichenko, Martinez-Vasquez, and Peter (2009) go a step further. They first supplement the Ivanova, Keen, and Klemm analysis by using the same data source but with a few extra years of data (more than the first two years after the reform),

[2] See, for example, Mitchell (2007), which rails against an IMF report asserting that the flat-tax reforms – given their specific parameters – were unlikely to have an impact on labor-supply or tax compliance, and Bashevska (2014), which brands Macedonia a “workers’ hell” and charges its flat-tax reform with increasing poverty.

confirming the lack of any significant labor-supply/productivity response to the tax reform. However, they are able to isolate the effect of the reform on evasion from the effect of increased enforcement by using a difference-in-difference design which takes advantage of the fact that some income brackets did not experience a marginal rate change as a result of the reform (and hence would only have experienced an enforcement change) while others did, restricting the sample to those near the marginal rate discontinuity for robustness. They find that there was indeed a strong and significant impact of the flat-tax reform on evasion, although 30% smaller than implied by the approach of Ivanova, Keen, and Klemm.

Easterbrook (2008) delves into the tax code data for eight of the countries that implemented flat tax reforms and uses said data to calculate the actual change in average marginal income tax rates. She uses these calculations to calibrate the Prescott (2004) model of labor supply for each of the countries, finding that the model predicts a substantial labor-supply increase in most of the countries—excepting two that actually experienced increased average marginal income tax rates as a result of the reform. However, when she compares the predictions of the model to actual labor supply changes (using data on hours worked from the International Labor Office), true responses appear negligible in most cases.

Adhikari and Alm (2016) use the synthetic control method to study the effect of the flat-tax reforms on the *level* of GDP in the case of 8 specific flat-tax reforms, finding effects in each country that are positive, albeit not strongly significant[3]. Theirs is the closest existing work to this paper. However, the fact that Adhikari and Alm examine only a selection of 8 of the 25 flat-tax reforms/repeals in Eastern Europe means that it is difficult to regard any pooled estimates of the GDP effects as comprehensive. Also, Adhikari and Alm do not use data on the tax code of the countries they study, nor on the fiscal size of any of the reforms, and hence they do not

[3] It is worth noting that these are not all the same countries as examined by Easterbrook (2008).

separate the effects of a tax-cut-induced stimulus from a decrease in the average marginal tax rate (shifting downward of the tax schedule) from that of flat taxation per se (flattening of the tax schedule). Furthermore, the application of a data-driven big-data methodology to a decidedly small empirical macroeconomic dataset means, concretely, that Adhikari and Alm are assuming that a synthetic Kazakhstan (for example) which captures all of the *time-varying unobservable* characteristics of real Kazakhstan can be created from a data-driven weighted-average of the observables of mostly non-transition economies (since nearly all transition economies were flat-tax adopters), a bold assumption. In short, there is plenty of room for a further contribution to this topic.

World Bank (2005) examines the effects of the Slovak flat tax reform on inequality by simulating tax payments under the pre- and post-reform tax systems using household survey data. They curiously find that the Kakwani index of tax progressivity (in terms of pre-tax incomes) actually increased substantially, while the Reynold-Smolensky index (relevant for post-tax incomes) did not increase[4]. As such, while the tax reform made the distribution of tax payments more unequal—in that sense increasing progressivity—the reduction in overall revenue meant that the impact of the tax system on equalizing after-tax incomes was not affected. These results suggest flat tax reforms may have complex and by no means unambiguous distributional effects.

No paper has yet either (i) examined the effect of the flat-tax reforms on longer-term economic growth, (ii) used the full panel flat-tax implementing countries, (iii) studied the outcomes of a flat-tax repeal, or (iv) conducted a systematic (i.e., more than single-country) examination of the mechanism of any such effect – particularly vis-à-vis the relative importance

[4] This is made less counterintuitive when one notes that Slovakia is one of the countries identified by Easterbrook as having had effectively no change in average marginal tax rates as a result of the reform.

of reduced average tax rates and reduced tax progressivity. I hope to fill these gaps in the literature.

3 A Simple Model of Flat Taxation

To provide some intuition for the plausible effects of flat taxation, I present a simple two-period model of investment under varying tax progressivity. Individuals are endowed with wealth ω and live for two periods. They obtain utility from consumption and, consequently, decide how much of their wealth to consume in the first period and how much to invest in a (risk-free) asset with return R which allows them to transfer their wealth to the next period. In other words, any endowment not consumed, c_1, in the first period will be consumed in the second period, such that $c_2 = (1 + R)(\omega - c_1)$. They must pay income tax on their investment income, where the tax rate is given by

$$\tau(y) = \alpha + \beta y$$

for any level of income, y. Notice that, for any $\beta > 0$, the income tax rate is increasing in income – the standard definition of a progressive tax schedule. When $\beta = 0$, the income tax rate reduces to the flat base rate of α (the standard setup in models of investment under taxation). It can be shown that, with a very weak assumption on the form of the utility function, decreased tax progressivity decreases investment.

Proposition: Consider a two-period model of investment wherein individuals are endowed with wealth ω and choose how much of that wealth to consume, c_1, in the first period and how much to invest in an asset with return R for consumption in the next period. Investment income, y, is taxed at the rate $\alpha + \beta y$. In this framework, for any utility function u satisfying $u'(c) > 0$ and

$u''(c) > 0$, the individual's chosen level of investment, $s = \omega - c_1$,

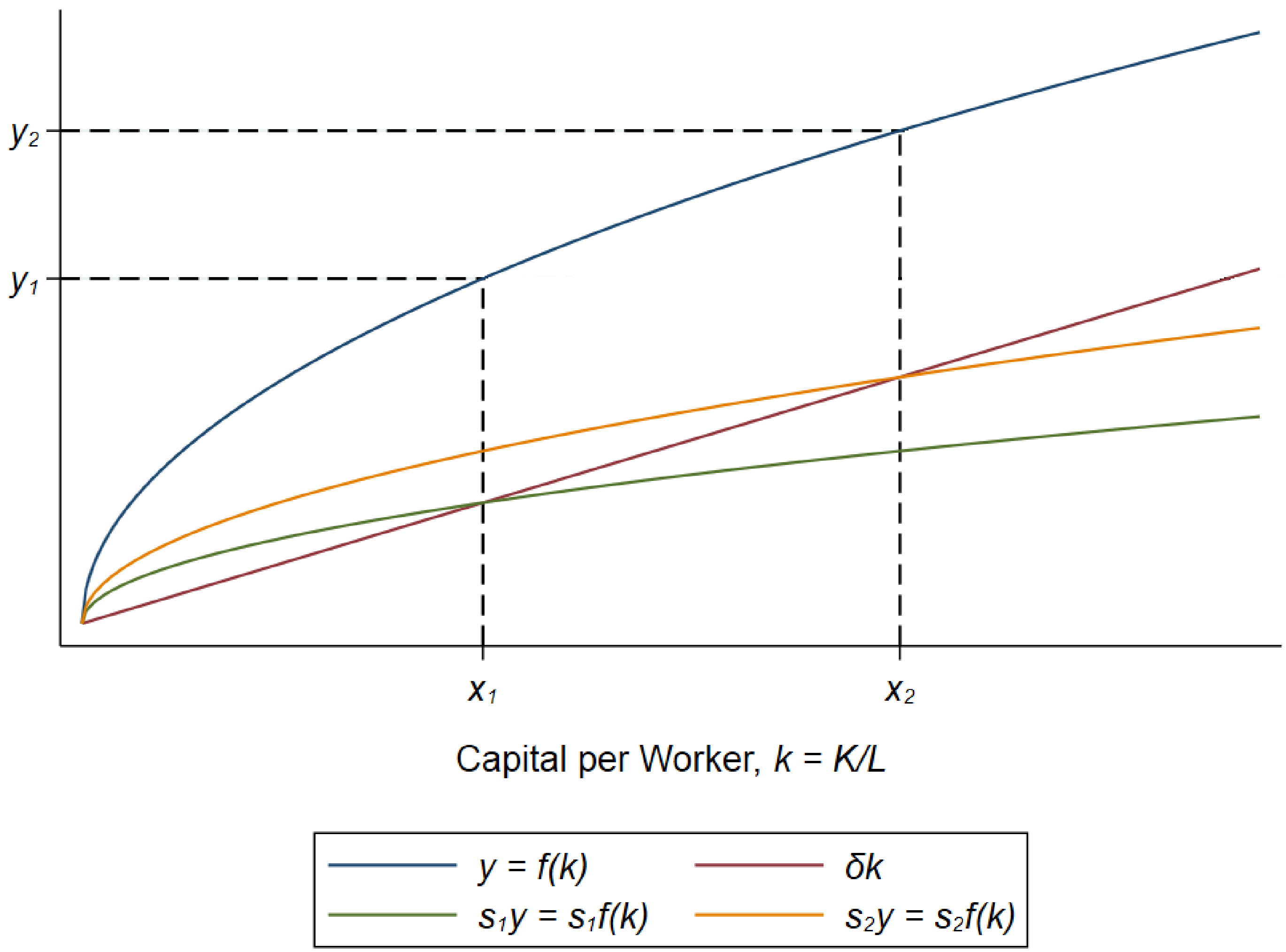
(i) declines, if risk aversion is sufficiently low, as the flat base rate α increases (holding constant progressivity) OR

(ii) declines *unambiguously* as the progressivity of the tax schedule increases (holding constant the average tax rate).

This proposition is proven in Appendix C.1 of the paper. In short and intuitive terms, it says that while shifting the tax schedule vertically may only lead to boosted investment under certain circumstances, changing its slope (even conditional on the average tax rate) *unambiguously* boosts investment. This result is important because it indicates that progressivity of the tax schedule has implications above and beyond simple changes in the base average marginal tax rate. If changes in the income tax schedule have any effect on investment, this result suggests that flat-tax reforms are precisely where we would be most likely to detect them.

But what are the implications for economic growth, if any? To answer this question, I consider a simple Solow (1956) growth model. In the Solow model, net investment in a period is equal to output in that period times the saving rate minus any depreciation of the pre-existing capital stock. That is, $\dot{K}_t = sY_t - \delta K_t$, where Y_t denotes aggregate output in period t, K_t denotes the capital stock, s denotes the saving rate, and δ denotes depreciation. In the steady-state of the model net investment is zero, as $sY_t = \delta K_t$. Increased savings on the part of the populace lead to positive net investment and movement to a higher steady-state level of the capital stock. Since output in the Solow model is produced according to the aggregate production function $Y_t = K_t^{\alpha}(AL_t)^{1-\alpha}$, this translates to a higher level output. This higher level of output, however, is not realized instantaneously. In each period, the new, higher level of investment expands the capital

Figure 2: Solow Model Example



stock slightly, expanding productive capacity and thus output along with it. As the capital stock continues to expand, total depreciation in each period also increases. Eventually, the higher level of depreciation catches up with the higher level of investment and the economy settles at its new steady-state. In this manner, there is a *transitionary*, non-permanent increase in economic growth as a result of the tax reform. Figure 2 plots a graphical example of this dynamic.

It should be noted that this framework is relatively broad in its applicability. That is, while literal financial investment in an asset which pays some return is the most obvious form of investment, a broad class of actions fit under the umbrella of sacrificing utility in the present period in order to obtain an improved payoff in the latter period. To the extent that working harder at a job increases one's future income (through promotions) or to the extent that getting a higher education improves one's future income, the above framework is applicable with only mild adjustment. These, too, are "investments", and while financial investment may be the foremost amongst them, the takeaway is that a broad set of economic activities may be affected by tax progressivity. And to the extent this broad class of additional variables affect economic growth – through endogenous growth models in the case of increased labor or human capital models in the case of improved education – limiting one's viewing lens to standard investment through the Solow model may yield but a lower bound on the importance of tax progressivity.

4 Data and Empirical Framework

4.1 Data

To conduct my analysis, I acquire quarterly GDP data from the Economist Intelligence Unit (EIU), which has collected said data from the Central Statistical Bureaus of the respective countries. For a few of the countries, the data are not available from the EIU, so I obtain it directly from the nation's Central Statistical Bureau[5]. In some cases, the data do not come seasonally-adjusted. As such, I apply the standard x13 seasonal adjustment procedure to these series. For many of the countries in my panel, quarterly GDP data are not available before 1995, so I supplement this with interpolated annual data from the Penn World Table where said quarterly data are missing.

From the World Income Inequality Database (WIID), I obtain data on income share by decile of population and the Gini coefficient at repeated cross-sections. The WIID collates this data from numerous sources, but one source that attempts to measure such indicators in a consistent manner across almost all countries is the World Bank, so I use the World Bank estimates within the WIID. For many countries, the World Bank has annual estimates of Gini and population-by-income stretching back for decades. For other countries, the frequency is less regular. In these cases, I use the estimate from closest year.

Also from the World Bank, I obtain the World Development Indicators (WDI) dataset, which includes data on foreign direct investment, sectoral shares in the economy, population growth, and many other useful indicators. I obtain data on patents from the World Intellectual Property Organization (WIPO). From the IMF Government Financial Statistics (GFS) dataset, I obtain data on tax revenue by source. Again from the Penn World Table, I obtain data on employment and annual average hours worked.

With regard to legislated changes in the tax code, I refer to and digitize information in the annual international tax guides published by Ernst & Young (*Worldwide Personal Income Tax Guide, Worldwide Corporate Income Tax Guide,* and *Worldwide VAT, GST, and Sales Tax Guide*), which detail the tax code in each country for each year since 2006. For the earlier reforms, I obtain this information from the data appendix in Easterbrook (2008), where it was

[5] I do this for Albania, Kyrgyzstan, Turkmenistan, Bosnia and Herzegovina, and Montenegro.

collected from analogous annual tax-code reports that PricewaterhouseCoopers published at the time[6]. In a procedure described in the appendices of this paper, I pair the tax code data with the data on income distributions in order to compute measures of the average marginal tax rate, tax progressivity, and the fiscal size of each flat-tax reform.

4.2 Empirical Framework

The fact that the timing of flat-tax adoption varied substantially across countries suggests a difference-in-differences identification strategy. As discussed extensively above, this series of tax changes consists of policies adopted for ideological reasons, rather than for other factors likely to influence output in the near future. This reduces concerns of systematic correlation between these tax changes and other determinants of output growth. Regardless, it could possibly be the case that individuals are more likely to vote for the center-right parties advocating flat tax reforms at certain points in their local business cycle. Adding lags of output growth controls for the state of the economy and helps to address this possibility of policy endogeneity. As such, the main regression specification is a difference-in-differences approach which controls for lags of output growth[7].

$$\Delta Y_{t,j} = \alpha + \phi F_{t,j} + \sum\nolimits_{i=0}^{M} \beta_i \Delta T_{t-i,j} + \sum\nolimits_{j=1}^{J} \gamma_j + \sum\nolimits_{y=1}^{Y} \psi_y + \theta n_{y,j} + \varepsilon_{t,j} \, ,$$

where $\Delta Y_{t,j}$ denotes quarter-over-quarter real GDP growth (calculated using , ΔT is the measure of tax changes associated with the flat-tax reform (assigned to the quarter of its implementation), $F_{t,j}$ is an indicator variable equal to 0 when country j does not have a flat tax system in effect and

[6] For Albania, which is missing information on its pre-reform tax code, I supplement this with IMF (2005), which provides said information. Similarly, for Macedonia before its reform, I refer to OECD (2003). I am unable to find the tax code for the year immediately before the reform in each of these two countries, so I must assume that they did not change in the couple of years leading up to the reform—an assumption that holds true for the other countries in the panel.

[7] Such controls are the norm in the literature pertaining to the macroeconomic effects of tax changes. See, for example, Romer and Romer (2010).

equal to 1 when it does, γ_j is the fixed-effect for country j, ψ_y is the fixed-effect for year-quarter yq, and $n_{y,j}$ is population growth for country j in year y. Standard errors are clustered at the level of treatment: the country level. X denotes a vector of control variables.

It is worth taking a moment to reflect upon the identification assumptions implicit in this approach. This difference-in-differences specification relies on a parallel trends assumption – that, conditional on a set of controls, if a country which implemented a flat tax in year t had, counterfactually, not done so, then its economic growth would have evolved along the same trajectory as those countries which actually had not implemented a flat tax reform by t. Without controlling for lagged growth, this assumption may have been relatively unpalatable, but doing so helps address the key reverse endogeneity concern described above. Even after this, however, there may remain some additional concerns. Correlated policymaking is one other particular concern: what if the flat taxes tended to be passed simultaneously with other major economic reforms? To deal with these concerns and others, I conduct a number of robustness checks that account for various potential confounds.

Econometricians have recently raised concerns about the reliability of results from static difference-in-difference specifications which are run in fundamentally dynamic settings. Consequently, I also run an analogous dynamic difference-in-differences specification. The second main specification adds lagged output growth to this setup:

$$\Delta Y_{t,j} = \alpha + \phi F_{t,j} + \sum_{i=0}^{N} \beta_i \Delta T_{t-i,j} + \sum_{i=0}^{N} \rho_i \Delta Y_{t,j} + \sum_{j=1}^{J} \gamma_j + \sum_{y=1}^{Y} \psi_y + \theta n_{y,j} + \varepsilon_{t,j}.$$

5 Empirical Results

5.1 Main Results

Table 2 shows the results of the baseline regression and various modifications thereof.

Table 2: Main Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Main	Controls Dropped	*Doing Business*	Close Elections	Fiscal Size Controls	Converg -ence	No Country FEs	Annual	Annual, No FEs
Dependent Variable:	ΔY	ΔY	ΔY	ΔY	ΔY	ΔY	ΔY	ΔY	ΔY
Flat Tax indicator, F	1.362***	2.860***	1.904**	0.900**	1.291**	1.488***	0.765***	2.508***	1.619***
	(0.424)	(0.746)	(0.754)	(0.415)	(0.501)	(0.410)	(0.233)	(0.744)	(0.509)
Observation Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Annual	Annual
Lags of GDP Growth	Yes; 20	No	Yes; 20	Yes; 20	Yes; 20	Yes; 20	Yes; 20	Yes; 20	Yes; 20
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1288	1288	608	432	1288	1288	1288	322	322

Note: * $0.05 < p \leq 0.1$, ** $0.01 < p \leq 0.05$, *** $p \leq 0.01$

Column (1) corresponds directly to the baseline specification, which finds that adopting a flat tax leads to an increase in economic growth of 1.36 percentage points annually. While sizeable, it is worth noting that average annual GDP growth during the 2000s in the countries in my panel was over 5%. Consequently, a boost of 1.36% is large but not unthinkable. Compounded over the course of 10 years, this corresponds to a roughly 3-year boost in growth due to flat tax adoption. Column (2) drops the controls for lagged GDP growth and population growth to show that the result is not being driven by the inclusion of specific controls. The results are only strengthened (albeit non-significantly so).

Column (3) addresses the concern of correlated policymaking. The idea here is that the party introducing the flat-tax reform may also introduce correlated reforms, which could be what are actually responsible for the growth. It is worth noting that, in most all of the countries in my panel, the flat tax has been (or was) in effect for a sufficiently long time such that a different party led the government for at least as many years as the party which introduced the flat tax, somewhat reducing the magnitude of this concern. Still, in order to deal with it, I turn to the Ease-of-Doing Business Index. The Ease-of-Doing-Business Index is compiled annually by the World Bank for a panel of nearly all countries in the world. Its aim is to capture the institutional quality of the environment for starting and operating a business with 10 sub-indices[8]. I find no evidence for this conjecture, as the growth effect actually becomes somewhat *larger* (though the difference is not significant) once these controls are added[9].

Column (4) restricts the sample to those countries wherein the flat tax reform was implemented after the close election victory of the party advocating flat taxation. The idea here

[8] These are starting a business, dealing with construction permits, getting electricity, registering property, getting credit, protecting investors, paying taxes, trading across borders, enforcing contracts, and resolving insolvency.
[9] Note that the sample size is noticeably lower for this column. That is because the Ease-of-Doing-Business Index did not exist prior to 2004, limiting the sample somewhat.

is that a country where 80% of the populace favors a flat tax and 20% is against is plausibly a quite different place than one where support was 50/50 when the issue came up for debate, and differences in outcomes amongst the latter group are more likely to reflect differences in policy adoption rather than idiosyncratic factors correlated with high enthusiasm for center-right policies. In other words, any policy endogeneity still left over after controlling for lagged GDP growth is likely to be further ameliorated or eliminated by such a strategy. Here, too, the effect remains strongly significant and of a similar magnitude.

Column (5) asks whether the effects are truly an enduring consequence of the flat-tax policy itself or simply a short-term Keynesian stimulus effect that has far more to do with deficit spending from any source than the particulars of a flat tax. Using the procedure described in Appendix C.2, I compute a measure of the fiscal size of each tax reform a la Romer and Romer (2010), and I add as controls to the main specification this variable and 20 of its lags. Statistical significance is retained, and the magnitude of the effect barely budges.

Column (6) adds a control for the log of GDP, in acknowledgement of the existence of convergence effects and the fact these countries tended to have less developed economies when they had progressive taxes than when they had flat taxes (since the latter is the more recent system in most of these countries). It should be noted that any bias induced by this factor should bias the effect in the main specification toward zero – downward, not upward. Regardless, the inclusion of this control does not substantially change the situation. In acknowledgement of the finding of Barro (2015) that effect sizes in regressions such as this with convergence terms may actually be biased by the inclusion of country fixed-effects, I run a version without said FEs and, in column (7), again find a significant (albeit non-significantly smaller) effect size.

Column (8) runs the regressions at the annual level using the Penn World Table data. To

Table 3: Varying Lag Lengths

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	4 Lags	8 Lags	12 Lags	16 Lags	20 Lags	24 Lags	28 Lags
Dependent Variable:	ΔY	ΔY	ΔY	ΔY	ΔY	ΔY	ΔY
Flat Tax indicator, F	1.042***	0.809***	1.237***	1.216***	1.362***	1.326**	1.260***
	(0.315)	(0.249)	(0.394)	(0.392)	(0.424)	(0.477)	(0.420)
Observation Frequency	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly	Quarterly
Lags of GDP Growth	Yes; 4	Yes; 8	Yes; 12	Yes; 16	Yes; 20	Yes; 24	Yes; 28
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1560	1492	1424	1356	1288	1224	1156

Note: * $0.05 < p \leq 0.1$, ** $0.01 < p \leq 0.05$, *** $p \leq 0.01$

the extent that the quarterly data released by these countries are less reliable than the annual data or that the seasonal-adjustment process induces any oddities, annual data from a highly standard source should abstract from such concerns. Here too I drop the fixed-effects in column (9). In both cases, again, the result remains statistically-significant.

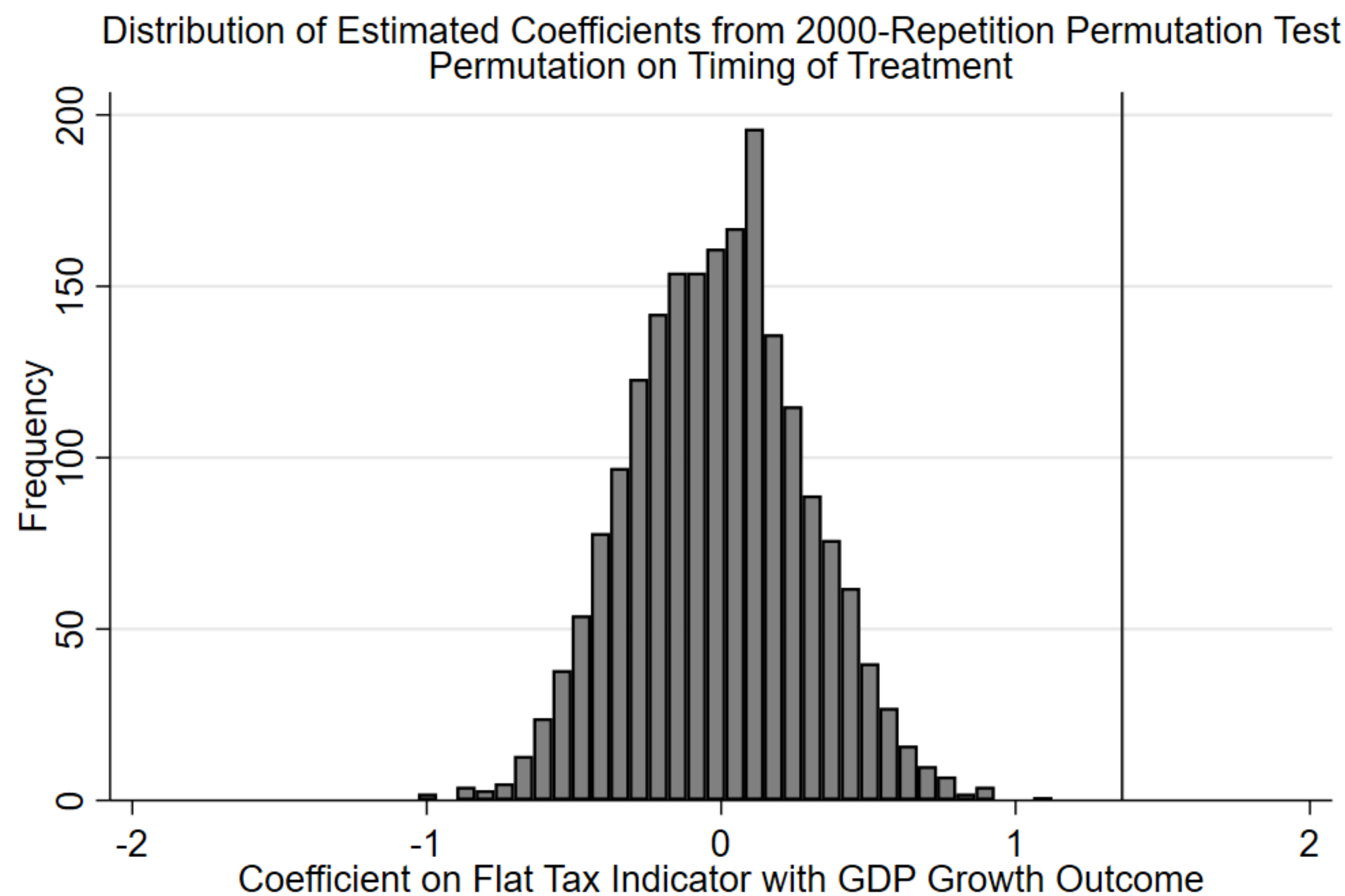
In Table 3, I re-run the baseline specification with a varying lag length – 4 lags, 8 lags, 12 lags, 16 lags, 20 lags, 24 lags, and 28 lags. In each case, the result remains statistically significant and the magnitude is little changed. This demonstrates that it is not a specific lag length driving the results.

The fact that the effect remains strongly statistically significant (and barely reduced in magnitude) even with a full set of country fixed-effects, a full set of year fixed-effects, varying lags of GDP growth, and a restriction to the countries that experienced flat-tax reform implementation after a close election is a strong statement indeed, made even stronger by the aforementioned robustness checks. Regional business-cycle fluctations, within-country fluctuations, national idiosyncracies, the potential presence correlated pro-growth developments, and concerns related to electoral endogeneity are all addressed—the latter in multiple ways.

However, a substantial concern remains. The relative paucity of clusters (i.e., 17) leads to the concern that clustered standard errors may lead be inaccurately narrow and consequently be over-reject the null hypothesis. To answer this concern, one can run a permutation test which generates p-values within-sample. I randomly re-assign the timing of treatment across countries in my sample 2000 times and run regressions on these placebo treatment variables, plotting the resulting coefficients in Figure 3. As can be seen, the implied p-value is less than 0.001 – the result is just as strongly significant as the clustered standard errors would imply.

As pointed out by recent applied econometrics papers such as Borusyak and Jaravel (2017),

Figure 3: Permutation Test



Note: Red vertical line corresponds to coefficient estimated from actual data.
Implied p-value < 0.001.

Figure 4: Dynamic Difference-in-Differences Specification

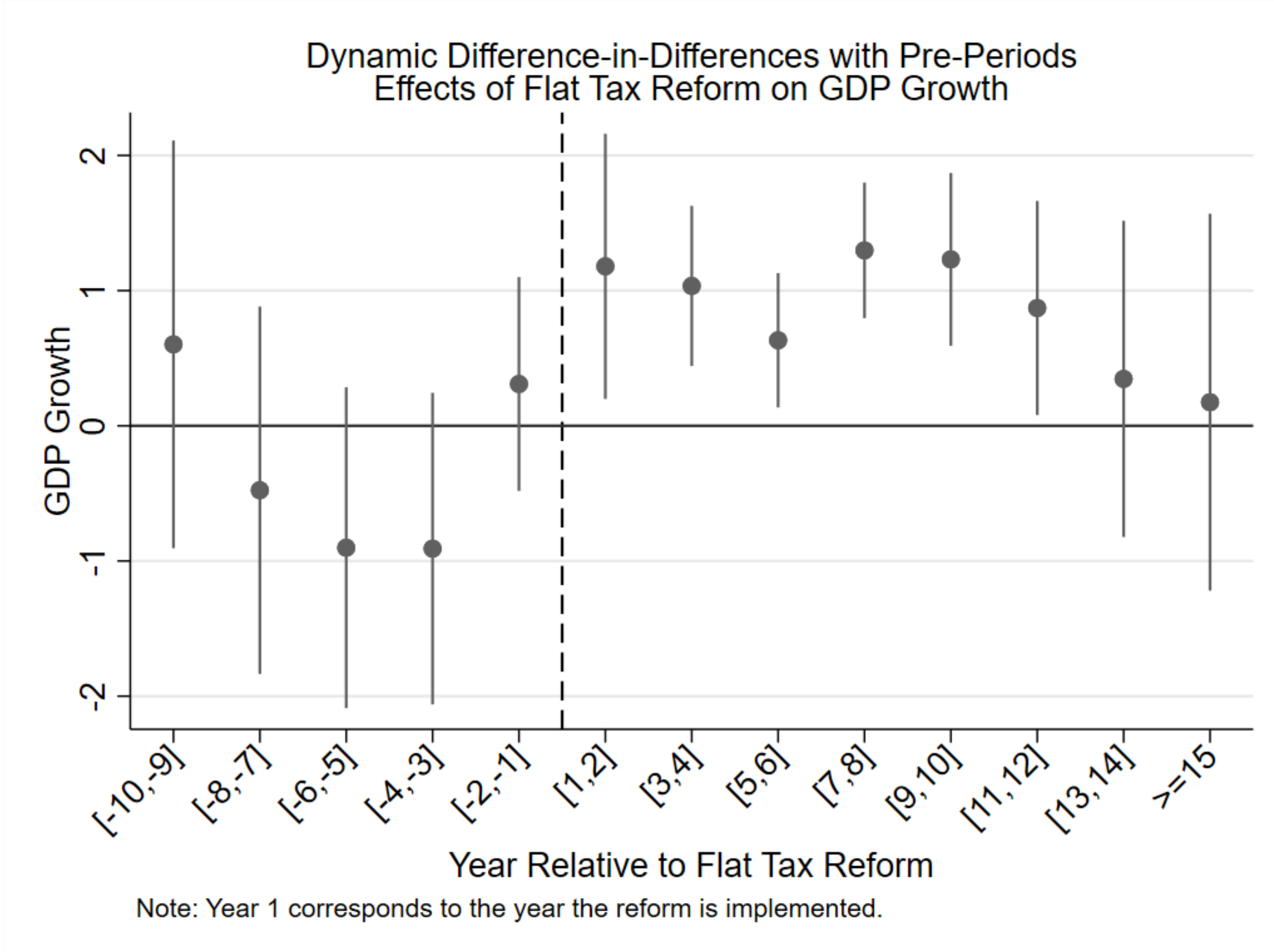


Table 4: Gini Specifications

	(1) Gini, 1	(2) Gini, 2
Dependent Variable:	Gini Growth	Gini Growth
Flat Tax indicator, *F*	0.150	-0.855
	(0.978)	(0.918)
Observation Frequency	Annual	Annual
Lags of GDP Growth	Yes; 5	No
Country FEs	Yes	No
Year FEs	Yes	Yes
Observations	356	390

Note: * $0.05 < p \leq 0.1$, ** $0.01 < p \leq 0.05$, *** $p \leq 0.01$

coefficients from static difference-in-difference regressions may be unreliable if dynamic coefficients exhibit little stability over time. Fortunately, as can be seen in Figure 4, the effect is quite stable for about a decade, and its magnitude during that time period is stable and consistent with the static specification. It is also worth highlighting that the fact the effect is not permanent but rather transitionary is exactly as predicted by the model discussed in Section 3.

Next, I run a year-level specification analogous to the main specification, except with the change in the Gini coefficient as the dependent variable. The results are displayed in Table 4. As seen in column (1), I find no statistically-significant evidence of any effect on inequality as measured by the Gini coefficient. Even if I drop the right-hand-side control variables – which dramatically increased the measured effect on the GDP growth outcome – I still find no evidence of an effect on inequality, as seen in column (2). One potential reason for this puzzling result is the fact that tax compliance was known to be very low in Eastern European and Central Asian countries prior to the flat tax reforms. If the reforms substantially boosted compliance, it would not necessarily be surprising to find a lack of any significant effect on inequality. While this cannot be verified with macro data, the micro-level exploration conducted by Gorodnichenko, Martinez-Vasquez, and Peter (2009) found strong evidence in favor of an effect of this sort in the Russian case, so it is not at all a stretch of the imagination to expect this to happen in these other similar settings.

5.2 Mechanism – Channel of the Effect

Thus flat taxation on income in Eastern Europe appears to have had a positive, robust, and rather large effect on economic growth. A key question remains: through what economic channel(s) was this effect realized? Theory and the assertions of Eastern European flat-tax proponents suggest a few possibilities:

- Domestic Investment: This is the primary channel suggested by the model. Reduction of the tax on high incomes should motivate individuals to re-allocate their income toward saving/investment.
- Labor Supply: Reduction of the income tax on high-income individuals should motivate said individuals (and individuals who might believe they could potentially be high-income in the future) to supply more of their own labor and thus generate more economic output. While such a level effect is theoretically straightforward and well-founded, an effect on economic growth rates through this channel could only be realized through an endogenous growth framework, if perhaps high-income individuals are more likely to work in professions that would contribute to such endogenous growth. Admittedly, this is considerably less straightforward and well-founded than the preceding channel.
- Foreign Direct Investment: Eastern European proponents of flat-taxation suggested it would attract foreign investors to their countries, persuading said individuals to invest, start a business, and move there, bringing themselves along with their financial interests. Such investment could spur economic growth.
- Systematic Budget Deficit: Most of the reforms represented a reduction in the general level of taxation. If government expenditure was not reined in by a commensurate amount, it could be the case that the flat-tax reforms have represented systematic budget deficits, which—viewed as repeated Keynesian stimuli—could result in debt-fueled (and likely unsustainable) economic growth.
- Shadow Economy Size: A notable characteristic of the Eastern European economies is the extremely large size of their underground/shadow sectors, estimates of which tend to

be in the range of 40-50%, depending on the country. If the crucial effect of reducing marginal tax rates on high-income individuals in these countries was to make it cheaper and easier to simply report one's income and pay one's taxes than to hire a team of "creative accountants", then it may in fact be the case that the measured economic growth is actually movement of the shadow sector out of the shadows.

- Removal of Sectoral Distortions: A key feature of the Communist-era Eastern European economies was an inordinately high share of heavy industry in the overall economy. Furthermore, member states of the CMEA—the Communist equivalent of the EEC—were strongly encouraged to specialize in certain areas (e.g., Romania was directed to specialize in agriculture, East Germany in tech, etc.). If an environment of high taxes and subsidies in the aftermath of this period kept sectors distorted in such a way that that the economies were not attaining allocative efficiency, transition to a low, flat-tax regime could induce economic growth.

All of the aforementioned hypotheses have testable implications and can be addressed here. I run difference-in-differences regressions precisely analogous to the main specification, albeit with differing left-hand-side variables. First, with regard to labor supply, the flat-tax reforms could potentially have had an effect on the extensive margin or the intensive margin. As can be seen in columns (1) and (2) of Table 5, there is no significant evidence of an effect on the extensive margin, but there is some significant evidence of an effect on the intensive margin – the flat-tax reform is associated with an increase of 13.45 hours in the growth of annual hours worked.

Column (3) examines the effect on investment growth. Here, too, there is a significant increase – to be specific, an increase of 4.9 percentage points. This finding – and the preceding

Table 5: Channel-of-Effect Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Labor Supply, Extensive	Labor Supply, Intensive	Investment	FDI	Budget Balance	Shadow Economy Share	Patents	Structural Change	Structural Change
Dependent Variable:	Empl. Growth	Δ Hours Worked	log(*I*)	log(FDI)	Budget Balance	Δ Share Shadow	Patent Growth	Lilien Index, *E*	Lilien Index, *Y*
Flat Tax indicator, *F*	0.092	13.454*	0.049**	-0.45	0.55	-0.37	0.061	-0.006	-0.057
	(0.063)	(7.226)	(0.020)	(0.56)	(0.51)	(0.63)	(0.123)	(0.050)	(0.058)
Observation Frequency	Annual	Annual	Annual	Quarterly	Annual	Annual	Annual	Annual	Annual
Lags of GDP Growth	Yes; 3	Yes; 3	Yes; 3	Yes; 12	Yes; 3	Yes; 3	Yes; 3	Yes; 3	Yes; 3
Country FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	356	188	356	815	312	327	318	295	266

Note: * $0.05 < p \leq 0.1$, ** $0.01 < p \leq 0.05$, *** $p \leq 0.01$

one on labor supply – are consistent with the implications of the simple model discussed earlier in this paper. It was precisely these variables through which the flat tax effects on output were mediated. Column (4) examines foreign direct investment (FDI). No significant effect is found here. Although a potential effect on FDI was much-touted by Eastern European flat-tax advocates, such an effect would have to occur through a much more circuitous pathway. For example, US citizens who invest money in an Eastern European country would still need to pay some US taxes on any income resulting from such investments, unless they became a resident of the country in which they are investing – a very hefty and costly decision.

Column (5) turns to the budget balance. No evidence is found of any effect on the budget balance resulting from the flat tax reforms. Budgetary concerns were cited by some Eastern European flat-tax opponents, but these do not appear to have been borne out. The increased labor supply and investment resulting from the reforms likely ameliorated direct revenue decreases, and given the findings of Gorodnichenko, Martinez-Vasquez, and Peter (2009), increased tax compliance may also be partially responsible for the lack of budget deficits. In any case, this finding makes it unlikely that repeated Keynesian stimuli induced by budgetary shortfalls are responsible for the boost in growth.

Column (6) turns to the matter of the shadow economy. I use estimates of shadow economy size from Hassan and Schneider (2016). Schneider has produced the most well-recognized, well-cited estimates of shadow economy size in the literature, and the most recent update of this dataset covers the period 1999 – 2013 for nearly all countries, which fortunately overlaps with the adoption (and repeal) of the vast majority of flat taxes in my panel. These estimates are imperfect, but for countries where more accurate estimates based on the tax gap can be calculated, they match very closely with the Schneider data. In column (6), shadow economy share – the fraction of economic activity esimated to be due to the shadow sector – is used as the outcome variable. I find no statistically-significant effect, which suggests the flat-tax reforms neither significantly shrunk or grew the shadow economy.

Column (7) analyzes the WIPO data on patenting. While the point estimate suggests a 6% increase in the amount of patents filed due to the flat tax reforms, the result is not remotely statistically-significant, and thus it cannot be said that the flat tax reforms are leading to an explosion of innovation, at least as measured by patent data[10].

Columns (8) and (9) examine the sectoral distortion hypothesis, its implication is that the introduction of flat taxation would result in systematically higher structural change. The canonical method for measuring structural chance is to use the Lilien Index, named for Lilien (1982), which measures structural change by summing squared changes in the output (or employment) share of each sector, weighted by that sector's size as a fraction of total output (or employment). Applying this technique to three-sector (agriculture, industry, services) data on employment and GDP shares, respectively, in columns (7) and (8), no statistically-significant effect of flat taxation on either measure of structural change is found. It is worth noting that if the structural change is occurring at a finer level (e.g., workers in the chemical industry becoming workers in the metal industry), it would not be detected by these measures. Regardless, the key distortion of the Communist-era economies was excessive industry and insufficient services, so one might expect movement along that margin, which would indeed be picked up by these measures.

5.3 Mechanism – AMTR, SDMTR, or Both?

[10] An alternative specification which analyzes patents which were granted, not merely patents which were filed similarly yields a non-significant positive coefficient.

The model has another important implication – that the increased economic growth is a result not merely of the fact that the flat-tax reforms reduced tax rates but of the fact that they flattened the whole tax schedule. To this end, I use the *Ernst & Young* data on annual tax schedules and the WIID data on income distributions in a procedure described in Appendix C.3 to compute the change in the average marginal tax rate (a measure of the average level of the tax schedule) and the change in the standard deviation of the marginal tax rate (a measure of the progressivity of the tax schedule) associated with each flat-tax reform. The former measure is quite standard and has a long history in the literature on taxation, dating back to Barro and Sahasakul (1983, 1986). The latter is a natural extension which measures progressivity – a country with a standard deviation of the marginal tax rate equal to zero is a country with a flat tax. The higher the value of this standard deviation, the more the marginal tax rate varies across individuals – i.e., the more progressive the tax schedule[11].

I regress GDP growth on these two measures in order to identify the effect of a downward shift in the tax schedule and the effect of a change in its slope. As in the baseline specification, I include lags of GDP growth, country fixed-effects, and year-quarter fixed-effects. The results are given in column (1) of Table 6. It can be seen that decreasing AMTR and decreasing SDMTR both increase GDP growth. In other words, consistent with the model, the flat-tax reforms induce growth not only through their impact on shifting the tax schedule downward and reducing the AMTR – a subject much-discussed in the existing literature on taxation. They also matter in that they reduce progressivity, itself evidently an important and understudied factor. Columns (2) and (3) repeat this regression for the investment growth and labor supply growth

[11] In theory, a non-zero standard deviation of the marginal tax rate could represent either a progressive tax schedule wherein low-income individuals pay a lower tax rate than high-income individuals *or* regressive tax schedule wherein low-income individuals pay a higher tax rate than high-income individuals. In the case of every single country in my panel, tax rates are monotonically increasing in income. As such, a higher value of the standard deviation of the marginal tax rate can only represent a higher level of progressivity.

Table 6: AMTR & SDMTR Regressions

	(1) GDP Growth	(2) Investment Growth	(3) Hours Worked Growth
Dependent Variable:	Empl. Growth	Δ Hours Worked	log(FDI)
AMTR	-0.154***	-0.443**	-0.809*
	(0.048)	(0.173)	(0.410)
SDMTR	-0.005*	-0.020*	-0.078**
	(0.003)	(0.012)	(0.026)
Observation Frequency	Quarterly	Annual	Annual
Lags of GDP Growth	Yes; 12	Yes; 3	Yes; 3
Country FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Observations	1288	357	189

Note: * $0.05 < p \leq 0.1$, ** $0.01 < p \leq 0.05$, *** $p \leq 0.01$

outcomes, again finding that both factors are significant.

6 Effect Size and Elasticities

Surveying these results, the bulk of the effect of flat taxation on economic growth in Eastern Europe appears to go through a boost in domestic investment. It is thus worthwhile to consider whether the magnitudes of the effect are reasonable and gel with existing macro estimates. First, with regard to the elasticity of output with respect to investment, the regressions suggest that an annual 4.9 percentage-point increase in investment is responsible for a 1.36 percentage-point annual increase in output – i.e., capital elasticity of output of 0.27. Across countries and time, published estimates of capital elasticity range from 0.2 to 0.4[12]. Thus the implied elasticity here is well within this range.

Second, with regard to the elasticity of investment with respect to the marginal tax rate, the average change in average marginal tax rate resulting from the reforms is approximately -5%. The average change in top marginal rate is -15%. The average change in after-tax income of top-bracket individuals is around +10%. Saving is disproportionately undertaken by high-income individuals; such individuals saving 50 cents out of each additional dollar of income they receive is not at all unreasonable and would yield a 5% increase in investment. Indeed, for the U.S. (c. 1990s), Dynan, Skinner, and Zeldes (2004) find that the top quintile of earners have an average MPS of 0.43. The figure should be even higher for the top decile, top 5%, and top 1% – and these categories of individuals make up the vast majority of saving in the economy. However, these estimates pertain to effects on the level of investment, whereas I find evidence of increased growth of investment. It is certainly true that the GDP growth resulting from the increase in the

[12] See, for example, Boskin and Lau (1990), Levy (1990), and Berndt and Hansson (1992).

level of investment leads to additional knock-on effects – further increased after-tax income, which will again lead to higher investment. However, even using estimates of the capital elasticity near the top of the aforementioned range, the total effect I estimate – cumulated over 10 years – of the flat tax reform on growth is 2 to 4 times the size of the total effect one would anticipate from the elasticities previously found in the literature[13].

Having said this, as was pointed out in the context of the model, the existing literature has focused on tax changes that induced a change in the AMTR without a major change in the SDMTR – i.e., tax changes that were not flat-tax reforms. As my model reveals, there are reasons to believe that an additional effect an investment would result from the reduced AMTR. Furthermore, as also hinted at in the model section, the effect through the capital investment channel is only one plausible avenue through which the growth effects of flattening the tax schedule may be realized. Similar logic works for any costly investment which yields a future payoff greater than the initial investment. As seen above, some evidence was found of an effect on the labor-supply channel. Increased schooling leading to a higher-quality, more productive workforce could potentially be another. On the whole, the point is that while increases in the capital stock may explain the largest share of growth effect, there are a multitude of other small channels through which the effect may be operating.

7 Conclusion

Between 1994 and 2011, the spectre of flat-taxation haunted Eastern Europe and Central Asia — and, despite flat-tax repeals in several countries, flat income taxation remains in effect in

[13] After the initial 5% increase in investment yields a 2.5% expansion in output/income, this should again yield a 1.25% increase in investment, which yields a 0.625% expansion in output/income, which yields a 0.2% expansion in income, and so on. The series sums to a cumulative effect of 3.3%. This is one-quarter the cumulative effect over a decade that I find of the flat-tax reforms in my main specification (one-half the effect in the specifications that find the smallest effect sizes).

most of the countries that introduced it during that era. The results of the analysis here demonstrate that flat income taxation had significant, robust, and economically large effects on GDP growth — an annualized 1.3 percentage-point effect in the main specification, which controls for lags of GDP growth, population growth, country fixed-effects, and year fixed-effects. Although the effect varies somewhat depending on the precise specification used, it is always strongly significant, and it is found to endure for approximately one decade. Robustness checks aimed at controlling for the possibility that parties which introduce flat taxes are conceivably more likely to foster a pro-growth environment in other ways, controlling for electoral endogeneity with a restriction of the panel to countries where the flat-tax was introduced (repealed) after a close electoral victory, and combating potential econometric bias all retain strong significance of the aforementioned effect. Finally, deeper analysis of the channels through which the growth rate effect could possibly proceed reveals that domestic investment is the key element. A moderate effect on intensive-margin labor supply is also uncovered. However, no evidence is found for increased FDI, systematic budget deficit, or removal of sectoral distortions as a result of the flat-tax reforms.

Decomposing the flat-tax reforms into a reduction in the average marginal tax rate and a reduction in progressivity (the standard deviation of the marginal tax rate), I find that both of these play a statistically-significant role. In other words, in terms of boosting investment and (transitionary) economic growth, tax progressivity matters above and beyond simply the average level of the tax rate, consistent with the implications of my simple model of consumption and saving under varying tax rates and progressivity.

The extent to which these findings have applicability outside of Eastern Europe is certainly open to discussion. On the one hand, all of these countries have very similar shared histories over the course of the past three-quarters of a century – being devastated by World War II, then transformed into a Communist-led planned economy, and finally beginning a turmoil-ridden transition to market economics in the early 1990s. Because developed Western countries did not suffer from massive amounts of capital depreciation in the 1990s, they may not necessarily have quite as much to gain from boosts to capital accumulation. On the other hand, one could argue that the developing world does indeed have much to gain from such a boost. As such, a potential avenue for fruitful future research could be examining the effects of flat income taxation (and other types of flat taxation) in the developing countries of Latin America and Africa where such taxes have recently begun to be adopted.

References

Adhikari, B. and J. Alm (2016). “Evaluating the economic effects of flat tax reforms using synthetic control methods.” *Southern Economic Journal* 83(2): 437-463.

Auerbach, A.J. and L. Kotlikoff (1987). *Dynamic Fiscal Policy*. Vol. 11. Cambridge: Cambridge University Press.

Bashevska, M. (2014). “Macedonia: Tax Haven – and Workers’ Hell.” *LeftEast*. LeftEast, 18 Sep 2014. Web. 20 May 2017.

Barro, R.J. (2015). “Convergence and modernisation.” *The Economic Journal*, 125(585): 911-942.

Barro. R.J. and C. Sahasakul (1983). “Measuring the average marginal tax rate from the individual income tax.” *Journal of Business* 56: 419-457.

Barro. R.J. and C. Sahasakul (1986) “Average marginal tax rates from social security and the individual income tax.” *Journal of Business* 59: 555-566.

Berndt, E.R. and B. Hansson (1992). “Measuring the Contribution of Public Infrastructure Capital in Sweden.” *Scandinavian Journal of Economics* 94: 151-168.

Blau, F. and L. Kahn (2007). “Changes in the Labor Supply Behavior of Married Women: 1980-2000.” *Journal of Labor Economics*, 25: 393-438.

Boskin, M.J. and L.J. Lau (1990). “Post-War Economic Growth in the Group-of-Five Countries: A New Analysis.” NBER Working Paper No. 3521.

Clementi, F. and M. Gallegati (2005). “Pareto’s law of income distribution: Evidence for Germany, the United Kingdom, and the United States.” *Econophysics of wealth distributions*: 3-14. Milan: Springer Milan.

Dynan, Skinner, and Zeldes (2004). “Do the Rich Save More?” *Journal of Political Economy* 112(2): 397-444.

Easterbrook, K.F. (2008). “Flat Taxes and Labor Supply in Central and Eastern Europe.” Senior Honors Thesis, Stanford University.

Gorodnichenko, Y., J. Martinez-Vazquez, and K.S. Peter (2009). “Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion and Productivity Response in Russia.” *Journal of Political Economy* 117: 504-554.

Hassan, M. and F. Schneider (2016). “Size and Development of the Shadow Economies of 157 Countries Worldwide: Updated and New Measures from 1999 to 2013.” IZA Discussion

Paper No. 10281.

Hall, R.E. and A. Rabushka (1983). *Low tax, simple tax, flat tax*. New York: McGraw-Hill Companies.

IMF (2005). “Albania: Selected Issues and Statistical Appendix.” IMF Country Report No. 05/90.

Ivanova, A., M. Keen, and A. Klemm (2005). “The Russian Flat Tax Reform.” *Economic Policy*, 20: 397-444.

Keen, M., Y. Kim, and R. Varsano (2008). “The ‘Flat Tax(es)’: Principles and Experience.” *International Tax and Public Finance*, 15: 712-751.

Levy, D. (1990). “Aggregate output, capital, and labor in the post-war U.S. economy.” *Economics Letters* 33(1): 41-45.

Lilien, D. M. (1982). “Sectoral Shifts and Cyclical Unemployment.” *Journal of Political Economy*, 90 (4): 777–793.

Mitchell, D.J. (2007). “The IMF’s Remarkably Shoddy Flat Tax Study.” *FoxNews.com*. Fox Entertainment Group, 5 Jan 2007. Web. 20 May 2017.

OECD (2003). "Tax Policy Assessment and Design in Support of Direct Investment–A Study of Countries in South East Europe." *Stability Pact: SEE Compact for Reform, Investment, Integrity, and Growth*. OECD Directorate for Financial, Fiscal, and Enterprise Affairs.

Pechman, J.A. (1984). “Review: ‘Low Tax, Simple Tax, Flat Tax’ by Robert E. Hall; Alvin Rabushka.” *Journal of Political Economy*, 92(2): 340-342.

Pencavel, J. (1986). “Labor Supply of Men: A Survey.” *Handbook of Labor Economics*, Vol. 1.

Pencavel, J. (2002). “A Cohort Analysis of the Association between Work Hours and Wages among Men.” *The Journal of Human Resources*, 37(2): 251-274.

Prescott, E.C. (2004). “Why Do Americans Work So Much More than Europeans?” *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1): 2-13.

Romer, C.D. and D.H. Romer (2010). “The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Schocks.” *American Economic Review*, 100(3): 763-801.

Romer, C.D. and D.H. Romer (2014). “The incentive effects of marginal tax rates: Evidence from the interwar era.” *American Economic Journal: Economic Policy*, 6(3): 242-281.

Saez, E., J. Slemrod, and S.H. Giertz (2012). “The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review.” *Journal of Economic Literature*, 50(1): 3-50.

Sinelnikov-Murylev, S.G. et al. (2003). “Analysis of Certain Results of the Personal Income Taxation Reform.” *Russian Economy in 2002: Trends and Perspectives*, 24.

Rebelo, S. and N.L. Stokey (1995). “Growth Effects of Flat-Tax Rates.” *Journal of Political Economy*, 103(3): 519-550.

Werning, I. (2007). "Optimal fiscal policy with redistribution." *The Quarterly Journal of Economics*, 122(3): 925-967.

World Bank (2005). “The Quest for Equitable Growth in the Slovak Republic.” World Bank Report No. 32433_SK. Washington: World Bank

Chapter 4

Minimum Wages and the Rigid-Wage Channel of Monetary Policy[1]

1 Introduction

Minimum wages and monetary policy are major features of the modern U.S. economy and the economies of many other nations today. A core source for efficacy of monetary policy is the existence of price and wage rigidities, and the minimum wage is an example of an important legislatively-set wage rigidity. It may thus come as a surprise that little attention has been paid to the intersection of these two topics, and no systematic empirical investigation of the role that minimum wages play in mediating monetary policy efficacy has been undertaken as of yet. We aim to fill this gap in the literature.

We also argue that a systematic exploration of the minimum wage's implications for monetary policy is an ideal setting in which to study the importance of the wage-rigidity channel of monetary policy. The empirical literature on nominal wage rigidity has yielded mixed evidence about the extent to which wages are downwardly rigid, as we discuss in our literature review; for example, recent evidence from administrative data suggests that only 7-8% of job stayers experience no year-to-year wage changes during normal times, a lower fraction than the share of total U.S. employment near the minimum wage in the late 1970s and early 1980s, and up to 30% of job stayers experience wage cuts during recessions (Kurmann and McEntarfer 2019). Thus, focusing on the minimum wage – a wage which is known by definition to be nominally-rigid and is binding

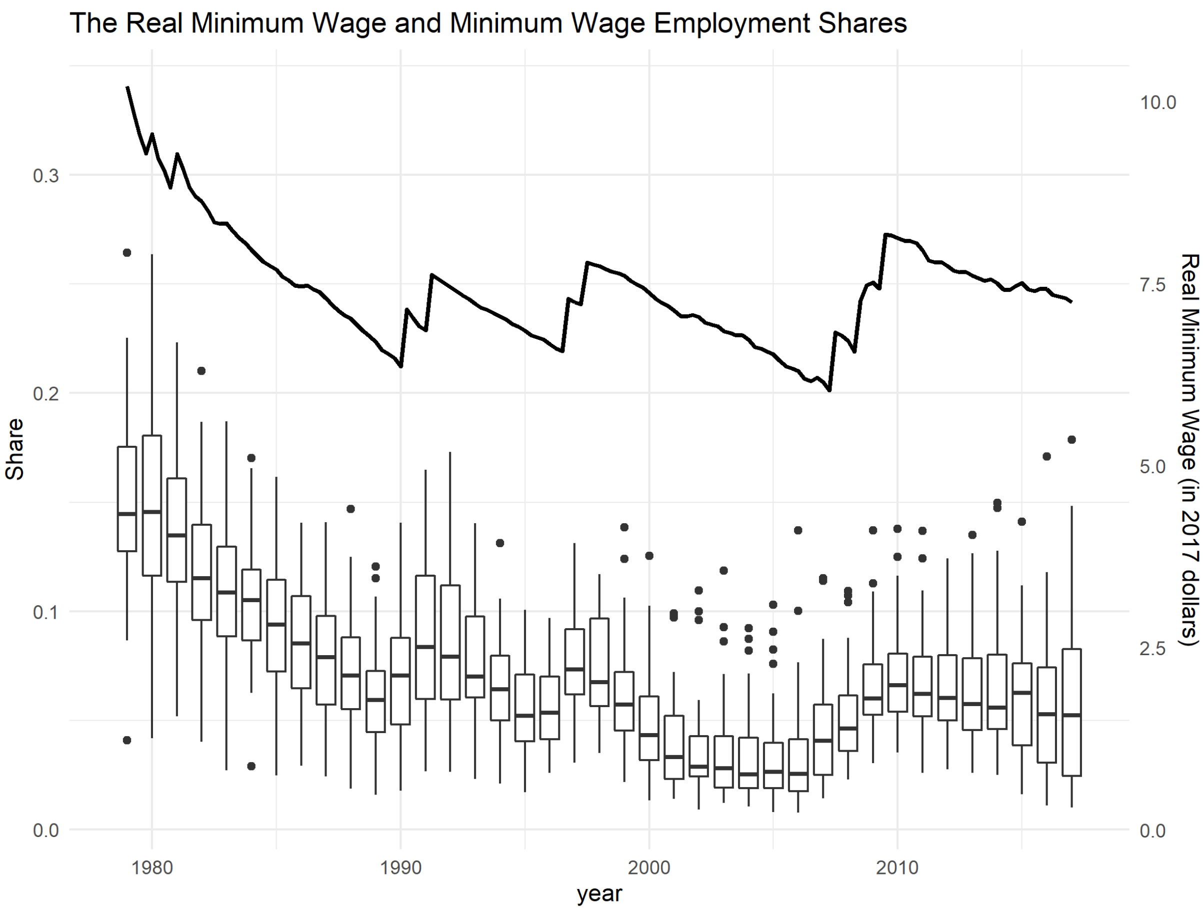
[1] Joint with Robert Minton, Harvard University.

for a non-trivial fraction of the population – allows for a directed examination of the wage-rigidity channel of monetary policy.

We begin by setting up a model of monetary policy in which minimum wages are the only source of non-neutrality. The key assumption is that minimum wages in each state are binding: low skill workers would like to supply more labor than firms demand. Formally, low skill workers take their labor supply as given and, because the minimum wage is exogenous, firm demand determines the quantity of low skill labor in equilibrium. Expansionary monetary policy increases capital rental rates, endogenous wages, and prices, leading to reductions in the real cost of low skill labor for firms. Factor price changes induce both substitution and scale effects: under expansionary policy, firms substitute towards more use of low skill labor and also scale up their operations. Because our mechanism is fundamentally a supply shock, the model predicts larger effects on tradable employment than non-tradable employment; intuitively, more production shifts towards the places where it has become relatively cheaper when that production can be consumed nationally rather than just locally. The tradable sector is also marked by large capital shares, which, combined with the large elasticity of substitution between capital and minimum wage labor, contribute to minimum wage labor demand being relatively more elastic in tradable sectors.

Next, we take the implications of the model to the data. To illustrate the variation we exploit, Figure 1 shows the federal real minimum wage over time and the distribution of state shares of wage and salaried employment near the minimum wage in each year. Defining a worker as near the minimum wage if their hourly wage is within 10% of their state minimum wage, it can be seen that, in 1981, 13.5% of wage and salaried employment was near the minimum wage in the median state (11.4% in the 25th percentile state and 16.1% in the 75th percentile state). 1981 would be the last time minimum wages were raised for nearly a decade, and by 1989, only 5.9% of wage

Figure 1: Real Minimum Wages and Minimum Wage Employment Shares



Note: Solid line represents real minimum wage over time (left vertical axis). Boxplots represent minimum wage employment shares over time (right vertical axis).

and salaried employment was near the minimum wage in the median state. This decline in minimum wage shares of employment tracks the decline in the real federal minimum wage quite well over this period – it was less common for individual states to set their minimum wages then than it is now. By 2005, just before the federal minimum wage increases of 2007-2009, only 2.7% of wage and salaried employment was near the minimum wage in the median state. Our empirical work exploits both time-series and cross-sectional variation in minimum wage shares and drives our conclusion that the declining minimum wage shares may have reduced the efficacy of monetary policy over time.

More specifically, for our baseline regression specifications, we obtain data on monetary policy shocks from Coibion et al. (2017), who expanded the original Romer and Romer (2004) narrative shock series beyond 1996. We obtain monthly data on state-level employment from the Quarterly Census of Employment and Wages (QCEW). And we compute the minimum-wage labor share of total costs by state and year using the Current Population Survey (CPS) and GDP and employee compensation data from the Bureau of Economic Analysis (BEA). Combining these sources, we run a regression specification that has much in common with canonical national-level monetary policy regressions, albeit adapted for a state panel setting by featuring two-way clustering on both the time variable (year and month) and the state variable in order to robustly account for complex autocorrelation structures. That is,

$$\Delta E_{s,t} = \alpha + \sum_{i=0}^{48} \beta_i Shock_{t-i} + \gamma MWShare_{s,t} + \sum_{i=0}^{48} \delta_i Shock_{t-i} \cdot MWShare_{s,t} + \sum_{i=1}^{48} \eta_i \Delta E_{s,t-i} + \varepsilon_{s,t} \ ,$$

where $\Delta E_{s,t}$ represents month-over-month employment growth and $MWShare_{s,t}$ represents the minimum-wage labor share of total costs. From this specification, we find evidence that the short-run employment fluctuation induced by monetary policy is significantly higher in states where the share of the minimum wage workers is higher. The peak effect on employment of a 1 percentage-point Romer and Romer Federal Funds Rate shock on employment growth is appoximately 2.5 percentage points stronger where the minimum wage share is at its 90th-percentile value compared to its 10th-percentile value.

We apply a battery of robustness checks to this finding as well. We run a version analogous to a difference-in-differences specification, adding state and year fixed-effects to the baseline regression. Observing that changes in the share of minimum wage workers can be driven either by plausibly-exogenous factors such as minimum wage changes or by more endogenous factors such as uneven growth of low-wage and high-wage industries, we construct a Bartik-type variable that accounts for the latter effect and add it to our baseline regression. In an alternative approach to isolating the plausibly-exogenous variation, we run an IV specification instrumenting the state minimum wage share with the state minimum wage. Our main result is robust to all these alternative specifications, and the magnitude of the effect is scarcely modified. Comparing our findings to the total effect size of monetary policy during the Volcker era, we find that the rigid-minimum-wage channel accounts for 41 to 69% (depending on the specification) of monetary policy's total effect on employment.

Additionally, we replace the Romer and Romer narrative shocks in our main specification with VAR shocks, and the result is not much changed. We run the baseline specification in the Canadian context – using Canadian data on provincial minimum wages, employment, and monetary shocks – again finding the same significant relationship. We proceed even further with our robustness checks, using QCEW county-level data and the publicly-available 5% samples of the 1980/1990/2000 Censuses to compute the share of minimum wage workers at the county level. Equipped with this data, we add state-by-time fixed-effects in order to pursue a within-state county-level identification strategy. Once again, the result remains statistically significant.

To test the mechanism suggested by the model more clearly, we also run a within-state specification that compares near-minimum-wage employment to higher-wage employment, finding that the employment increases are primarily driven by near-minimum-wage workers, just as the model suggests. Finally, we separately examine the effects on employment in tradable versus non-tradable sectors, finding that the effect is somewhat larger amongst tradable sectors, a result consistent with the implications of our model and inconsistent with the competing explanation that all effects we measure are driven by differences in the MPC across states.

We conclude that minimum wages are an overlooked but important factor in determining the efficacy of monetary policy, confirming the more general hypothesis that wage rigidity is a key contributor to monetary non-neutrality. Indeed, our empirical magnitudes suggest that a sizeable fraction of monetary policy's effectiveness is filtered through precisely this channel. This suggests two policy implications. On the one hand, minimum wages appear to function as an additional dimension of policy space. A higher fraction of minimum wage workers induced by a higher minimum wage unleashes greater effectiveness of monetary stabilization policy. On the other hand, monetary policy may primarily be functioning to erode distortions that were themselves previously put in place by the government. The fact that this channel accounts for a non-trivial amount of monetary policy effectiveness suggests that the Fed – often conceived of as an agency fully independent from the political process – is actually relaxing *legislated* policies and is thus working in close conjunction with the political process.

2 Literature Review

There is an extensive literature, with a diverse methodological history, devoted to studying the effects of monetary policy on economic outcomes. A key bifurcation in the literature on the effects of monetary policy is between those papers which use a vector autoregression (VAR) framework and those which use the narrative approach. While these literatures are both impressive in depth, defying a systematic listing here, key examples of VAR papers include Bernanke and Blinder (1992), Leeper, Sims, and Zha (1996), Bernanke and Mihov (1998), Christiano, Eichenbaum, and Evans (1999), Uhlig (2005), and Bernanke, Boivin, and Eliasz (2005). Key examples of narrative papers include Romer and Romer (1989), Romer and Romer (2004), and Coibion et al. (2017). Both branches of this literature find significant effects of monetary policy on real outcomes – but the effects found in the narrative literature are typically much larger. Ultimately, our regressions will interact monetary policy shocks derived in these literatures with minimum wage shares that we compute in the data.

Our findings contribute directly to the literature on the rigid nominal wage channel of monetary policy. While models generating non-neutrality of monetary policy through nominal wage rigidity are common in the literature, there are no empirical tests of this channel in settings where the extent of wage rigidity is not in question. This is important because the empirical evidence on the extent to which nominal wages are rigid is quite mixed.

Early microdata evidence on downward nominal wage rigidity from the PSID, which contains individual-level wage changes, was relatively unfavorable. Fallick, Villar, and Wascher (2020) describe this evidence: McLaughlin (1994) and Lebow, Stockton, and Wascher (1995) do not find strong evidence of downward nominal rigidity, though Kahn (1997) finds some evidence for hourly wage workers. Later work, e.g. Altonji and Devereux (2000), found that the mixed evidence on downward nominal wage rigidity might be due to measurement error in reported wages.

The evidence on downward nominal wage rigidity in small employer surveys and case studies has also been mixed. While Wilson (1999) and Altonji and Devereux (2000) find supporting evidence, Blinder and Choi (1990) find that five of the nineteen interviewed firms had recently cut wages, despite the booming economy.

Studies using the CPS, e.g. Daly, Hobijn, and Lucking (2012) and Daly and Hobijn (2014, 2015), find an increase between 2007 and 2011 in the fraction of workers in the same job (hereafter, “job stayers”) who report no change in their wage relative to the previous year. These studies are reassuring for the rigid nominal wage hypothesis, since the ORG component of the CPS, like the PSID, contains reported hourly wages, where we may be most likely to find rigidity. One issue with these studies is they focus on the fraction of workers with no wage change rather than focusing on the fraction of workers who receive wage cuts.

More recent studies turn to large surveys of employers that are less likely to suffer from measurement error. An early example is Lebow, Saks, and Wilson (2003). They use microdata from the BLS’s Employment Cost Index (ECI) program, which collects information on compensation for thousands of jobs across thousands of establishments, and find stronger evidence of downward nominal wage rigidity than was typically found in panel data on individual wages: from 1981 to 1999, about 14.5% percent of year-to-year wage and salary changes were negative, and about 18.5% were 0. Fallick, Villar, and Wascher (2020) turn again to this data and find increased downward nominal wage rigidity during and after the Great Recession.

Administrative data point to the importance of analyzing wage cuts and wage freezes separately. Kurmann and McEntarfer (2019) use data collected by the unemployment insurance office in Washington state, which covers over 95% of private-sector employment in the state. They find that, during the Great Recession, the fraction of job stayers who are paid the same wage as a year earlier increases from 7-8% to 16% and then gradually returns to its pre-recession average. The fraction of job stayers who experience wage cuts increases during the recession from 20% to

30%, and the fraction of stayers who experience declines in annual earnings increases to 39%, suggesting some role for composition effects in hours. Jardim, Solon, and Vigdor (2019) find, using the same data, that for every quarter of year-to-year wage changes they in their data, at least 20% of job stayers experienced nominal wage reductions.

Elsby and Solon (2019) survey evidence from employers' payroll records and pay slips in multiple countries, which includes the research from Kurmann and McEntarfer (2019) and Jardim, Solon, and Vigdor (2019) cited above. They find that, except during periods of high inflation or when nominal wage cuts are legally prohibited, an average of 15-25% of job stayers receive nominal wage cuts from one year to the next.

We also contribute to the literature that develops tests of underlying economic mechanisms relying on differential effects of shocks on tradable and non-tradable employment. Intuition and our model suggest larger effects of our channel on tradable employment than non-tradable employment. If, on the other hand, the minimum wage share in a region is correlated with the marginal propensity to consume (MPC) of a region, and this MPC channel is the true underlying mechanism, we might expect that monetary policy leads to larger demand shocks in these regions. Research shows that local demand shocks often lead to larger effects on non-tradable employment than on tradable employment, the opposite of what we would expect from our minimum wage channel. In two papers, Mian, Rao, and Sufi (2013) and Mian and Sufi (2014) develop local demand shocks using changes in housing market wealth and argue these shocks have effects on non-tradable employment but no effects on tradable employment. Chodorow-Reich et al. (2020) similarly argue that local demand shocks generated from changes in stock market wealth affect non-tradable employment but not tradable employment. We think our work further validates the usefulness of analyzing tradable and non-tradable employment when testing underlying economic mechanisms.

Finally, our research is related to an extensive literature on the effects of minimum wage changes on employment. There is limited consensus in this literature on the effect of minimum wage changes on employment (Neumark 2017). It is beyond the scope of this paper to summarize this literature, but we will point to some key research. Well-known papers such as Card and Krueger (1994) and Dube, Lester, and Reich (2010) find no adverse effects of minimum wage increases on employment. More recent evidence includes Cengiz et al. (2019), which also finds no evidence of negative effects on overall employment but does find some effect on employment in tradable sectors. Neumark and Wascher (1992), on the other hand, find that a 10% increase in the minimum wage causes a 1-2% decline in employment among target groups such as teenagers and yound adults. More recent work by Clemens and Wither (2019) finds that a 9% minimum wage increase reduces employment by as much as 9% in a key target group. Reich, Allegretto, and Godoy (2017) analyze Seattle's 2015-16 minimum wage increase from $9.47 to $11 and find it led to no disemployment effects on the food services industry (argued to have a high share of minimum-wage workers). Conversely, Jardim et al. (2019) use administrative data beyond the food-services sector to study the same minimum wage increase, finding the data points to an elasticity of -0.9, and the subsequent increase to $13 point to large disemployment effects, an elasticity of -2.6.

3 Model

Since minimum wage workers make up a relatively small fraction of employment, how large should the effects of our channel of monetary policy be? Further, how much heterogeneity across states should monetary policy generate through our channel? We address these points formally in the model, which provides quantitative estimates of how large an effect monetary policy should generate through the minimum wage channel alone. Throughout this section, the “minimum wage share” in a sector refers to the total payroll of minimum wage workers in that sector divided by total cost in that sector.[2]

The share of minimum wage workers is correlated with numerous other variables that may lead to differential effects of monetary policy across regions and time, so the model also provides an opportunity for us to generate the unique implications of our channel relative to competitor explanations. The model focuses on one confound in particular: the share of minimum wage workers may be high in regions where a higher share of households is credit constrained. In this case, any effects we attribute to monetary policy relaxing the minimum wage may be due instead to monetary policy alleviating or exacerbating credit constraints. More generally, higher minimum wage share regions may be regions where there is a higher marginal propensity to consume (MPC). We address this concern by analyzing tradable and non-tradable sectors in the model. As discussed in our literature review, shocks going through the MPC channel should lead to larger effects on employment in non-tradable sectors than on employment in tradable sectors. The minimum wage channel should lead, in contrast, to larger effects in tradable employment, a result we will confirm in the model and in our empirical analysis.

3.1 Households

The representative agent in each state $s = 1, \ldots, S$ purchases tradables and non-tradables to produce a commodity, which can be invested or consumed. So, though there are two types of goods available in each state, there is only one type of capital, produced out of both non-tradables and

[2] Full details on how these are computed from CPS and BEA data, see data section.

tradables, in each state. Agents cost minimize over tradable and non-tradable inputs when producing the commodity, yielding the expenditure function

$$E_s\left(P_t^T, P_{s,t}^{NT}, Y_{s,t}\right) = \min_{Y_{s,t}^T, Y_{s,t}^{NT}} P_t^T Y_{s,t}^T + P_{s,t}^{NT} Y_{s,t}^{NT} \quad s.t.\ F_s\left(Y_{s,t}^T, Y_{s,t}^{NT}\right) = Y_{s,t}$$

Where F_s is assumed to be constant returns to scale for each s. We will call $Y_{s,t}$ "demand" in state s at time t. The commodity price is an ideal price index given by

$$P_{s,t} = E_s\left(P_t^T, P_{s,t}^{NT}, 1\right).$$

We assume the steady state elasticity of substitution between tradables and non-tradables, denoted by $\sigma_{NT,T}$, does not vary by state, an assumption driven by the absence of measurements of the parameter at this level of granularity.

We use lower-case variables to refer to the natural log of their upper-case variants, e.g. $k_{s,t} = \ln K_{s,t}$. A hat denotes a variable's deviation from its steady state value. The existence of steady state is shown later. Finally, a boldface variable refers to a vector or matrix.

The cost minimization conditions can be log-linearized as

$$\hat{\boldsymbol{y}}_t^{NT} - \hat{\boldsymbol{y}}_t^T \approx -\sigma_{NT,T}(\hat{\boldsymbol{p}}_t^{NT} - \mathbf{1}\,\hat{p}_t^T)$$

$$\hat{\boldsymbol{y}}_t \approx \boldsymbol{\eta}_{NT}\,\hat{\boldsymbol{y}}_t^{NT} + \boldsymbol{\eta}_T\,\hat{\boldsymbol{y}}_t^T.$$

Where endogenous boldface variables are $S \times 1$ vectors. The steady state cost shares $\boldsymbol{\eta}_{NT}$ and $\boldsymbol{\eta}_T$ are diagonal $S \times S$ matrices giving the share of a state's GDP in the state's non-tradable and tradable sectors, respectively.

The price index defined as above can be log-linearized as

$$\hat{\boldsymbol{p}}_t \approx \boldsymbol{\eta}_{NT}\,\hat{\boldsymbol{p}}_t^{NT} + \boldsymbol{\eta}_T\,\mathbf{1}\,\hat{p}_t^T.$$

The agent's dynamic problem can now be defined at the commodity level, abstracting from tradables and non-tradables.

The representative agent supplies capital K and two types of labor to the production side of the economy. Labor type L is subject to a binding wage floor $\overline{W}$, and labor type H is paid an endogenous wage W. We will refer to labor type L as "low skill" or "minimum wage" labor and to labor type H as "high skill" or "endogenous wage" labor. The model contains no uncertainty. The results are not meaningfully changed if we permit two representative agents in each state, (1) low skill agents who consume hand-to-mouth and face a binding minimum wage and (2) high-skill agents who perform all investment in the state and whose wage is endogenous. The budget constraint is

$$P_{s,t}Y_{s,t} = P_{s,t}\left(C_{s,t} + I_{s,t}\right) = \overline{W}_{s,t}L_{s,t} + W_{s,t}H_{s,t} + R_{s,t}K_{s,t},$$

Where the first equality links the budget constraint to the previously described cost minimization component of the consumer problem. The law of motion for capital is

$$\dot{K}_{s,t} = I_{s,t} - \delta K_{s,t},$$

Where a dot refers to the time derivative of a variable, and δ is the depreciation rate of capital. The utility function is separable and does not vary by region or time:

$$U\left(C_{s,t}\right) - V\left(H_{s,t}\right) - V_L\left(L_{s,t}\right).$$

Our key assumption is that the wage floor is binding in each state. Thus, the representative consumer in each state would like to choose a higher value of $L_{s,t}$ than the state can support. In the maximization problem, $L_{s,t}$ will therefore be taken as exogenous. This is a simple application to the disequilibrium framework of Barro and Grossman (1971). We use the budget constraint to unconstrain the maximization problem, which we write as

$$\max_{\{\dot{K}_{s,t}\}_{t=0}^{\infty},\{H_{s,t}\}_{t=0}^{\infty}} \int_0^\infty e^{-\rho t}\left(U\left(\frac{\overline{W}_{s,t}}{P_{s,t}}L_{s,t} + \frac{W_{s,t}}{P_{s,t}}H_{s,t} + \left(\frac{R_{s,t}}{P_{s,t}} - \delta\right)K_{s,t} - \dot{K}_{s,t}\right) - V\left(H_{s,t}\right)\right)dt$$

Where the initial capital stock is given in each state. Though the utility function does not vary by state, the inverse elasticity of intertemporal substitution, denoted γ, and the Frisch-elasticity of labor supply, denoted ϵ, may still vary by state, since they depend on the level of consumption and skilled labor supply, respectively. We assume they do not vary by state, which could easily be micro-founded using CRRA forms for consumption utility and skilled labor disutility. Optimization yields an intratemporal and intertemporal Euler equation in each state, which we can log-linearize, respectively, as

$$-\gamma\, \hat{\boldsymbol{c}}_t + \widehat{\boldsymbol{w}}_t - \widehat{\boldsymbol{p}}_t \approx \frac{1}{\epsilon} \widehat{\boldsymbol{h}}_t$$

$$\dot{\boldsymbol{c}}_t \approx \frac{\rho + \delta}{\gamma} (\hat{\boldsymbol{r}}_t - \widehat{\boldsymbol{p}}_t)$$

All of the boldface objects are endogenous $S \times 1$ vectors.

3.2 Non-tradable Sector

There is a firm in each state that produces non-tradables for use in that state. The sector first solves the cost minimization problem

$$E_s^{NT}\left(R_{s,t}, W_{s,t}, \overline{W}_{s,t}, Y_{s,t}^{NT}\right)$$

$$= \min_{K_{s,t}^{NT}, H_{s,t}^{NT}, L_{s,t}^{NT}} R_{s,t} K_{s,t}^{NT} + W_{s,t} H_{s,t}^{NT} + \overline{W}_{s,t} L_{s,t}^{NT} \quad s.t.\ F_s^{NT}\left(K_{s,t}^{NT}, H_{s,t}^{NT}, L_{s,t}^{NT}\right) = Y_{s,t}^{NT}.$$

Where F_s^{NT} is assumed to be constant returns to scale for each s. Profit maximization then yields non-tradable prices in each state,

$$P_{s,t}^{NT} = E_s^{NT}\left(R_{s,t}, W_{s,t}, \overline{W}_{s,t}, 1\right).$$

We assume the elasticities of substitution in steady state, denoted by σ_{HL}^{NT}, σ_{LK}^{NT}, and σ_{HK}^{NT}, do not vary by state. The non-tradable firms' cost minimization conditions can be log-linearized and stacked as

$$\hat{\boldsymbol{l}}_t^{NT} \approx \sigma_{HL}^{NT} \boldsymbol{\eta}_H^{NT} \left(\widehat{\boldsymbol{w}}_t - \widehat{\overline{\boldsymbol{w}}}_t\right) + \sigma_{LK}^{NT} \boldsymbol{\eta}_K^{NT} \left(\hat{\boldsymbol{r}}_t - \widehat{\overline{\boldsymbol{w}}}_t\right) + \hat{\boldsymbol{y}}_t^{NT}$$

$$\widehat{\boldsymbol{h}}_t^{NT} \approx \sigma_{HL}^{NT}\boldsymbol{\eta}_L^{NT}\left(\widehat{\overline{\boldsymbol{w}}}_t - \widehat{\boldsymbol{w}}_t\right) + \sigma_{HK}^{NT}\boldsymbol{\eta}_K^{NT}(\widehat{\boldsymbol{r}}_t - \widehat{\boldsymbol{w}}_t) + \widehat{\boldsymbol{y}}_t^{NT}$$

$$\widehat{\boldsymbol{k}}_t^{NT} \approx \sigma_{LK}^{NT}\boldsymbol{\eta}_L^{NT}\left(\widehat{\overline{\boldsymbol{w}}}_t - \widehat{\boldsymbol{r}}_t\right) + \sigma_{HK}^{NT}\boldsymbol{\eta}_H^{NT}(\widehat{\boldsymbol{w}}_t - \widehat{\boldsymbol{r}}_t) + \widehat{\boldsymbol{y}}_t^{NT}$$

The endogenous variables and minimum wage variable are again $S \times 1$ vectors, and the $\boldsymbol{\eta}_i^{NT}$ are diagonal $S \times S$ matrices with entries given by the cost share of input i in non-tradable production in the relevant state.

Note that these equations are the standard Slutsky equations for the firm. The first two terms denote substitution effects, and the final term denotes the scale effect. If production were Leontief, then each σ would be 0, and we would be left only with the scale effect.

3.3 Tradable Sector

There is one, national tradable firm that produces in all states. We find this setup more realistic than permitting distinct tradable sectors in each state that produce a homogeneous output, since we will be able to allow parsimoniously for differences in state-level tradable output. The sector operates by producing a commodity in each state and then combining these commodities to produce final tradable output. In the first stage, it cost minimizes over production in each state:

$$E_s^T\left(R_{s,t}, W_{s,t}, \overline{W}_{s,t}, Y_{s,t}^T\right)$$
$$= \min_{K_{s,t}^T, H_{s,t}^T, L_{s,t}^T} R_{s,t}K_{s,t}^T + W_{s,t}H_{s,t}^T + \overline{W}_{s,t}L_{s,t}^T \quad s.t.\ F_s^T\left(K_{s,t}^T, H_{s,t}^T, L_{s,t}^T\right) = Y_{s,t}^T.$$

The production function F_s^T exhibits constant returns to scale in each s. This generates an ideal price index for the price of the state commodities required in production of the national tradable:

$$P_{s,t}^T = E_s^T\left(R_{s,t}, W_{s,t}, \overline{W}_{s,t}, 1\right).$$

The firm then minimizes national-level costs:

$$E^T\left(P_{1,t}^T, \dots, P_{S,t}^T, Y_t^T\right) = \min_{\{Y_{s,t}^T\}_{s=1}^S} \sum_{s=1}^{S} P_{s,t}^T Y_{s,t}^T \qquad s.t.\ F^T\left(Y_{1,t}^T, \dots, Y_{S,t}^T\right) = Y_t^T$$

Finally, profit maximization yields the national tradable price,

$$P_t^T = E^T\left(P_{1,t}^T, \dots, P_{S,t}^T, 1\right).$$

The cost minimization conditions can be combined and stacked. This is a more complicated procedure than in the non-tradable sector, but we will give intuition after defining the relevant objects. Assume the elasticities of substitution between the state commodities in producing the national tradable are all equal and given by σ_S. This could be micro-founded by assuming F^T has a CES form with a single elasticity of substitution, σ_S. Note that, had we modeled the tradable sector with distinct tradable sectors in each state that produce homogeneous output, we would implicitly be letting $\sigma_S \to \infty$, the case of perfect substitutes. Further, define η_s^T as the cost share of the state s commodity in producing the national tradable, measurable by tradable GDP in that state divided by tradable GDP in the U.S. Denote the $S \times S$ diagonal matrix of these shares by $\boldsymbol{\eta}^T$. The diagonal $S \times S$ matrices of cost shares $\boldsymbol{\eta}_i^T$ are analogous to those defined in the non-tradable sector: their entries are given by the cost share of input i in tradable production in the relevant state. We define the $diag$ operator, which retrieves the diagonal entries of a matrix as a column vector, and the $*$ operator, which performs elementwise multiplication between two objects of the same dimension.

First, we define the scale effect

$$\boldsymbol{s}_t = \boldsymbol{S} * \left(\boldsymbol{\eta}_L^{T^*} \widehat{\overline{\boldsymbol{w}}}_t + \boldsymbol{\eta}_H^{T^*} \widehat{\boldsymbol{w}}_t + \boldsymbol{\eta}_K^{T^*} \hat{\boldsymbol{r}}_t\right) + \mathbf{1}\, \hat{y}_t^T,$$

where

$$\boldsymbol{S} = \sigma_s \begin{pmatrix} -\frac{1-\eta_1^T}{\eta_1^T} & 1 & \cdots & 1 \\ 1 & -\frac{1-\eta_2^T}{\eta_2^T} & \cdots & 1 \\ \vdots & 1 & \ddots & \vdots \\ 1 & 1 & \cdots & -\frac{1-\eta_S^T}{\eta_S^T} \end{pmatrix}, \qquad \boldsymbol{\eta}_i^{T^*} = \begin{pmatrix} diag(\boldsymbol{\eta}^T \boldsymbol{\eta}_i^T)' \\ diag(\boldsymbol{\eta}^T \boldsymbol{\eta}_i^T)' \\ \vdots \\ diag(\boldsymbol{\eta}^T \boldsymbol{\eta}_i^T)' \end{pmatrix}$$

And we note that elementwise multiplication of $\boldsymbol{S}$ with the $\boldsymbol{\eta}^*$ matrices must occur before the $\boldsymbol{\eta}^*$ matrices multiply the factor prices. Then it follows that

$$\hat{\boldsymbol{l}}_t^T \approx \sigma_{HL}^T \boldsymbol{\eta}_H^T \left(\widehat{\boldsymbol{w}}_t - \widehat{\overline{\boldsymbol{w}}}_t\right) + \sigma_{LK}^T \boldsymbol{\eta}_K^T \left(\hat{\boldsymbol{r}}_t - \widehat{\overline{\boldsymbol{w}}}_t\right) + \boldsymbol{s}_t$$

$$\widehat{\boldsymbol{h}}_t^T \approx \sigma_{HL}^T \boldsymbol{\eta}_L^T \left(\widehat{\overline{\boldsymbol{w}}}_t - \widehat{\boldsymbol{w}}_t\right) + \sigma_{HK}^T \boldsymbol{\eta}_K^T (\hat{\boldsymbol{r}}_t - \widehat{\boldsymbol{w}}_t) + \boldsymbol{s}_t$$

$$\widehat{\boldsymbol{k}}_t^T \approx \sigma_{LK}^T \boldsymbol{\eta}_L^T \left(\widehat{\overline{\boldsymbol{w}}}_t - \hat{\boldsymbol{r}}_t\right) + \sigma_{HK}^T \boldsymbol{\eta}_H^T (\widehat{\boldsymbol{w}}_t - \hat{\boldsymbol{r}}_t) + \boldsymbol{s}_t.$$

As in the non-tradable sector, these equations all represent Slutsky equations for the tradable sector, except now there are two substitution effects. On the one hand, when a particular input in a state becomes more expensive, the first substitution effect, and the same one we saw in the non-tradable sector, drives substitution to the cheaper inputs in that state. The second substitution effect, contained in $\boldsymbol{s}$, drives the tradable firm to substitute away from the commodity in the state that has seen the factor price increase. Thus, for tradable production in the state where the factor price has increased, the substitution effect in $\boldsymbol{s}$ is a scale effect, whereas for tradable production construed nationally, it is just another substitution effect away from the more expensive input, which in $\boldsymbol{s}$ is the state-level commodity. Finally, $\boldsymbol{s}$ also contains the standard output scale effect we also saw in the Slutsky equation for the non-tradable sector.

Note that if production at the state level and national level were both Leontief, all elasticities of substitution would be 0, and we would have $\hat{\boldsymbol{l}}_t^T \approx \boldsymbol{1}\, \hat{y}_t^T$, $\widehat{\boldsymbol{h}}_t^T \approx \boldsymbol{1}\, \hat{y}_t^T$, and $\widehat{\boldsymbol{k}}_t^T \approx \boldsymbol{1}\, \hat{y}_t^T$. Note that this scale effect, unlike the scale effect in the non-tradable sector, is the same for all states. In this case, the model predicts no heterogeneity in log employment effects by state in the tradable sector. If production at the state level were Leontief, but we kept the general form for national production, we would still have $\hat{\boldsymbol{l}}_t^T \approx \boldsymbol{s_t}$ and be unable to simplify $\boldsymbol{s}_t$ to just $\mathbf{1}\, \hat{y}_t^T$ (and similarly for the other inputs).

For the shocks we consider in our model, increasing σ_s from arbitrarily close to 0 will have only a modest effect. This is because our national shocks will change all states' real minimum wages at the same time. Recall that σ_s provides a measurement of how the national firm's relative use of state commodities varies with the relative price of those commodities. When the minimum wage increases by the same amount in all states, relative commodity price changes are governed by differential cost shares of minimum wage workers across states. If minimum wages do not increase in all states simultaneously, there will be a force for relative commodity prices to change more dramatically, particularly when cost shares are not too close to 0 and minimum wage changes are large.

3.4 Equilibrium and Steady State

The national money supply M_t, which is the quantity of money times its velocity, is the numeraire. It satisfies

$$M_t = \sum_s P_{s,t}^{NT} Y_{s,t}^{NT} + P_t^T Y_{s,t}^T = \sum_s E_s\left(P_{s,t}^{NT}, P_t^T, Y_{s,t}\right) = \sum_s P_{s,t} Y_{s,t} \equiv GDP_t$$

Where GDP_t is nominal gross domestic product.

The rest of the equilibrium is standard. Goods markets and labor markets clear. It is worth mentioning that labor market clearing in the low skill labor market means that the low skill labor demand taken as given by households in each state equals the combined low skill labor demand of the tradable and non-tradable sectors in that state. We will solve a log-linearized version of the model and so will only be concerned with local versions of the transversality conditions.

It is shown in the appendix that a steady state in which all nominal variables grow at the same rate as the money supply exists. A key feature here is that nominal minimum wages in each state all grow at the same rate, the same rate as the money supply. Without this feature, the real minimum wage may change, leading to shifts in real variables.

3.5 Calibration and Solution

The log-linearized equations described above can be simplified to a $(3S+1)\times(3S+1)$ system of differential equations given by

$$\begin{pmatrix} \dot{\boldsymbol{c}}_t \\ \dot{\boldsymbol{k}}_t \\ \dot{\overline{\boldsymbol{w}}}_t \\ \dot{m}_t \end{pmatrix} \approx A \begin{pmatrix} \hat{\boldsymbol{c}}_t \\ \widehat{\boldsymbol{k}}_t \\ \widehat{\overline{\boldsymbol{w}}}_t \\ \widehat{m}_t \end{pmatrix}.$$

It is not difficult to solve this system numerically. Now the matrix A must be calibrated. Where possible, we use standard parameter values, as outlined in Table 1.

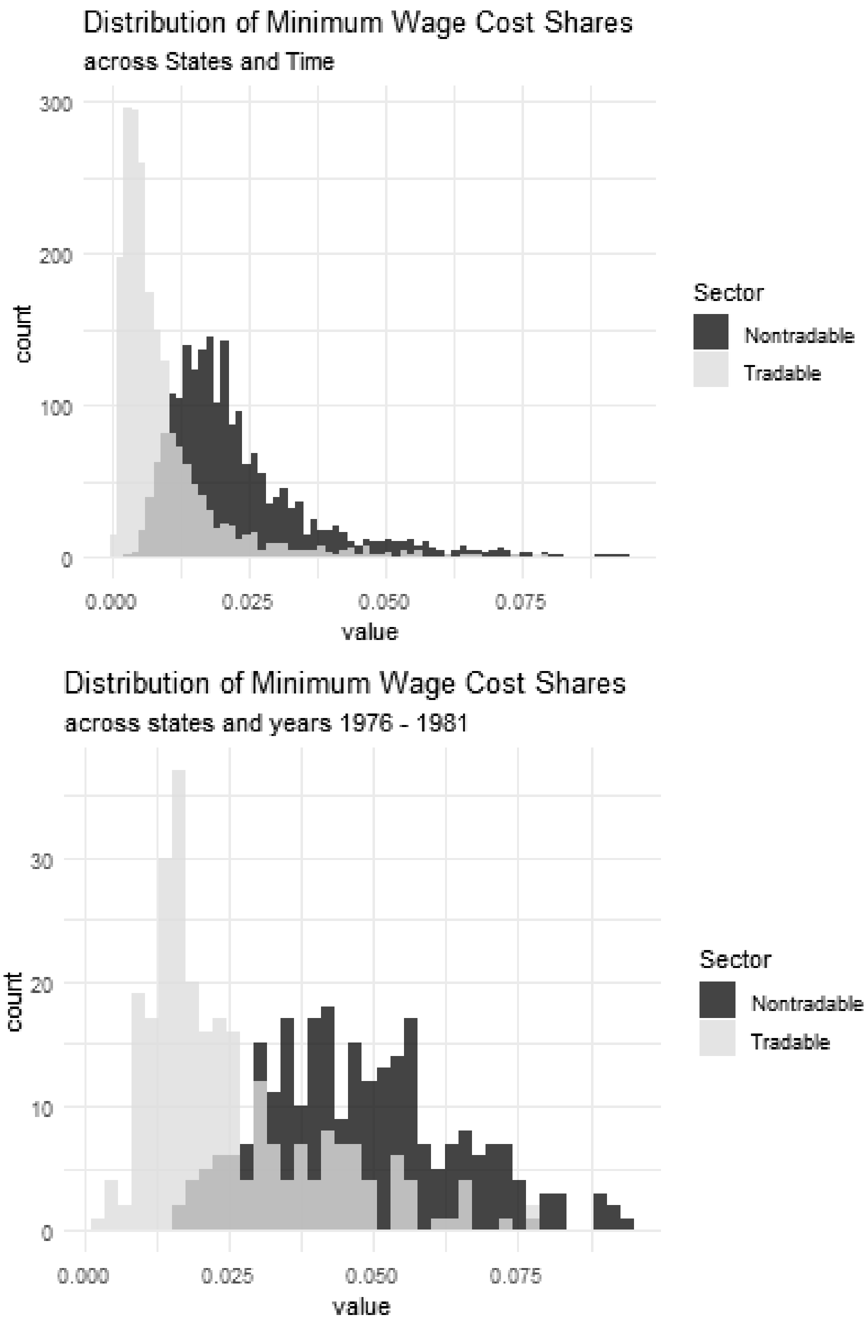
Note that we use the same elasticities of substitution in the table for both the tradable and non-tradable sectors. Our code can handle setting these separately. Now the elasticity σ_s is somewhat nonstandard. We use the Cobb-Douglas case, $\sigma_s = 1$, as our baseline. In the appendix, we show that the results do not change much when reducing σ_s to 0.1, a calibration close to Leontief that minimizes the ability of substitutions across state commodities in the tradable sector to drive the result that monetary policy has a larger effect on tradable employment than non-tradable employment through the minimum wage channel.

We also set $\sigma_{NT,T}$, the elasticity of substitution between tradables and non-tradables in consumption, to a value consistent with the Cobb-Douglas case, or $\sigma_{NT,T} = 1$. This is used by

Table 1: Parameter Values for Model Calibration

Parameter	***Description***	***Value***	***Notes***
γ	Risk aversion in steady state	2	Upper bound suggested by Chetty (2006)
ρ	Discount rate	.01	Quarterly
δ	Depreciation rate	.025	Quarterly
ϵ	Frisch elasticity of high skill labor supply in steady state	.4	Whalen and Reichling (2016)
σ_{HL}	High/low skill labor elasticity of substitution in steady state	1.41	Katz and Murphy (1992)
σ_{HK}	High skill labor/capital elasticity of substitution in steady state	.5	Oberfield and Raval (2020)
σ_{LK}	Low skill labor/capital elasticity of substitution in steady state	1.67	Krusell et al. (2000)

Figure 2: Cost Share Calibrations



Mian and Sufi (2014), which motivated our analysis of tradable and non-tradable employment in the first place.

The only way our calibration will differ with the time period we analyze is in the cost shares η. Our calibrations for the minimum wage cost shares are shown in Figure 2. These are computed as the minimum wage share of total payroll of wage and salaried workers in a given state, multiplied by the labor share in the state—more details are provided in the data section. It is clear in panel 1 of the figure that minimum wage cost shares in the tradable sector are typically lower than those in the non-tradable sector. The bulk of cost shares in both sectors are below 0.025 across states and time. We show the shares from 1976-1981 in panel 2 to highlight how much higher they are: more than half of the minimum wage cost shares in the non-tradable sector are higher than 0.05, and occasionally the shares reach the 0.10 range. Thus, in this period in certain states, 10% of production cost is subject to a price floor above the equilibrium price.

A key calibration is the magnitude of the shock to the effective money supply we feed into the model. We recall that $M_t = GDP_t$, which can be measured in the data as $P_t Y_t$. Romer and Romer (2004) studied the effects of shocks to the federal funds rate on prices and output, separately. Thus, the cumulative effects on a shock to $\ln M_t$ can be measured as the sum of the cumulative effects on $\ln P_t$ and $\ln Y_t$. The effect of a 1 percentage point increase in the federal funds rate, measured in this way, accumulates to a 4% decline in M over two years (a 4% decline in output and 0% decline in prices) and a 7% decline in M over four years (a 1% decline in output and a 6% decline in prices). To be conservative, since shocks other than those from Romer and Romer (2004) usually lead to smaller effects, and because we are not particularly interested in the impulse response functions resulting from this model, we will calibrate a 1 percentage point shock to the federal funds rate as an unanticipated and permanent 4% shock to the money supply.

Figure 3: Model Outcomes

Panel 1

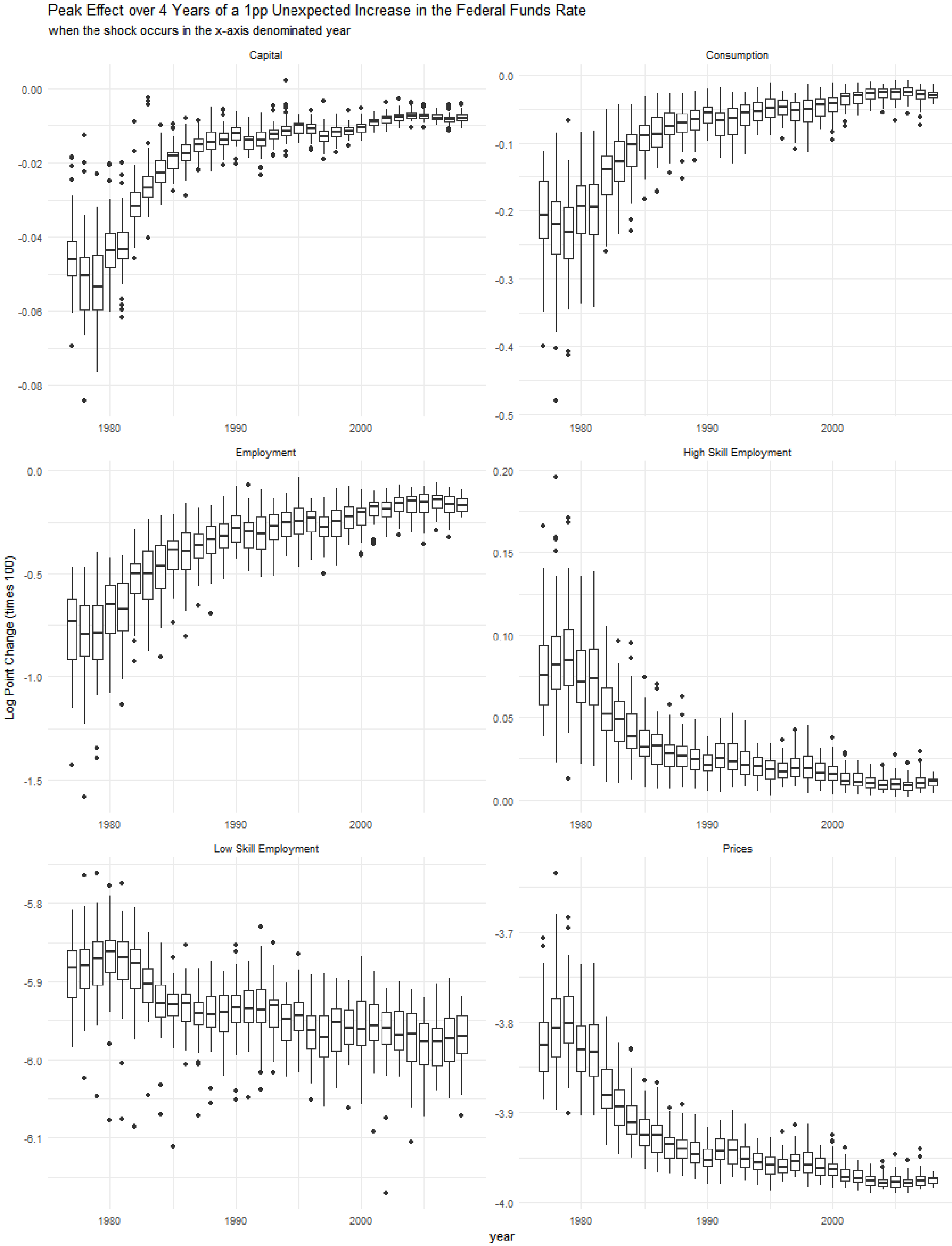
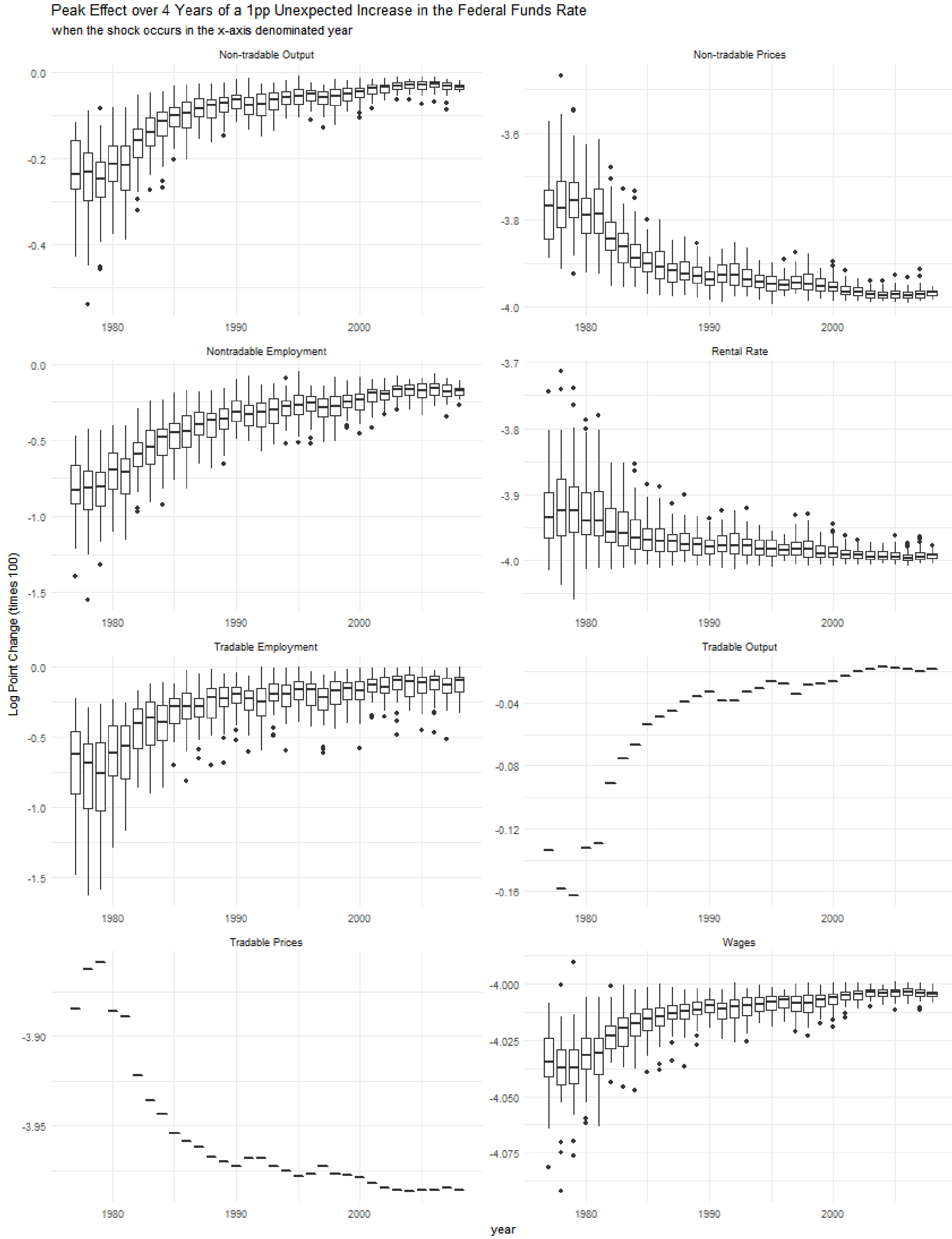


Figure 4: Model Outcomes (cont'd)

Panel 2



3.6 Model Results

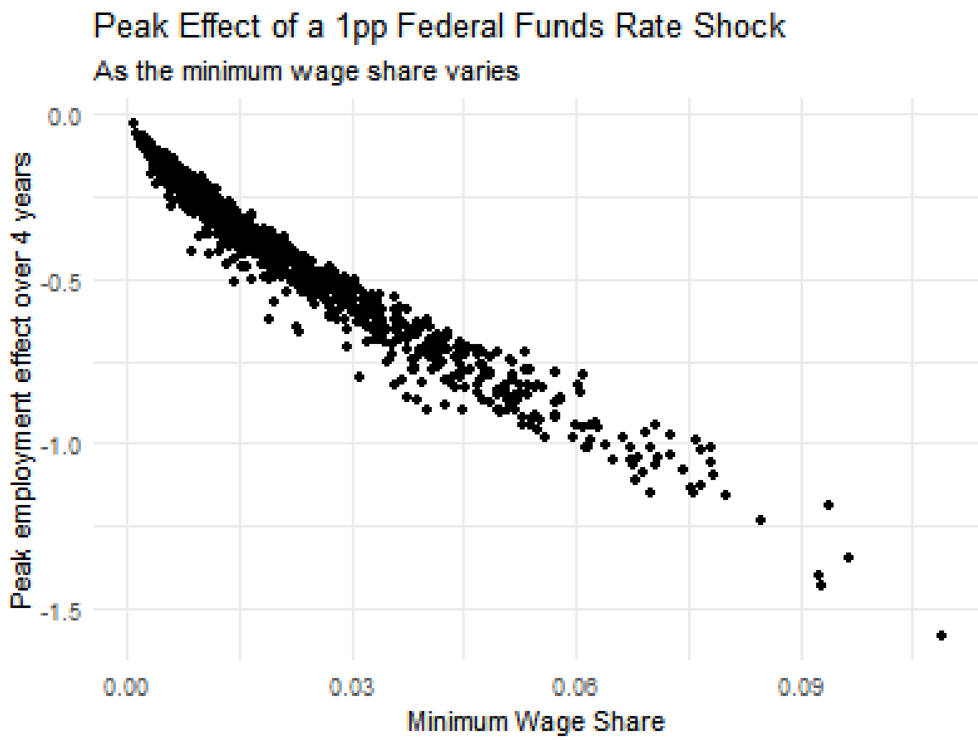
We vary our cost share calibrations by year and include all 50 states in the model, so it is infeasible to present impulse response functions to summarize our results. Instead, for each year of calibration and each outcome variable of interest, we compute the impulse response functions over a 4-year horizon and take the largest magnitude effect achieved over that 4-year horizon for each state. We summarize these maximal results in each year using a boxplot. Figure 3 shows these graphs for various outcomes of interest in two panels.

We wish to highlight several outcomes of interest, particularly in the late 1970s and early 1980s. In these years, the minimum wage channel of monetary policy contributes an overall employment decline of about -0.75% in response to a 1 percentage point increase in the federal funds rate. This effect is driven by low wage employment, which declines by nearly 6%. With some outlier exceptions, high skill employment actually increases slightly, by around 0.075%. Capital and consumption both fall somewhat, by -0.05% and -0.2%, respectively. Prices fall by less than 4%, the magnitude of the shock to the money supply, highlighting that our channel is fundamentally a supply shock. Put differently, contractionary monetary policy raises the real cost of production by increasing the nominal minimum wage relative to other prices, and so prices fall by less than what would be predicted under monetary neutrality. The heterogeneity in these effects over time is less than the heterogeneity in employment, however.

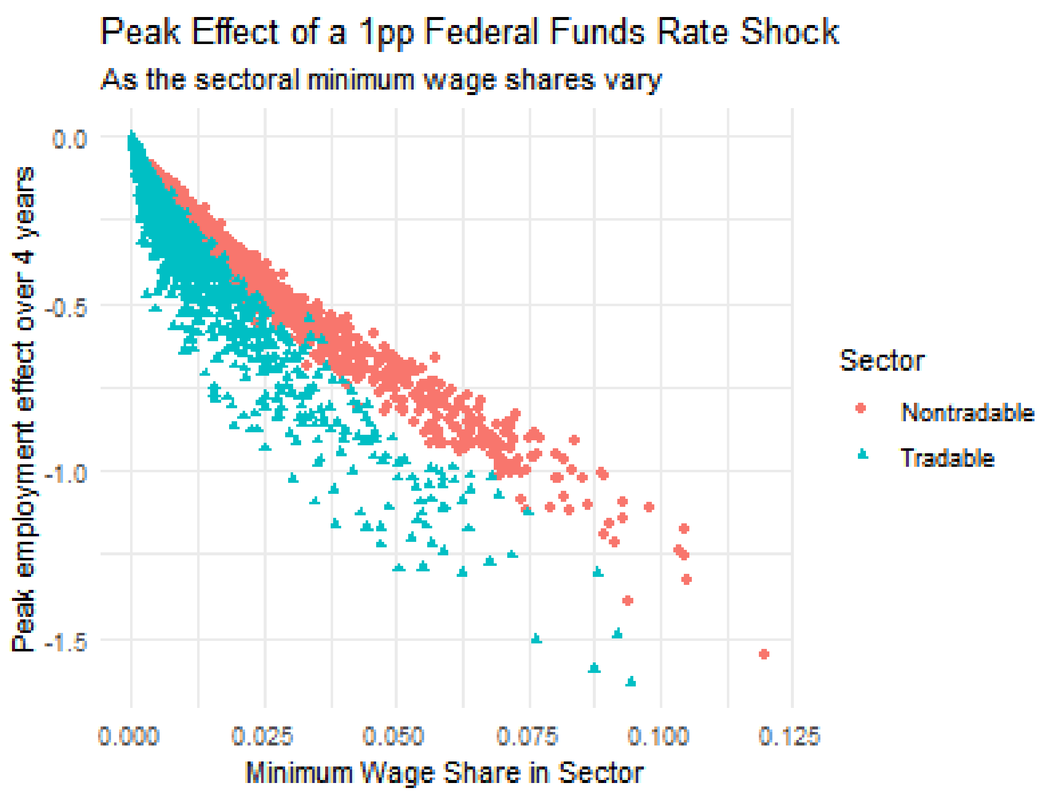
We also want to highlight how our results look much closer to monetary neutrality from the late 1980s onward. This is the period where the minimum wage share has become small, meaning less economic cost is at the binding wage floor. We think the fact that a declining proportion of economic cost has been at the binding minimum wage over time contributes to findings that the effects of monetary policy may have fallen over time.

Figure 5: How Employment Effects Vary with the Minimum Wage Share

Panel 1



Panel 2



Panel 2 of Figure 3 shows that median states experience smaller declines in tradable employment than non-tradable employment, but the minimum wage cost shares in tradable employment are often much smaller; it is also apparent that heterogeneity across states in tradable employment is much larger than in non-tradable employment. Figure 5 highlights how our employment effects depend on the minimum wage share in a state. Panel 1 shows how our overall employment effects increase in magnitude with the minimum wage share in the state. Panel 2 shows how our tradable employment effects increase in magnitude more quickly with the tradable minimum wage share than the non-tradable employment effects do with the non-tradable minimum wage share. As we discussed above, this is not driven by cross-state substitution in the tradable sector as much as it is by large capital shares, combined with the large elasticity of substitution between capital and minimum wage labor.

4 Empirical Framework

But does this channel of effect for monetary policy highlighted by the model actually exist in practice? To answer that question, it is necessary to conduct some empirical analysis. We begin by using standard data sources for employment – the QCEW and the CPS – and an adapted version of a very simple and standard specification from the narrative monetary policy literature. We subsequently branch out from this specification and run a broad variety of robustness checks intended to encapsulate many potential critiques of the baseline specification.

4.1 Data

We obtain data on narrative monetary policy shocks from Coibion et al. (2017). Coibion et al. follow the technique devised by Romer and Romer (2004), who obtained narrative records of the Federal Reserve's intentions for the federal funds rate around FOMC meetings and regressed this series on internal Fed forecasts "to derive a [monthly] measure free of systematic responses to information about future developments." The Romer and Romer (2004) series of monetary shocks has become one of the canonical sources of exogenous variation used in the monetary policy literature. Because the series initially terminated in 1996, however, Coibion et al. extended it through 2015. We also obtain an alternative VAR shocks series from Coibion (2012), a paper dedicated in part to explaining why the Romer and Romer (2004) shocks generate such large effects of monetary policy. This VAR series yields effects of monetary policy close to those found in Christiano, Eichenbaum, and Evans (1999), somewhat smaller than those found in Leeper, Sims, and Zha (1996) and somewhat larger than those found in Bernanke and Blinder (1992); in particular, these shocks lead to output effects that are roughly six times smaller than those found in Romer and Romer (2004). Further, the VAR series is much less sensitive to the inclusion of monetary policy episodes in 1980 that drive the estimated Romer and Romer (2004) shock effects to be large.

For two key reasons, we will not use monetary policy shocks derived from the high-frequency identification literature, which includes Cook and Hahn (1989), Kuttner (2001), Cochrane and Piazzesi (2002), Gürkaynak, Sack, and Swanson (2005), and, more recently, Gertler and Karadi (2015) and Nakamura and Steinsson (2018). First, we have the most power to detect our results in the late 1970s and early 1980s, when there was a relatively high share of minimum wage workers. The earliest futures data used in Nakamura and Steinsson (2018) begins in 1995, and their measurements of real interest rates require TIPS data – TIPS were issued beginning in 1997. We cannot use the futures rate surprises from Gertler and Karadi (2015) because the key data is available from 1991 to 2012, leaving out the core period during which we want strong, exogenous monetary policy variation. Second, we question the power of these shocks in our context become some of these series are able to detect effects of monetary policy on financial variables but not real variables such as output and employment.

For data on employment, we turn to the Quarterly Census of Employment and Wages (QCEW), which has collected population data on employment by county by industry in the United States since 1937. Despite the name of the dataset, employment data is available at the monthly level[3]. Digitized data from January 1975 onward is readily available for download on the Bureau of Labor Statistics (BLS) website. For our baseline regression specification, we use the state-level figures aggregated across all industries, but for certain alternative specifications – such as our specification with as our within-state county-level design – we make use of the underlying county-level and/or industry-specific data.

We use the CPS Outgoing Rotation Groups (CPS-ORG) to compute the share of minimum wage workers as a proportion of all workers by state and year. Households in the CPS sample respond to the questionnaire for four months in a row; they are then out of the sample for eight months; finally, they return to the sample for another four months. At the end of each of the two four-month blocks during which a household is present in the CPS sample, they are asked a specific set of questions not asked in other months. These questions – which include amongst them an explicit question about what the respondent's hourly wage is – make up the Outgoing Rotation Groups questionnaire. Because the monthly size of the CPS is 60,000, this means that the monthly size of the ORG is (approximately) 15,000 – an annual sample size of 180,000. By merging this data with Vaghul and Zipperer's (2016) dataset on historical state and federal minimum wages, we can identify minimum wage workers as any wage or salaried worker whose computed hourly

[3] The reason for the apparent discrepancy is that the census is conducted quarterly but asks employers how many workers were on their payroll at the end of each of the three months of the preceding quarter.

wage is within a one dollar band around their state's binding minimum wage (i.e., between 50 cents below and 50 cents above). Our results are practically identical if we instead use percentages—defining near minimum wage workers as those within 10% of the minimum wage—and they are scarcely changed if we widen to (20%) or narrow (to 5%) the band. We can then compute the minimum wage share of payroll in a state as the total payroll to minimum wage workers in the state divided by total payroll in the state.

To compute the minimum wage share in the state, we multiply the payroll share computed above by the labor share in the state. The Bureau of Labor Statistics (BLS) describes how they compute the national labor share, and we implement this procedure at the state level. The statistic is computed by dividing total compensation in a state by total GDP in a state, using data published by the Bureau of Economic Analysis (BEA). This share is then adjusted upwards for proprietors' incomes due to their own work at their businesses, which is not included in total compensation as measured by the BEA. We show how our computed labor share at the national level compares to the BLS's labor share in the appendix.

It is worth noting that an alternative approach to using the CPS-ORG is to use the Current Population Survey's Annual Social and Economic Supplement (CPS-ASEC) to compute the share of minimum wage workers. The ASEC commenced in 1962 and since its inception has asked respondents their total wage income, weeks worked, and hours worked per week over the last year. From this, it is possible to compute each individual's hourly wage. However, prior to 1977, the aforementioned variables were binned, so weeks and hours worked – and thus hourly wage – can only be approximately known. And compared to the ORG's annual sample size of 180,000, the ASEC has an annual size of 60,000. As a consequence of its lower sample size, the approximation implicit necessary as a part of the preceding process, and the fact that the QCEW data is only available from 1975 onward anyway, we use the ORG instead of the ASEC. Having said this, our results are virtually unchanged if we instead use the ASEC data.

As noted, neither the ASEC nor the ORG are of sufficient size to calculate county-level annual minimum wage shares. As such, given that state minimum wage shares are a very slow-moving variable, we turn to the Census. Using the IPUMS 5% public-use samples of the 1980, 1990, and 2000 Censuses, we compute county-level minimum wage payroll shares for use in our within-state specifications. Because BEA data on GDP at the county level does not start until 2001, we cannot easily compute our full minimum wage share at the county level. This will motivate us showing regressions using the county-level minimum wage payroll shares and alternative specifications using minimum wage shares computed using county minimum wage payroll shares multiplied by the state labor share.

We also obtain data on some additional control variables for robustness checks. We obtain data on per-capita bank deposits by county from the FDIC and data on personal income per-capita by county from the Bureau of Economic Analysis (BEA). We obtain a Canadian narrative monetary shocks series (constructed analogously to the Romer and Romer shocks) from Champagne and Sekkel (2018), and we obtain Canadian data on monthly employment and the share of minimum wage workers by province from Statistics Canada's Monthly Labour Force Survey Public Use Microdata File (PUMF).

4.2 Identification

Our baseline specification is an adapted version of the standard narrative-shocks monetary policy regression. We add interaction effects between the shock variable and the minimum wage share, and we additionally two-way cluster our standard errors at the state and time level in order to account by complex correlation structures induced by the fact that our dataset is a panel dataset but the monetary shock series is state-invariant.

$$\Delta E_{s,t} = \alpha + \sum_{i=0}^{48} \beta_i Shock_{t-i} + \gamma MWShare_{s,t} + \sum_{i=0}^{48} \delta_i Shock_{t-i} \cdot MWShare_{s,t} + \sum_{i=1}^{48} \eta_i \Delta E_{s,t-i} + \varepsilon_{s,t} \quad (1)$$

where $\Delta E_{s,t}$ denotes employment growth in state s at time t, $MWShare_{s,t}$ denotes the minimum-wage labor share of costs in state s at time t, and $Shock_t$ denotes the (nationwide) monetary policy shock at time t.

In various robustness checks, we enhance this specification with additional control variables and/or different approaches to identification. First, we add state and time fixed-effects to the specification to account for all time-varying, state-invariant and state-varying, time-invariant confounds. Separately, we add controls for the interaction effect of a couple of other variables with the shock series: bank deposits per-capita and per-capita income (proxying for the marginal propensity to consume). The idea is that there may remain some crucial variables correlated with minimum wage share that could plausibly be the true channel for monetary policy efficacy, rather than the minimum wage share itself.

Observing that changes in the minimum wage share in a state can be driven either by plausibly-exogenous factors such as minimum wage changes or by more endogenous factors such as changes in the share of each industry in that state's employment, we construct a Bartik-type variable that controls for the latter effect and add it to our baseline regression. In particular, we construct the variable by computing

$$\Delta S_{s,t} = \sum_j Shift_{-s,j,t} \cdot Share_{s,j,t-1}, \quad (2)$$

where $Shift_{-s,j,t}$ represents the national-level growth of employment in industry j over time period t (calculated as a leave-one-out average) and $Share_{s,j,t-1}$ represents the employment share of industry j in total state-s employment in the preceding time period $t-1$. This shift-share isolates the national-level, non-idiosyncratic component of growth in employment that stems from broader trends. Adding this control to the specification should help ensure that the effect we are finding is not driven by that more endogenous source of minimum-wage-share variation. As an alternative approach to isolating the plausibly-exogenous variation, we run an IV specification instrumenting the state minimum wage share with the state minimum wage itself. In particular, for our first-stage, we instrument the direct effect and the interaction effects involving the minimum wage share with the corresponding minimum wage variables as follows,

$$MWShare_{s,t} = \omega + \rho MinWage_{s,t} + \sum_{i=0}^{48} \theta_i Shock_{t-i} \cdot MinWage_{s,t} + u_{s,t} \qquad (3)$$

$$Shock_{t-i} \cdot MWShare_{s,t} = \chi_i + \rho MinWage_{s,t} + \sum_{i=0}^{48} \theta_i Shock_{t-i} \cdot MinWage_{s,t} + u_{s,t}^{i} \qquad (4)$$

and the second-stage constitutes placing the predicted values of the left-hand-side variables from these first-stage regressions back into our baseline specification.

As another enhancement of our baseline specification, to factor out potential concerns of correlated state policymaking, we add state-by-time fixed-effects and rely on the county-level data in order to pursue a within-state identification strategy. Additionally, we run another within-state specification – similar to a triple-differences specification – comparing near-minimum-wage employment growth to higher-wage employment growth to confirm that our effects are indeed driven by near-minimum-wage workers. That is,

$$\Delta E_{s,w,t} = \alpha + \sum_{i=0}^{16} \delta_i \left| Shock_{t-i} \right| \cdot \mathbf{1}\{NearMinWage\} + \sum_{i=1}^{16} \eta_i \Delta E_{s,w,t-i} + \omega_{s,t} + \theta_{s,w} + \varepsilon_{s,t} \qquad (5)$$

where $\omega_{s,t}$ denotes state-by-time fixed effects and $\theta_{s,w}$ denotes state-by-wage-group fixed-effects. Note that QCEW data on employment by wage group is unavailable. So, for this specification, as our left-hand-side variable, we compute separate series of near-minimum-wage and higher-wage employment growth using the CPS-ORG data. Also, observe that we are taking the absolute value of the shock series in this specification. This is because our model suggests that near-minimum-wage employment should respond more strongly *in magnitude* to monetary policy shocks – i.e., that expansionary shocks ($Shock_t < 0$) should induce more employment growth for this group and that contractionary shocks ($Shock_t > 0$) should induce more employment loss (*less* employment growth) for this group.

Finally, to be completely parallel with our model, we run the baseline specification (1) over tradables and non-tradables separately. And we similarly use another within-state specification comparing employment growth in tradable versus non-tradable sectors within-state:

$$\Delta E_{s,k,t} = \begin{array}{c} \alpha + \sum_{i=0}^{48} \beta_i \left| Shock_{t-i} \right| \cdot \mathbf{1}\{Tradable\} + \gamma MWShare_{s,k,t} \\ + \sum_{i=0}^{48} \delta_i \left| Shock_{t-i} \right| \cdot \mathbf{1}\{Tradable\} \cdot MWShare_{s,k,t} + \sum_{i=1}^{48} \eta_i \Delta E_{s,k,t-i} + \omega_{s,t} + \theta_{s,k} + \varepsilon_{s,t} \end{array} \tag{6}$$

where $\omega_{s,t}$ denotes state-by-time fixed effects and $\theta_{s,k}$ denotes state-by-tradability fixed-effects. Note that in this setting the coefficients δ_i record the interaction effect of monetary policy shocks on tradable employment relative to non-tradable employment as a function of the minimum-wage labor share, and the "level effect" coefficients β_i measure the effect of monetary policy shocks on tradable employment relative to non-tradable employment independent of the minimum wage share.

5 Results

5.1 Main Results

Beginning with the baseline specification, Figure 6 depicts its results in the form of an impulse response function cumulating the interaction effect over time. The error bands represent 90% confidence intervals. Note that the magnitude of the interaction effect peaks at -1. Thus the figure can be interpreted as follows: a 1 percentage-point *higher* minimum wage share corresponds to 1 percentage-point *lower* employment growth (at peak) as a result of a 1 percentage-point

Figure 6: Baseline Specification

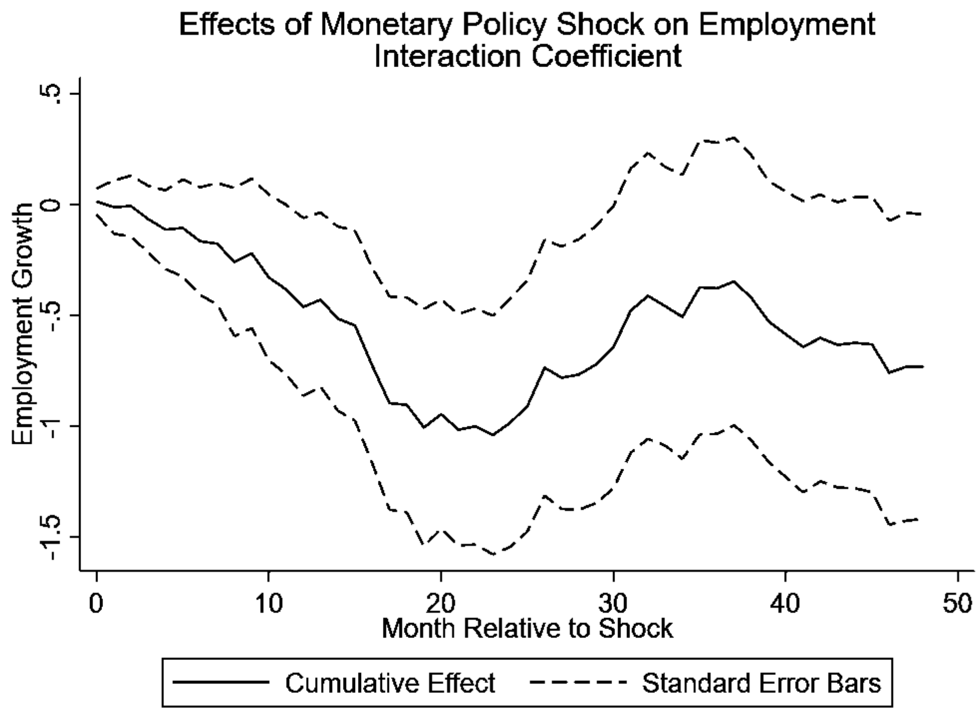
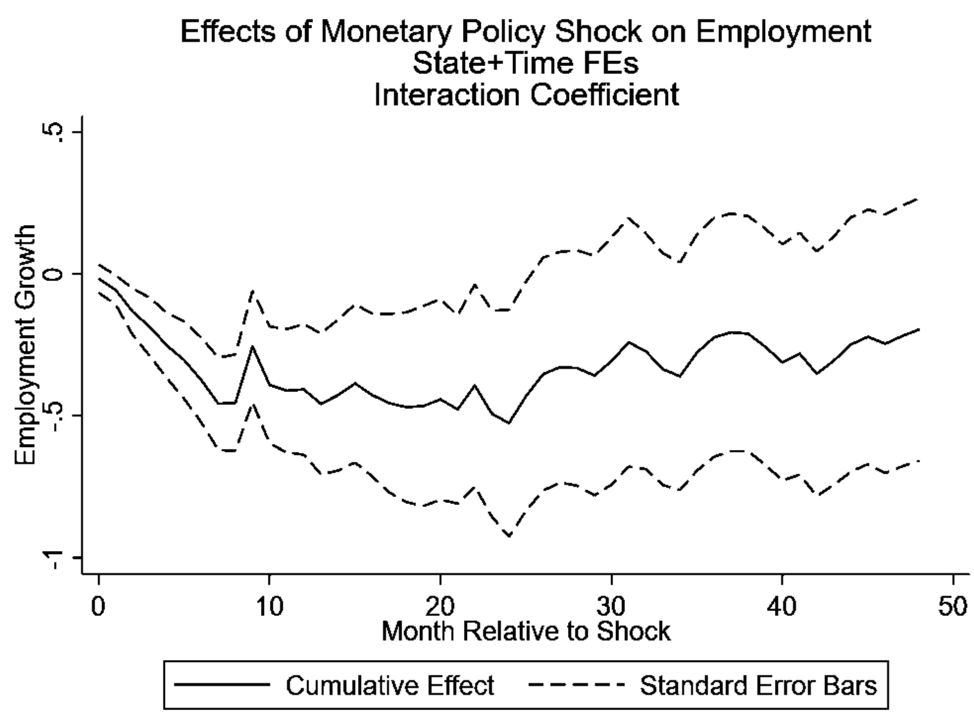


Figure 7: Baseline with State and Time FEs



Romer and Romer contractionary monetary policy shock (i.e., a 1 percentage-point unexpected increase in the Fed Funds Rate). Stated more intuitively, a state at the 90th percentile of the minimum wage share will experience a peak employment effect of a 1 percentage-point Romer and Romer federal funds rate shock that is approximately 2.5 percentage-points higher than a state at the 10th percentile of the minimum wage share.

Now, suppose certain states are more responsive to monetary policy for reasons unrelated to the minimum wage. For example, poorer states are likely to have a higher average marginal propensity to consume, which should boost monetary policy efficacy through more traditional channels. Similarly, suppose the efficacy of monetary policy is declining over time for, again, reasons unrelated to the minimum wage. Because the share of minimum wage workers is also declining over time, this could plausibly pollute the coefficients we estimate in the baseline model. We address these potential concerns by adding state and time fixed-effects. The resulting impulse response function is plotted in Figure 7. Notably, the effect not only survives – it is made more strongly significant than in the baseline specification. The magnitude, however, is (non-significantly) smaller by a factor of one-half.

Another concern is that the industries which have the highest share of minimum wage workers might just be the industries that are most affected by monetary policy – for reasons unrelated to the minimum wage share itself. If these industries are concentrated in specific states, that could be driving our results. To deal with this concern, Figure 8 turns to the Bartik controls, adding them to the baseline specification. The idea here is that controlling for the Bartik instrument purges the component of employment growth driven by broad industrial trends; the remaining unexplained left-hand-side variation in economic growth is not a consequence of which industries happen to be concentrated in which states. As can be seen in the left panel of Figure 8, the effect survives and,

Figure 8: Bartik Specification

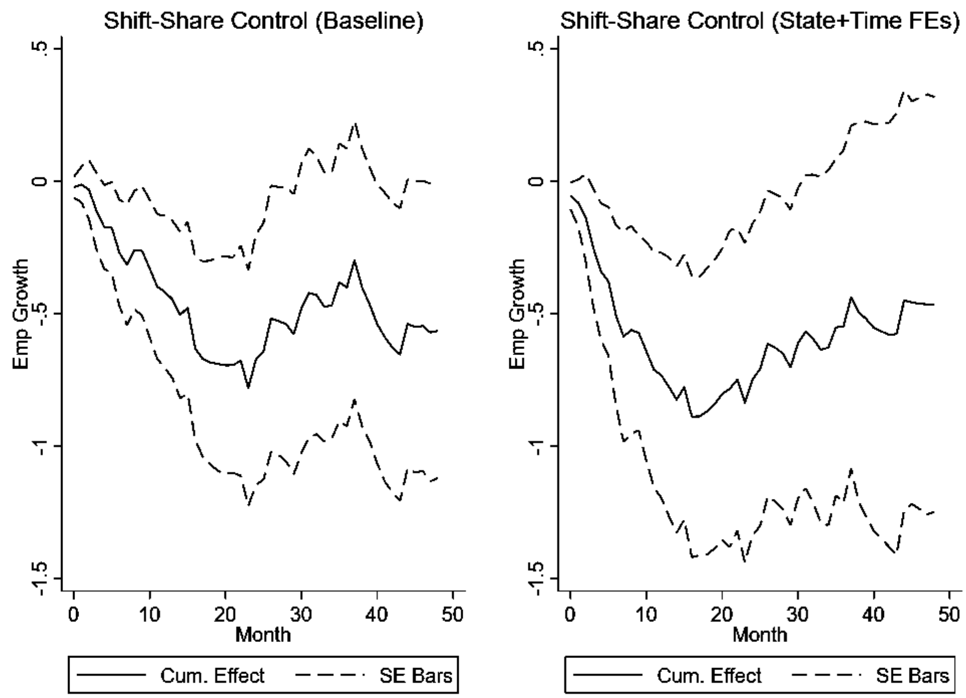
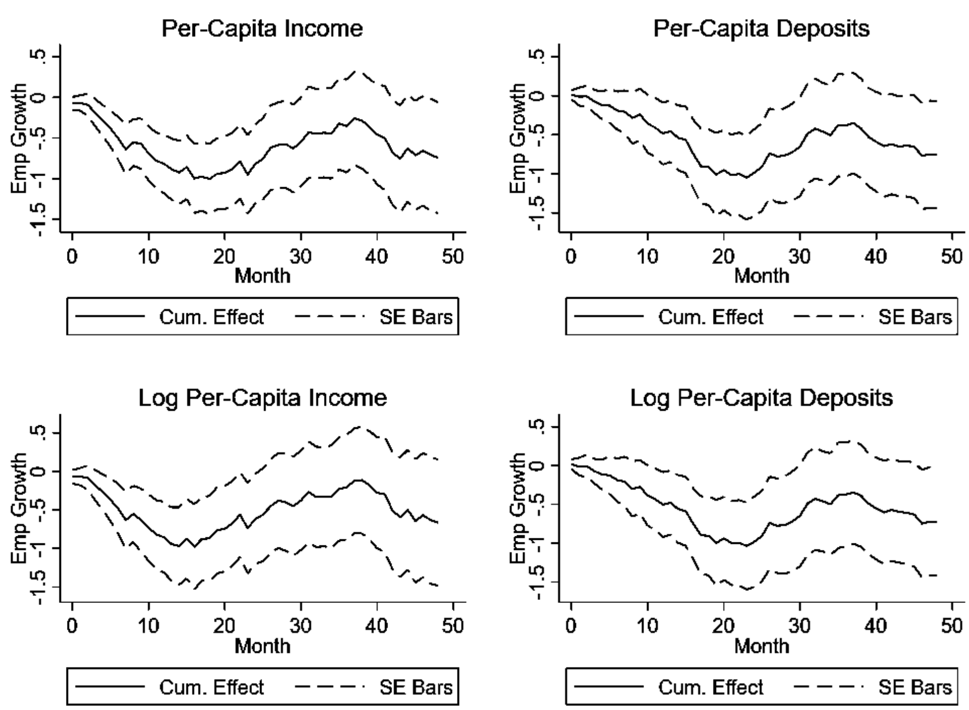


Figure 9: Other Controls



indeed, is little changed from baseline. The right panel of Figure 8 adds *both* the Bartik control and the state and time FEs to our baseline specification, combining the desirable characteristics of both of these robustness checks. Here, too, the effect retain significance. The magnitudes are in-between the baseline specification and the specification with only state and time fixed-effects.

In Figure 9, we interact some control variables with the Romer and Romer shock series (and its 48 lags) to help demonstrate that the effect we find is not due to correlation of the minimum wage share with other important variables that affect monetary policy efficacy. We can see that doing this with bank deposits per-capita and per-capita income – two variables which are particularly likely to correlate with MPC – do not materially change our result from baseline.

Variation in the state minimum wage share may come from a variety of sources – including changing industry shares within the state. Again, this source of variation may be somewhat endogenous. As a result, a somewhat different approach from the Bartik control of factoring out this industry-correlate-driven variation in the minimum wage share is to instrument for the minimum wage share with the legislated state minimum wage. This isolates variation driven by political decisions on the part of the state legislature, plausibly a more exogenous source of variation than changing industry shares. Figure 10 turns to this IV strategy. Again, the result survives; the magnitude of the point estimate, however, increases by a factor of approximately 2, though our previous results remain within the standard error bars of this point estimate.

The specification represented in Figure 11 makes use of the VAR shocks from Coibion (2012) instead of the Romer and Romer shocks. The monetary policy literature has proceeded along two main strands – one pursuing a narrative approach and the other pursuing a VAR-based approach. We aim to show that our result goes through regardless of which shocks series we use; it's not an artefact of one approach or the other. While the shape of the impulse response function

Figure 10: IV Specification

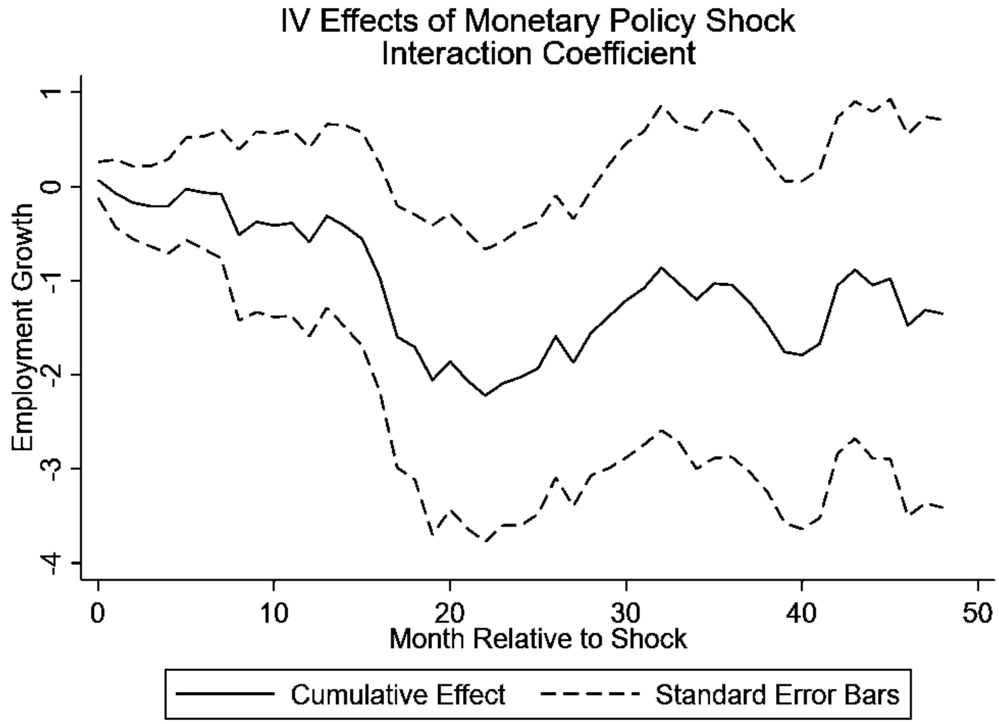
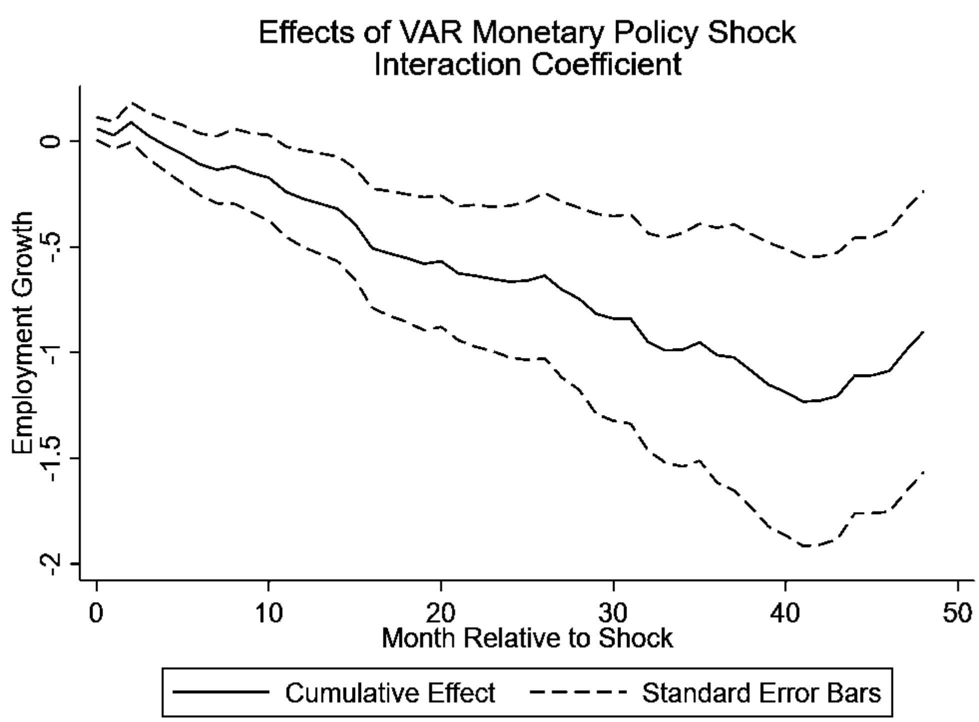


Figure 11: VAR Shocks Specification



is somewhat different – with the effect peaking some two years later than in the baseline specification – the magnitude is the same, as is the takeaway: monetary policy is significantly more effective where the share of minimum wage workers is higher.

Figure 12 plots the results of running the baseline specification on the province-level Canadian data. As in the case of the VAR robustness check, we do not necessarily have a reason to believe that the US data is inferior to the Canadian data (or vice versa) – we merely regard it as a second laboratory in which to test our hypothesis and provide evidence of its generality. Again, despite a somewhat modified shape of the impulse response function, the peak magnitude is nearly the same and the evidence remains that a higher share of minimum wage workers significantly boosts monetary policy efficacy.

We next turn to the within-state specification. Suppose, for example, that expansionary monetary policy causes states with a low share of minimum wage workers (i.e., where the minimum wage isn't very binding) to increase the state minimum wage in order to prevent it from being further devalued by the price level increases induced by the monetary policy. Insofar as the increased minimum wage reduces employment growth, we could be picking up this effect in our baseline regressions. Now, it's worth noting that this is still part of the causal chain – in this example, expansionary monetary policy is still technically causing higher employment growth in high minimum-wage-share states relative to low minimum-wage-share states – so this conjecture would not mean that the effects we estimate in the baseline specification cannot be interpreted causally. However, we might nonetheless be interested in honing in on the component of the overall effect mirroring our model, factoring out the aforementioned side-channel. Figure 13 presents the specification that adds a state-by-time fixed-effect to the baseline specification in order to exploit county-level variation in the minimum wage share. This factors out any state

Figure 12: Canada Specification

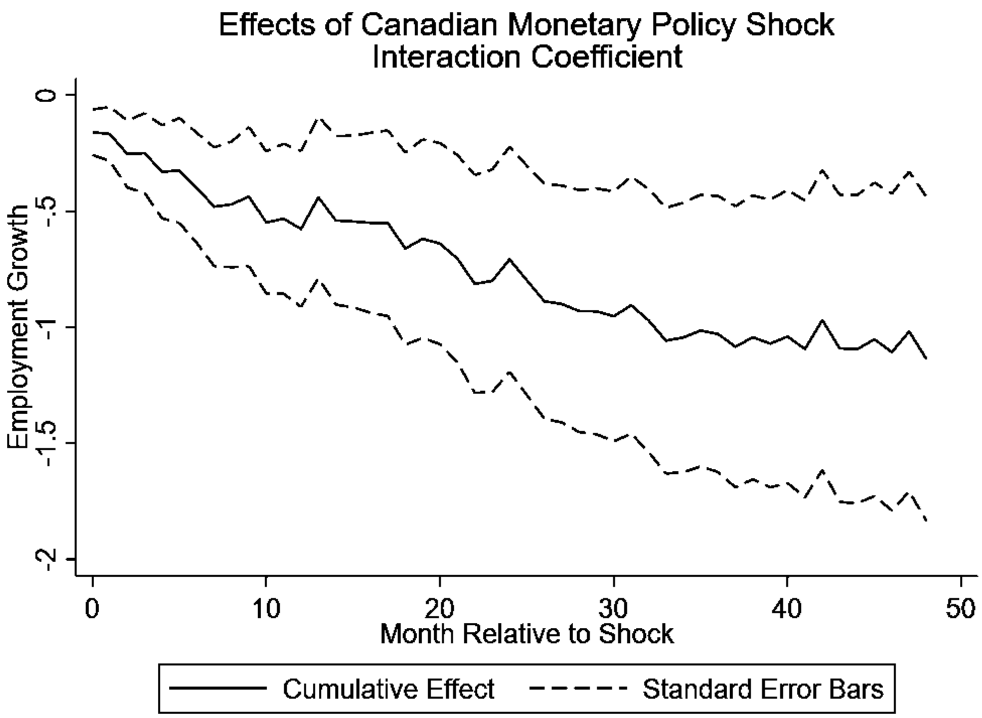
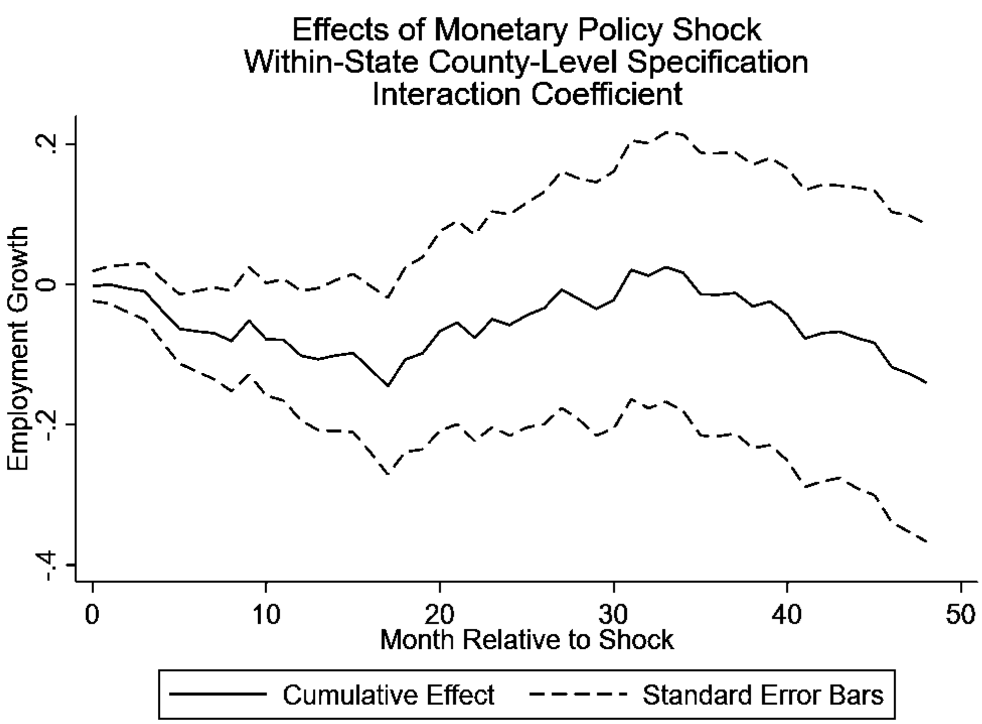


Figure 13: County-Level Within-State Specification



policy responses. Here, too, an impulse response function similar to the baseline is reproduced – notably, though, the effect size is smaller, perhaps suggesting some evidence that the side-channel does exist. It is vital to mention that, because we need to calculate the county-level minimum wage shares from the Decennial Census data, we use 1980 shares through 1989, 1990 shares through 1999, and 2000 shares through 2009[4]. This introduces potentially substantial measurement error in our minimum wage shares and, consequently, is likely to bias the result downward (due to attenuation bias) – another potential reason for the reduced magnitude.

5.2 The Mechanism: Testing Model Implications

The within-state specification comparing the effects of monetary policy on near-minimum-wage employment to higher-wage employment is slightly more complex. The prediction of our model is that expansionary monetary policy should have more positive effects on near-minimum-wage employment than on higher-wage employment, whereas contractionary monetary policy should have more negative effects on near-minimum-wage employment than on higher-wage employment. Consequently, in our regression specification, we examine whether the *absolute value* of the effect is higher on near-minimum-wage labor. As can be seen in Figure 14, the effects on near-minimum-wage labor are massively higher than those on higher-wage labor, consistent with the mechanism laid out by the model. In the left panel, we define "near-minimum-wage" workers as those less than 125% of their state's minimum wage; in the right panel, we use 150% as the threshold. Note that the magnitudes are quite high, but near-minimum-wage labor makes up less than 10% of total labor, so the very large coefficients in this specification reflect that the

[4] Because our employment data begins in 1975 but the publicly-available sample of the 1970 Census is much smaller (1% sample) than the 1980 Census – and thus leaves virtually all counties unidentified in the data – we extend the 1980 minimum wage share back through 1975. Instead omitting years 1975-1979 does not substantially alter the results.

Figure 14: Near-Minimum-Wage vs. Higher-Wage Within-State Specification

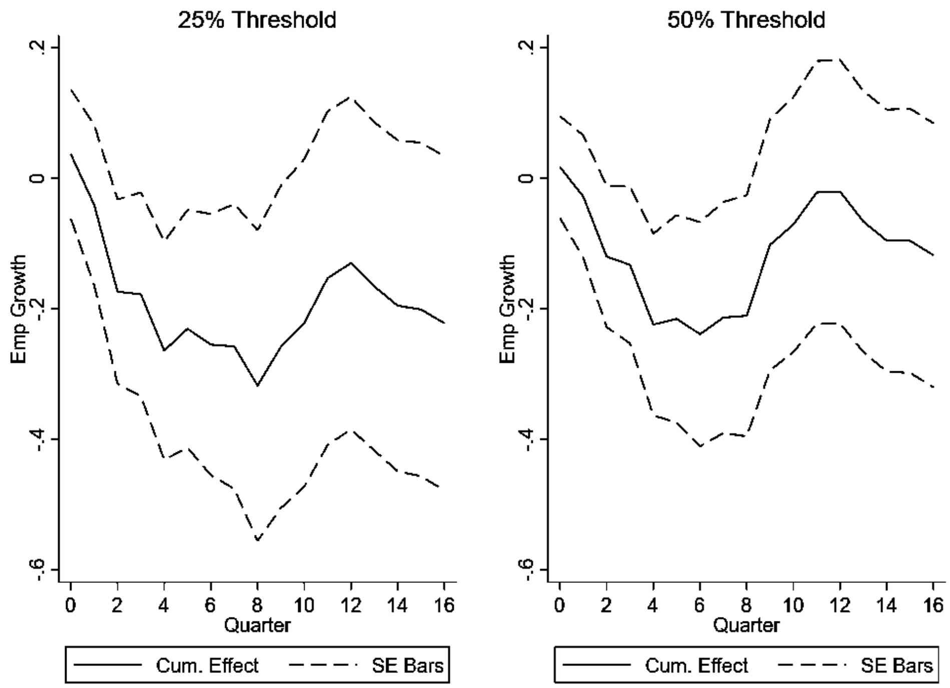
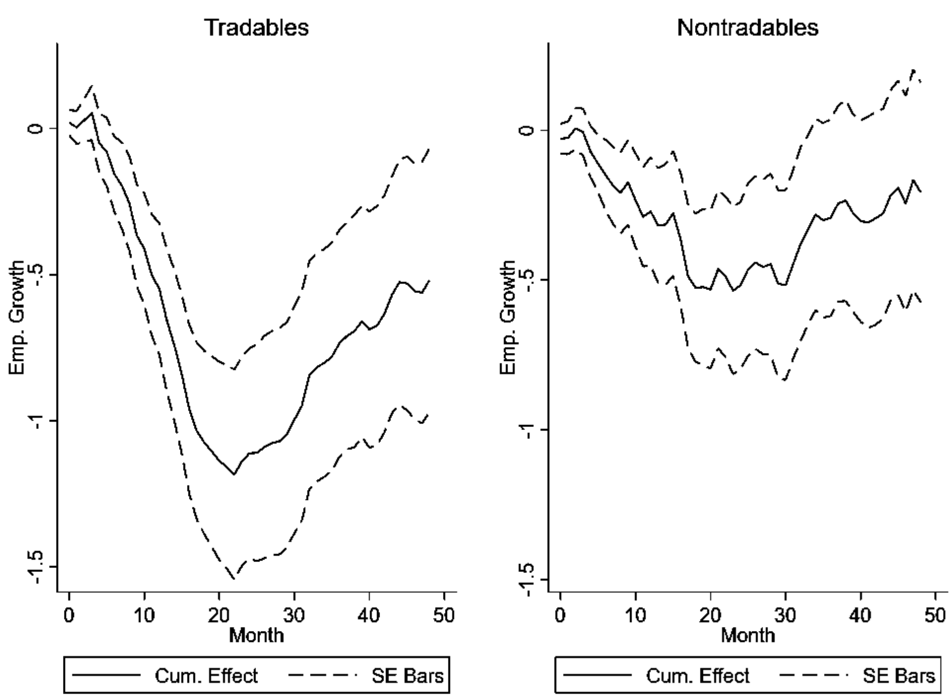


Figure 15: Baseline Specification for Tradable and Non-Tradable Employment



overall 5% increase is being driven quite disproportionately by near-minimum-wage workers, as one would expect.

In order to further validate the mechanism of the effect, we decompose employment growth into employment growth in tradables and employment growth in non-tradables[5]. We run a version of the baseline specification that interacts the Romer and Romer shocks and their 48 lags with the tradable minimum wage share and, separately, a version of the baseline specification with the non-tradable minimum wage share as the interaction term. As we show in Figure 15, a higher minimum wage share significantly boosts monetary policy efficacy in both tradables and non-tradables. The effect is not driven by non-tradables. Indeed, if anything, the effect is stronger on *tradable* employment, precisely the opposite of what the MPC channel would suggest. Figure 16 runs a within-state version of this specification – comparing tradables to non-tradables – and makes it even more directly clear that the effect is not driven by non-tradables. To the extent the effects on the two groups are ever significantly different, the effect on tradable employment is larger.

One remaining concern is that the effect on tradables is only larger (or the same size) as the effect on non-tradables because of business-stealing effects. That is, since tradables can more easily be produced in one state and then transported/sold in another state, a state with a low minimum-wage labor share may simply siphon business from states with high minimum-wage labor shares in response to a contractionary monetary policy shock. Because the minimum wage is (mostly) non-binding in the former state and binding in the latter, an increase in the real minimum wage reduces the relative cost of business in the former state. If this dynamic drives the

[5] We define "Tradables" as the Agriculture, Mining, Manufacturing, and Finance sectors. We define "Non-tradables" as the Construction, Transportation, Communications, Utilities, Retail Trade, Wholesale Trade, Services, and Public Administration sectors. There is no one authoritative definition of these two terms, and some classifications omit Finance from the Tradables category and/or omit Wholesale Trade and parts of Services from the Non-tradables category. Omitting some or all of these sectors from our classification does not materially change our results.

Figure 16: Tradables vs. Non-Tradables Within-State Specification – Interaction Effect

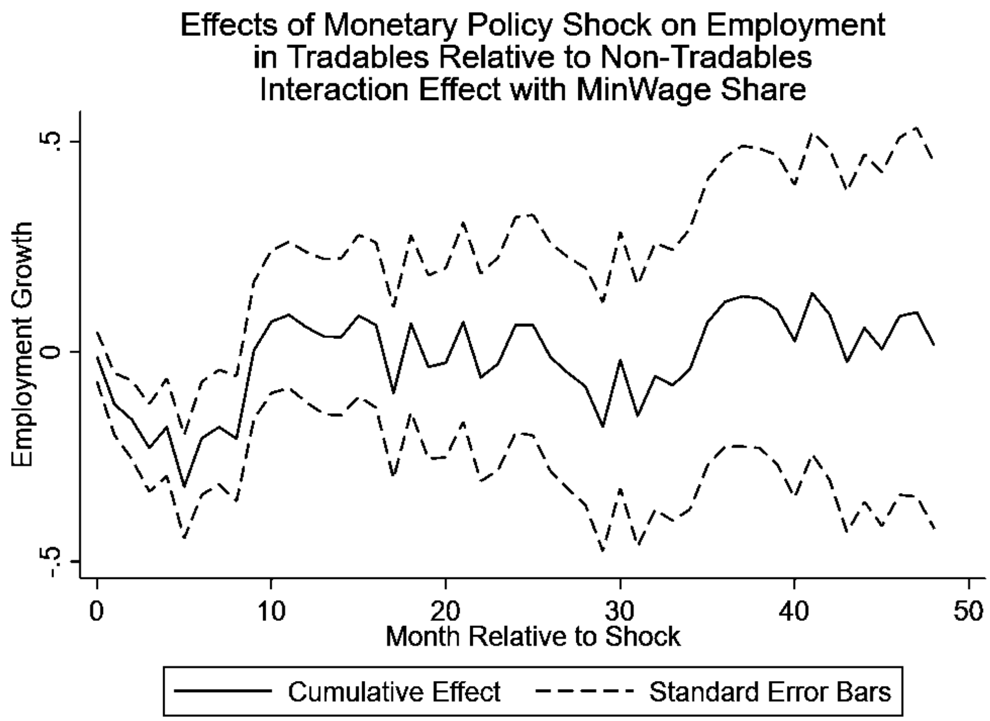
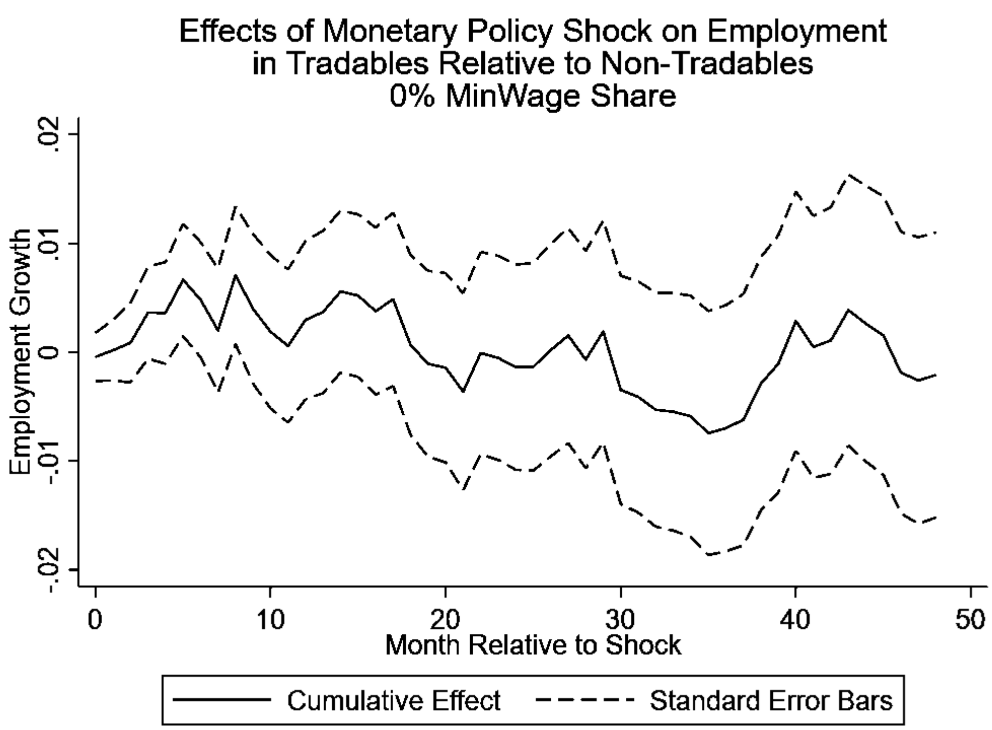


Figure 17: Tradables vs. Non-Tradables Within-State Specification – Level Effect



effects we find on tradables, we might remain concerned that any overall, non-zero-sum stimulus effect is entirely a consequence of the non-tradable sector and therefore the MPC channel. To determine whether this is the case, we compare the effect of a monetary policy shock on tradable versus non-tradable employment *in a 0% minimum-wage share state*. As discussed above, business-stealing would imply positive effects of a contractionary shock on tradable employment relative to non-tradable employment in this setting. As seen in Figure 17, there is no statistically-significant effect on tradable employment relative to non-tradable employment when the minimum wage share is 0%. This suggests that business-stealing effects are of minimal importance here and do not drive our results. This finding agrees with the results of our model, which suggest that stronger within-state input substitution drives larger effects in the tradable sector. Intuitively, there is a somewhat limited role for cross-state substitution when the shock changes all states' real minimum wages simultaneously.

5.3 Comparison of Effect Magnitudes: Model vs. Empirics

At this point, we have presented many different empirical specifications, some yielding differing magnitudes for how the effect of monetary policy on employment varies with the minimum wage share. While most of these magnitudes are statistically indistinguishable, we wish to discuss them in-depth here and compare them to the magnitudes from the model.

The effect size from our model can be read off of Figure 5, Panel 1, which plots the peak employment effect over 4 years in a state as a function of the minimum-wage labor share of total costs in the state. Clearly, the effect on employment of a 1 percentage point increase in the federal funds rate is decreasing and convex in the minimum wage share; it also passes through 0. To get our model's effect size, we can regress, with no constant, the employment data in that plot against a second order polynomial in the minimum-wage labor share of total costs. This yields a maximal effect size when we consider increasing the minimum wage share a small amount from 0: to a first-order, in response to a 1pp increase in the federal funds rate, a state with no minimum wage workers will experience effect a .24 percentage point smaller change in employment than a state with a 0.01 minimum-wage labor share of total costs.

This maximal effect size is four times smaller than the effect size from our baseline regression, displayed in Figure 6, and roughly two times smaller than the effect size from our cross-sectional and Bartik analyses in Figure 7 and Figure 8, though it is within the standard error bands of the latter analyses. It is important to note that the model is measuring the change in employment *hours*, whereas our empirics using the QCEW data are analyzing employment *counts*. This would cause magnitudes to differ to the extent that minimum wage workers work fewer hours on average than higher wage workers. The key driver of this difference in magnitudes, however, is that changes in the minimum wage may increase wages of workers higher up in the wage distribution (see, e.g., Autor et al. 2016); in this sense, the minimum wage shares we used in our model and in our empirical analyses, which focused on workers very close to the minimum wage, may be at times substantially less than the shares of all workers affected by changes in the minimum wage. Unlike in the model, where using smaller shares will lead to monetary policy effects closer to monetary neutrality, using smaller shares in the empirics will lead to effect magnification. We still prefer using shares of workers *near* the minimum wage because that object is much easier to measure than shares of workers *affected* by the minimum; the latter definition requires causally identified estimates of which workers' wages increase when the minimum wage increases.

5.4 Comparison of Effect Magnitudes: Minimum-Wage Channel vs. Overall

So far, we have provided substantial evidence that minimum wages – as one of the key sources of wage rigidity in the modern macroeconomy – are an important channel through which monetary policy is operationalized. We have shown evidence that monetary policy is significantly more effective where the minimum-wage labor share is higher. An important question remains: what fraction of monetary policy's total effect is due to the minimum wage channel? In this section, we provide an answer to this question.

To do so, it is first necessary for us to obtain an estimate of the total effect of monetary policy. An obvious choice is to run a specification directly in line with that of Romer and Romer (2004) and other papers which have utilized narrative monetary shock series:[6]

$$\Delta E_{s,t} = \alpha + \sum_{i=0}^{48} \beta_i Shock_{t-i} + \sum_{i=1}^{48} \eta_i \Delta E_{s,t-i} + \varepsilon_{s,t} \tag{6}$$

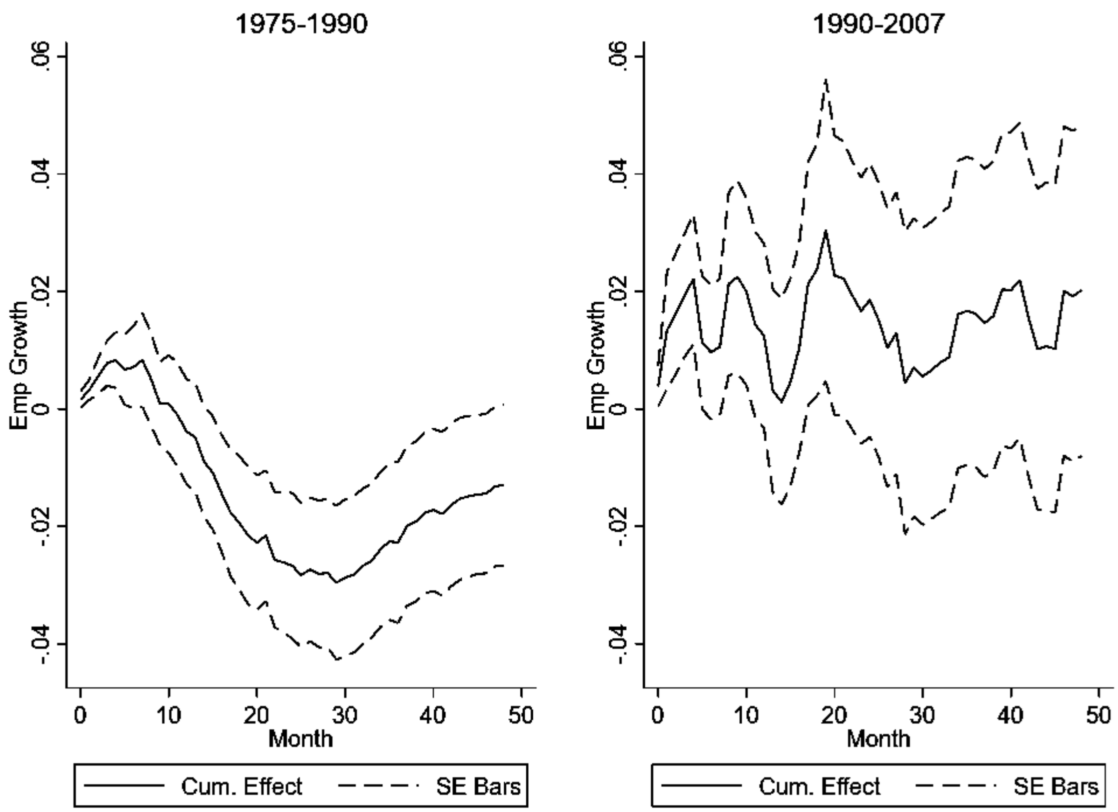
That is, we regress the growth in employment on its lags and the Romer and Romer monetary shock series in order to measure the overall effect of monetary policy on employment. We weight the regression by employment in this context in order to get an accurate measure of the national-level effect.[7] We can then use our data on the (nationwide) minimum-wage labor share along with the results from our regression specifications in Section 5.1 to compute the component of monetary policy's effect originating from the minimum wage channel in a given year. Dividing the latter by the former, we obtain estimates of the fraction of monetary policy's total effect that is due to the minimum wage channel.

There is an issue with using this specification for the purpose of measuring the fraction of monetary policy's total effect that is due to the minimum wage channel. Estimates resulting from

[6] Note also that this specification is equivalent to a version of our baseline specification with the minimum-wage share and interaction effects removed.

[7] We would not want to, for example, weight California and Vermont equally given their sharply different labor-force sizes.

Figure 18: Overall Effects of Monetary Policy on Employment



this regression treat the effect size of monetary policy as a fixed object, one which cannot change over time. In reality, there is some evidence that monetary policy has been less effective since the 1990s. Nakamura and Steinsson (2018) find evidence of a relatively flat Phillips curve over this period. Similarly, Mavroeidis, Plagborg-Møller, and Stock (2014) summarize later papers as typically finding statistically insignificant estimates of the Phillips curve slope; further, they mention how slope estimates in the well-known analysis of Galí and Gertler (1999) fall to insignificance when their sample is extended to include later data. And, indeed, we find the same. ***Figure 18*** displays the results of running the above specification over two non-overlapping samples: 1975-1990 and 1990-2007. As can be seen, monetary policy has a strongly significant effect on employment in the former time period and no discernible effect in the latter. Consequently we focus on the former period.[8] We call 1975-1990 the "Volcker era," despite Volcker's tenure extending from 1979 to 1987, since our regressions use four years of lags. Thus, our regression sample contains outcomes starting in 1979, with lagged regressors extending back to 1975, and concludes with 1990 outcomes using regressors extending back to 1987.

As seen in ***Figure 18***, the peak effect of a 1 percentage-point federal-funds rate shock during the 1975-1990 period is a 2.9 percentage-point reduction in employment. As seen in Figures 6 and 7, the peak interaction effect of a federal-funds rate shock with the minimum-wage labor share is -0.52 in the specification with state and time fixed-effects, -0.76 in the specification with a Bartik control, and -0.88 in the specification with both. Since the average minimum-wage labor share over this period is 2.28 percent, this implies that the minimum-wage channel of monetary policy is responsible for a 1.19, 1.73, or 2.01 percentage-point reduction in employment. That is, over

[8] Because there appears to be no statistically-significant effect of monetary policy in the latter period, attempting to obtain an estimate of the share of the effect driven by the minimum wage channel would constitute dividing by a (statistical) zero.

the 1975-1990 period, our estimates suggest that the minimum-wage channel of monetary policy is responsible for between 41% and 69% of monetary policy's total effect – a non-trivial share.[9]

This finding also suggests a further point. As noted, much of the existing literature – and we ourselves – have found that the overall effectiveness of monetary policy is declining over time. Because a substantial fraction of monetary policy efficacy was due to the minimum wage channel, our evidence suggests that declining real minimum wages – which induce a declining minimum wage labor share – are one important factor behind the reduced efficacy of monetary policy over time.

6 Conclusion

We observe that the standard theoretical and empirical understanding of monetary policy suggests that it should erode real minimum wages. Our model establishes this point formally, providing quantitative predictions about how differences in the minimum wage share across states and time generates heterogeneity in the effects of monetary policy. The model also predicts that our channel of monetary policy should lead to larger changes in low-wage and tradable employment compared to high-wage and non-tradable employment, respectively. To test empirically whether the minimum wage channel is indeed an important channel through which monetary policy is operationalized, we turn to the data. Using QCEW data on employment growth, CPS and BEA data on the share of minimum wage workers by state, and Romer and Romer narrative monetary policy shocks, we find that, indeed, this channel is crucial. This result is robust to a variety of different identification strategies – including a within-state county-level technique

[9] The within-state county-level specification would imply the minimum-wage channel accounts for a much smaller proportion of monetary policy's total effect – approximately 13% – but as argued in Section 5.1, the county-level specification only identifies a subset of the minimum-wage channel of monetary policy.

– and the inclusion of a variety of controls. The relationship also manifests itself using VAR shocks instead of Romer and Romer narrative shocks. It is present in the Canadian data as well. A within-state specification comparing near-minimum-wage to higher-wage employment reveals that the employment growth effect does indeed proceed primarily through near-minimum-wage workers, as expected. And analysis of tradable and non-tradable employment shows that our channel of monetary policy goes more strongly through the tradable sector, as predicted by our model. Our evidence suggests that this rigid-minimum-wage channel of monetary policy accounts for 41 to 69% of monetary policy's total effect (depending on the precise specification used).

Taken as a whole, these findings suggest that minimum wages are an overlooked but important factor in determining the efficacy of monetary policy. Our results imply that monetary policy is minimally effective in the absence of minimum wages. This suggests that, on the one hand, higher minimum wages function as an additional dimension of "policy space" that boosts the ability of monetary policy to stabilize the economy as desired by policymakers. But on the other hand, the Fed – typically thought of as an independent agency – is actually relaxing legislated policies and thereby highly dependent on the political process. In any case, these findings suggest that the interaction effects of monetary policy and minimum wages are a highly understudied topic with substantial room for future exploration.

References

Autor, D. H., Manning, A., and Smith, C. L. (2016). “The Contribution of the Minimum Wage to US Wage Inequality over Three Decades: A Reassessment.” *American Economic Journal: Applied Economics* 8(1): 58–99.

Barro, R. J., and Grossman, H. I. (1971). “A General Disequilibrium Model of Income and Employment.” *American Economic Review* 61(1): 82–93.

Basu, S., and House, C. L. (2016) “Allocative and Remitted Wages: New Facts and Challenges for Keynesian Models.” In *Handbook of Macroeconomics*, edited by John B. Taylor and Harald Uhlig, 2: 297–354. Elsevier.

Bernanke, B. S., and Blinder, A. S. (1992). “The Federal Funds Rate and the Channels of Monetary Transmission.” *American Economic Review* 82(4): 901–921.

Champagne, J, and Sekkel, R. (2018). “Changes in Monetary Regimes and the Identification of Monetary Policy Shocks: Narrative Evidence from Canada.” *Journal of Monetary Economics* 99: 72–87.

Chetty, R. (2006) “A New Method of Estimating Risk Aversion.” *American Economic Review* 96(5): 1921-1834.

Chodorow-Reich, G., Nenov, P. T., and Simsek, A. (2020). “Stock Market Wealth and the Real Economy: A Local Labor Market Approach,” Working Paper.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). “Monetary Policy Shocks: What Have We Learned and to What End?” In *Handbook of Macroeconomics*, 1:65–148. Elsevier.

Cochrane, J. H., and Piazzesi, M. (2002). “The Fed and Interest Rates - A High-Frequency Identification.” *American Economic Review* 92(2): 90–95.

Coibion, O. (2012). “Are the Effects of Monetary Policy Shocks Big or Small?” *American Economic Journal: Macroeconomics* 4(2): 1–32.

Coibion, O., Gorodnichenko, Y., Kueng, L., and Silvia, J. (2017). “Innocent Bystanders? Monetary Policy and Inequality.” *Journal of Monetary Economics* 88: 70–89.

Cook, T., and Hahn, T. (1989). “The Effect of Changes in the Federal Funds Rate Target on Market Interest Rates in the 1970s.” *Journal of Monetary Economics* 24(3): 331–351.

Daly, M. C., and Hobijn, B. (2014). “Downward Nominal Wage Rigidities Bend the Phillips Curve.” *Journal of Money, Credit and Banking* 46(S2): 51–93.

Daly, M. C., and Hobijn, B. (2015). “Why Is Wage Growth so Slow?” *FRBSF Economic Letter*

2015-01.

Daly, M. C., Hobijn, B., and Lucking, B. (2012). "Why Has Wage Growth Stayed Strong?" *FRBSF Economic Letter* 2012-10.

Devereux, P. J., and Altonji, J. G. (1999). "The Extent and Consequences of Downward Nominal Wage Rigidity." NBER Working Paper No. 7236.

Elsby, M. W., and Solon, G. (2019). "How Prevalent Is Downward Rigidity in Nominal Wages? International Evidence from Payroll Records and Pay Slips." *The Journal of Economic Perspectives* 33(3): 185–201.

Fallick, B., Villar, D., and Wascher, W. L. (2020). "Downward Nominal Wage Rigidity in the United States During and After the Great Recession." Federal Reserve Bank of Cleveland Working Paper No. 16-02R.

Galí, J., and Gertler, M. (1999). "Inflation Dynamics: A Structural Econometric Analysis." *Journal of Monetary Economics* 44(2): 195–222.

Gertler, M., and Karadi, P. (2015). "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7(1): 44–76.

Gürkaynak, R. S., Sack, B., and Swanson, E. (2005). "The Sensitivity of Long-Term Interest Rates to Economic News: Evidence and Implications for Macroeconomic Models." *American Economic Review* 95(1): 425–36.

Jardim, E., Solon, G., and Vigdor, J. (2019). "How Prevalent Is Downward Rigidity in Nominal Wages? Evidence from Payroll Records in Washington State." NBER Working Paper No. 25470.

Kahn, S. (1997). "Evidence of Nominal Wage Stickiness from Microdata." *American Economic Review* 87(5): 993–1008.

Katz, L. F., and Murphy, K. M. (1992). "Changes in Relative Wages, 1963-1987: Supply and Demand Factors." *Quarterly Journal of Economics* 107(1): 35–78.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*. New York, NY: Harcourt, Brace, and Company, 1936.

Krusell, P., Ohanian, L. E., Ríos-Rull, J., and Violante, G. L. (2000). "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68(5): 1029–1053.

Kuttner, K. N. (2001). "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics* 47(3): 523–544.

Lebow, D., Saks, R., and Wilson, B. (2003). "Downward Nominal Wage Rigidity: Evidence from the Employment Cost Index." *Advances in Macroeconomics* 3(1): 1–30.

Leeper, E., Sims, C., and Zha, T. (1996). "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* 27(2): 1–78.

Mavroeidis, S., Plagborg-Møller, M., and Stock, J. H. (2014). "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52(1): 124–188.

McLaughlin, K. J. "Rigid Wages?" (1994). *Journal of Monetary Economics* 34(3): 383–414.

Mian, A., Rao, K., and Sufi, A. (2013). "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics* 128(4): 1687–1726.

Mian, A., and Sufi, A. "What Explains the 2007–2009 Drop in Employment?" *Econometrica* 82(6): 2197–2223.

Nakamura, E., and Steinsson, J. (2018). "High-Frequency Identification of Monetary Non-Neutrality: The Information Effect." *Quarterly Journal of Economics* 133(3): 1283–1330.

Oberfield, E., and Raval, D. (2020). "Micro Data and Macro Technology." Working Paper.

Reich, M., Allegretto, S., and Godoy, A. (2017). "Seattle's Minimum Wage Experience 2015-16." Working Paper.

Romer, C. D., and Romer, D. H. (2004). "A New Measure of Monetary Shocks: Derivation and Implications." *American Economic Review* 94(4): 1055–1084.

Romer, C. D, and Romer, D. H. (2010). "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100(3): 763–801.

Solon, G., Barsky, R., and Parker, J. A. (1994). "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias." *Quarterly Journal of Economics* 109(1): 1–25.

Vaghul, K., and Zipperer, B. (2016). "Historical State and Sub-State Minimum Wage Data." Washington Center for Equitable Growth, 2016.

Whalen, C., and Reichling, F. (2017). "Estimates of the Frisch Elasticity of Labor Supply: A Review." *Eastern Economic Journal* 43(1): 37–42.

Wilson, B. A. (1999). "Wage Rigidity: A Look inside the Firm." Federal Reserve Finance and Economics Discussion Paper Series, May 1999.

Appendix A: Appendix to Chapter 1

A.1 Proofs

Proof of Proposition 1: First, substituting the equation for children's preferences into the utility function,

$$u_{i,t}(x_{i,t}) = -(b_{i,t} - x_{i,t})^2 - \alpha(b_{i,t} - \gamma x_{i,t} - (1-\gamma)L)^2 .$$

Differentiating the utility function and setting the result equal to zero in order to find a maximum,

$$\begin{aligned}
&\partial u_{i,t} \Big/ \partial x_{i,t} = -2x_{i,t} + 2b_{i,t} - 2\alpha\gamma\left(\gamma x_{i,t} + (1-\gamma)L - b_{i,t}\right) = 0 \\
&\Rightarrow x_{i,t}(1+\alpha\gamma^2) = b_{i,t}(1+\alpha\gamma) - L(\alpha\gamma(1-\gamma)) \\
&\Rightarrow x_{i,t}^* = \frac{1+\alpha\gamma}{1+\alpha\gamma^2} b_{i,t} - \frac{\alpha\gamma(1-\gamma)}{1+\alpha\gamma^2} L.
\end{aligned}$$

Note that the coefficient on $b_{i,t}$ is positive and the coefficient on L is negative. Consequently, $\partial x_{i,t}^* \big/ \partial b_{i,t} > 0,\ \partial x_{i,t}^* \big/ \partial L < 0$.

Proof of Proposition 2: From the proof of Proposition 1, we have that

$$\begin{aligned}
x_{i,t+1} &= \frac{1+\alpha\gamma}{1+\alpha\gamma^2} b_{i,t+1} - \frac{\alpha\gamma(1-\gamma)}{1+\alpha\gamma^2} L \\
&= \frac{1+\alpha\gamma}{1+\alpha\gamma^2}\left(\gamma x_{i,t} + (1-\gamma)L\right) - \frac{\alpha\gamma(1-\gamma)}{1+\alpha\gamma^2} L \\
&= \gamma\left(\frac{1+\alpha\gamma}{1+\alpha\gamma^2}\right)^2 b_{i,t} - \gamma\left(\frac{1+\alpha\gamma}{1+\alpha\gamma^2}\right)\left(\frac{\alpha\gamma(1-\gamma)}{1+\alpha\gamma^2}\right)L + \frac{(1-\gamma)(1+\alpha\gamma)}{1+\alpha\gamma^2} L - \frac{\alpha\gamma(1-\gamma)}{1+\alpha\gamma^2} L \\
&= \gamma\left(\frac{1+\alpha\gamma}{1+\alpha\gamma^2}\right)^2 b_{i,t} - \frac{\gamma(1+\alpha\gamma)\left(\alpha\gamma(1-\gamma)\right) + \left(\alpha\gamma(1-\gamma)\right)(1+\alpha\gamma^2) - (1-\gamma)(1+\alpha\gamma)(1+\alpha\gamma^2)}{1+\alpha\gamma^2} L.
\end{aligned}$$

Consequently,

$$
\begin{aligned}
&\partial x^*_{i,t+1} \big/ \partial L < 0 \\
&\Leftrightarrow \gamma(1+\alpha\gamma)\big(\alpha\gamma(1-\gamma)\big) + \big(\alpha\gamma(1-\gamma)\big)(1+\alpha\gamma^2) - (1-\gamma)(1+\alpha\gamma)(1+\alpha\gamma^2) > 0 \\
&\Leftrightarrow \gamma(1+\alpha\gamma)\alpha\gamma + \alpha\gamma(1+\alpha\gamma^2) - (1+\alpha\gamma)(1+\alpha\gamma^2) > 0 \\
&\Leftrightarrow \gamma(1+\alpha\gamma)\alpha\gamma + \alpha^2\gamma^3 - 1 - \alpha\gamma^2 - \alpha^2\gamma^3 > 0 \\
&\Leftrightarrow \alpha^2\gamma^3 - 1 \\
&\Leftrightarrow \alpha > \sqrt{\gamma} \big/ \gamma^2 .
\end{aligned}
$$

Proof of Proposition 3: Differentiating the utility function,

$$
\begin{aligned}
&\partial u_{i,t} \big/ \partial x_{i,t} = -2(x_{i,t} - b_{i,t}) - 2\alpha\gamma_P(\gamma_P x_{i,t} + \gamma_N \overline{x}_t + \gamma_L L - b_{i,t}) = 0 \\
&\Rightarrow (1+\alpha\gamma_P^2) x_{i,t} = (1+\alpha\gamma_P) b_{i,t} - \alpha\gamma_P\gamma_N \overline{x}_t - \alpha\gamma_P\gamma_L L \\
&\Rightarrow x_{i,t} = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} b_{i,t} - \frac{\alpha\gamma_P\gamma_L}{1+\alpha\gamma_P^2} L - \frac{\alpha\gamma_P\gamma_N}{1+\alpha\gamma_P^2} \frac{1}{N} \sum\nolimits_j x_{j,t}.
\end{aligned}
$$

This is not yet a closed-form solution, as $x_{j,t}$ itself depends on $b_{i,t}$, $b_{j,t}$, and L. So,

$$
\begin{aligned}
&\frac{\partial x_{i,t}}{\partial x_{k,t}} = -\underbrace{\frac{\alpha\gamma_P\gamma_N}{1+\alpha\gamma_P^2}}_{\equiv A} \frac{1}{N} \left(1 + (N-1) \frac{\partial x_{j,t}}{\partial x_{k,t}} \right) \\
&\frac{\partial x_{i,t}}{\partial x_{k,t}} \left(1 + \frac{N-1}{N} A \right) = -\frac{1}{N} A \\
&\frac{\partial x_{i,t}}{\partial x_{k,t}} = \frac{-\frac{1}{N} A}{\frac{N+(N-1)A}{N}} = -\frac{A}{N+(N-1)A}.
\end{aligned}
$$

Using this result to derive the crucial comparative statics,

$$
\begin{aligned}
&\frac{\partial x^*_{i,t}}{\partial b_{i,t}} = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} - A \frac{\partial x_{j,t}}{\partial x_{i,t}} \frac{\partial x_{i,t}}{\partial b_{i,t}} \\
&\frac{\partial x^*_{i,t}}{\partial b_{i,t}} \left(1 - \frac{A^2}{N+(N-1)A} \right) = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} \\
&\frac{\partial x^*_{i,t}}{\partial b_{i,t}} \left(\frac{N+(N-1)A - A^2}{N+(N-1)A} \right) = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} \\
&\frac{\partial x^*_{i,t}}{\partial b_{i,t}} = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} \frac{N+(N-1)A}{N+(N-1)A-A^2} > 0.
\end{aligned}
$$

And we have

$$\frac{\partial x_{i,t}^*}{\partial L} = -\frac{\alpha\gamma_P\gamma_L}{1+\alpha\gamma_P^2} - A\left(\frac{\partial x_{i,t}}{\partial L} + \frac{\partial x_{j,t}}{\partial x_{i,t}}\frac{\partial x_{i,t}}{\partial L}\right)$$

$$\frac{\partial x_{i,t}^*}{\partial L}\left(1 + A - \frac{A^2}{N+(N-1)A}\right) = -\frac{\alpha\gamma_P\gamma_L}{1+\alpha\gamma_P^2}$$

$$\frac{\partial x_{i,t}^*}{\partial L}\left(\frac{N+(2N-1)A+(N-2)A^2}{N+(N-1)A}\right) = -\frac{\alpha\gamma_P\gamma_L}{1+\alpha\gamma_P^2}$$

$$\frac{\partial x_{i,t}^*}{\partial L} = -\frac{\alpha\gamma_P\gamma_L}{1+\alpha\gamma_P^2}\frac{N+(N-1)A}{N+(2N-1)A+(N-2)A^2} < 0.$$

Lemma 1: $b_{j,t} = -b_{i,t} \wedge L = 0 \Rightarrow x_{j,t}^* = -x_{i,t}^*$

By definition of the optimum, we know that the utility of family *i* is maximized at $x_{i,t}^*$. That is,

$$-(y-b_{i,t})^2 - \alpha(\gamma y - b_{i,t})^2 - py^2$$

is maximized at $y = x_{i,t}^*$. Consider now the utility of family *j*. Since $b_{j,t} = -b_{i,t}$, we have

$$\begin{aligned}&-(x_{j,t}-b_{j,t})^2 - \alpha(\gamma x_{j,t} - b_{j,t})^2 - p(x_{j,t})^2\\ &= -(x_{j,t}+b_{i,t})^2 - \alpha(\gamma x_{j,t} + b_{i,t})^2 - p(x_{j,t})^2\\ &= -(-x_{j,t}-b_{i,t})^2 - \alpha(-\gamma x_{j,t} - b_{i,t})^2 - p(-x_{j,t})^2.\end{aligned}$$

So, this expression must be maximized at $-x_{j,t}^* = x_{i,t}^*$. That is, $x_{j,t}^* = -x_{i,t}^*$.

Proof of Proposition 4: It is without loss of generality to set $L = 0$ here since the bliss points of both types of families are defined relative to L.

Case (a): Homogeneous society – N families with ideology $b_{i,t}$

Since all families have the same underlying ideological preference, their problems are symmetric, and they will all have the same optimal action. In other words, the solution from Proposition 3 simplifies to

$$x_{i,t} = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} b_{i,t} - \frac{\alpha\gamma_P\gamma_N}{1+\alpha\gamma_P^2} x_{i,t}$$

$$x_{i,t}\left(\underbrace{1+\frac{\alpha\gamma_P\gamma_N}{1+\alpha\gamma_P^2}}_{=B}\right) = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} b_{i,t}$$

$$x_{i,t}^{hom,*} = \frac{1}{B}\frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} b_{i,t}.$$

Case (b): Heterogeneous society – $N/2$ families with ideology $b_{i,t}$, $N/2$ families with ideology $-b_{i,t}$

By Lemma 1, in this society we will have $\bar{x}_t = \sum_j x_{j,t} = \frac{N}{2} x_{i,t} + \frac{N}{2}(-x_{i,t}) = 0$. As such, the above solution simplifies to

$$x_{i,t}^{het,*} = \frac{1+\alpha\gamma_P}{1+\alpha\gamma_P^2} b_{i,t}.$$

Consider the persistence of actions into future generations. By definition of $b_{i,t}$, we have

$$x_{i,t+k}^{hom,*} = \frac{1}{B}\frac{1+\alpha\gamma_P+\alpha\gamma_N}{1+\alpha\gamma_P^2} b_{i,t+k} = \frac{1}{B}\frac{1+\alpha\gamma_P+\alpha\gamma_N}{1+\alpha\gamma_P^2}(\gamma_P+\gamma_N)x_{i,t+k-1} = \left[\frac{1}{B}\frac{1+\alpha\gamma_P+\alpha\gamma_N}{1+\alpha\gamma_P^2}(\gamma_P+\gamma_N)\right]^k x_{i,t},$$

$$x_{i,t+k}^{het,*} = \frac{1+\alpha\gamma_P+\alpha\gamma_N}{1+\alpha\gamma_P^2} b_{i,t+k} = \frac{1+\alpha\gamma_P+\alpha\gamma_N}{1+\alpha\gamma_P^2}\gamma_P x_{i,t+k-1} = \left[\frac{1+\alpha\gamma_P+\alpha\gamma_N}{1+\alpha\gamma_P^2}\gamma_P\right]^k x_{i,t}.$$

Consequently, if $(\gamma_P+\gamma_N)/B > \gamma_P$, then $x_{i,t+k}^{het,*}$ will be the closer to 0 of the two expressions for sufficiently large k. Observe that

$$\begin{aligned}
&\frac{(\gamma_P+\gamma_N)(1+\alpha\gamma_P^2)}{(1+\alpha\gamma_P^2+\alpha\gamma_P\gamma_N)} - \gamma_P \\
&= \frac{\gamma_P+\alpha\gamma_P^3+\gamma_N+\alpha\gamma_P^2\gamma_N-\gamma_P-\alpha\gamma_P^3-\alpha\gamma_P^2\gamma_N}{1+\alpha\gamma_P^2+\alpha\gamma_P\gamma_N} \\
&= \frac{\gamma_N}{1+\alpha\gamma_P^2+\alpha\gamma_P\gamma_N} > 0.
\end{aligned}$$

So, indeed, $|x_{i,t+k}^{hom,*} - L| > |x_{i,t+k}^{het,*} - L|$ for sufficiently large k.

Proof of Proposition 5: Differentiating the utility function of a given parent *i*,

$$\partial u_{i,t} \big/ \partial x_{i,t} = -\omega_i(x_{i,t} - b_{i,t}) - 2\alpha\gamma_i\omega_i(\gamma_i x_{i,t} + \gamma_j x_{j,t} + \gamma_L L_t - b_{i,t}) - 2p(x_{i,t} - L) = 0$$

$$= b_{i,t}(\omega_i + \alpha\gamma_i\omega_i) - x_{i,t}(\omega_i + \alpha\gamma_i^2\omega_i + p) - L_t(\alpha\gamma_i\gamma_L\omega_i - p) - x_{j,t}(\alpha\gamma_i\gamma_j\omega_i) = 0$$

$$\Rightarrow x_{i,t} = \omega_i \frac{1+\alpha\gamma_i}{\omega_i + \alpha\gamma_i^2\omega_i + p} b_{i,t} - \frac{\alpha\gamma_i\gamma_L\omega_i - p}{\omega_i + \alpha\gamma_i^2\omega_i + p} L - \frac{\alpha\gamma_i\gamma_j\omega_i}{\omega_i + \alpha\gamma_i^2\omega_i + p} x_{j,t}.$$

This is not yet a closed-form solution, as $x_{j,t}$ itself depends on $b_{j,t}$, L, and $x_{i,t}$. So, substituting the parallel expression for $x_{j,t}$ into the above expression yields

$$x_{i,t} = \frac{1}{C}\omega_i \frac{1+\alpha\gamma_i}{\omega_i + \alpha\gamma_i^2\omega_i + p} b_{i,t} - \frac{1}{C}\frac{\alpha\gamma_i\gamma_j\omega_i}{\omega_i + \alpha\gamma_i^2\omega_i + p}\frac{1+\alpha\gamma_j}{\omega_j + \alpha\gamma_j^2\omega_j + p} b_{j,t}$$
$$-\frac{1}{C}\left(\frac{\alpha\gamma_i\gamma_L\omega_i - p}{\omega_i + \alpha\gamma_i^2\omega_i + p} - \frac{\alpha\gamma_i\gamma_j\omega_i}{\omega_i + \alpha\gamma_i^2\omega_i + p}\frac{\alpha\gamma_j\gamma_L\omega_j - p}{\omega_j + \alpha\gamma_j^2\omega_j + p}\right)L,$$

where $C \equiv 1 - \frac{\alpha\gamma_i\gamma_j\omega_i}{\omega_i + \alpha\gamma_i^2\omega_i + p}\frac{\alpha\gamma_j\gamma_i\omega_j}{\omega_j + \alpha\gamma_j^2\omega_j + p}$.

That is,

$$\frac{\partial x_{i,t}^*}{\partial L} = -\frac{1}{C}\left(\frac{\alpha\gamma_i\gamma_L\omega_i - p}{\omega_i + \alpha\gamma_i^2\omega_i + p} - \frac{\alpha\gamma_i\gamma_j\omega_i}{\omega_i + \alpha\gamma_i^2\omega_i + p}\frac{\alpha\gamma_j\gamma_L\omega_j - p}{\omega_j + \alpha\gamma_j^2\omega_j + p}\right).$$

A few specific cases merit highlighting. First, consider the case where both parents are equal in all dimensions ($\omega_i = \omega_j$, $\gamma_i = \gamma_j$). In this case, because $\alpha\gamma_i^2\omega \big/ (\omega_i + \alpha\gamma_i^2\omega_i + p) < 1$, the above expression is unambiguously negative provided α, $\gamma_L > 0$ and both parents undergo backlash. Next, consider the case where one parent, *j*, has no influence in inculcating his/her child ideologically, $\gamma_j = 0$. This zeroes out the second term within the parentheses in the above expression, and thus as long as *p* is sufficiently small and α, ω_i, γ_i, $\gamma_L > 0$, parent *i* will undergo backlash while parent *j* will not.

$$\frac{\partial x_{i,t}^{*}}{\partial L}=-\frac{\alpha\gamma_i\gamma_L\omega_i-p}{\omega_i+\alpha\gamma_i^2\omega_i+p}<0,\ \frac{\partial x_{j,t}^{*}}{\partial L}=1$$

Similarly, consider the case where one parent, *j*, does not care about these ideological matters, $\omega_j = 0$. This leads to the same solution as above, and provided α, ω_i, γ_i, $\gamma_L > 0$, once again parent *i* will undergo backlash while parent *j* will not.

A.2 Extension: Endogenized Laws, Voting

Given that backlash is systematic, will any laws ever be passed in the first place? In order to answer this question, it is possible to fully endogenize the passage of laws. Consider a scenario where families, at the start of each generation, vote on changing the law in a referendum. They are given the choice between re-affirming the law that was in effect in the previous generation or replacing it with a law corresponding to the bliss point of the median voter, b_{median}. Families have the following utility function – a slight adaptation of the baseline utility function:

$$u_{i,t}(x_{i,t})=-(x_{i,t}-b_{i,t})^2-\alpha(b_{i,t+1}-b_{i,t})^2-\sum_j\beta(x_{j,t}-b_{i,t})^2$$

Ideological preferences are formed as in the baseline case. A third term is added to the utility function to indicate that families care about the extent to which *other* families take actions close to their preferences. For example, conservative families wish others behaved in a manner consistent with conservative ideology and liberal families wish others behaved in a manner consistent with liberal ideology. Note that we could think of this new third term as having been present in the baseline utility function as well – there it would have been a constant, as individuals had no influence over the contemporaneous actions of other families. Here, because changing the law changes the actions of families, such an influence does exist.

It can be shown that, indeed, despite the existence of backlash, as long as families are sufficiently forward-looking, in equilibrium they will vote for laws that are close to their bliss point in order to move society (and future generations of their family) toward the law. They will tolerate the short-term backlash in order to attain long-term convergence. If families are not forward-looking and care disproportionately about the present and near future, the law will not be changed in equilibrium.

Proposition A: For α sufficiently high, the existing law will be replaced in a majority vote with the new law, $L_{new} = b_{median}$.

Proof: First, note that once the law is chosen, the problem faced by families here is identical to that in the baseline case; families' actions do not affect the value of the third term. As such, the optimal action is identical.

Thus, in order to decide how to vote, each family will assess their utility under the existing law. Denote by C_b and C_L the coefficients on $b_{i,t}$ and L, respectively, in the solution for the optimal action, and note from the proof of Proposition 1 that $C_b + C_L = 1$.

$$
\begin{aligned}
u(x^*(L)) &= -(x^*_{i,t} - b_{i,t})^2 - \alpha(\gamma x^*_{i,t} + (1-\gamma)L - b_{i,t})^2 - \beta \sum\nolimits_j (x^*_{j,t} - b_{i,t})^2 \\
&= -\left(C_b b_{i,t} + C_L L - b_{i,t}\right)^2 - \alpha\left(\gamma C_b b_{i,t} + \gamma C_L L + (1-\gamma)L - b_{i,t}\right)^2 - \beta \sum\nolimits_j (C_b b_{j,t} + C_L L - b_{i,t})^2 \\
&= -\left((C_b - 1)b_{i,t} + C_L L\right)^2 - \alpha\left((-1+\gamma C_b)b_{i,t} + \left((1-\gamma)+\gamma C_L\right)L\right)^2 - \beta \sum\nolimits_j (C_b b_{j,t} + (1-C_b)L - b_{i,t})^2 \\
&= -\left((1-C_b)b_{i,t} - C_L L\right) - \alpha\left((1-\gamma C_b)b_{i,t} - (1-\gamma C_b)L\right)^2 - \beta \sum\nolimits_j \left((b_{i,t} - L) - C_b(b_{j,t} - L)\right)^2 \\
&= -\left(C_L b_{i,t} - C_L L\right) - \alpha\left((1-\gamma C_b)b_{i,t} - (1-\gamma C_b)L\right)^2 - \beta \sum\nolimits_j \left((b_{i,t} - L) - C_b(b_{j,t} - L)\right)^2.
\end{aligned}
$$

They will compare this to their utility under the new law and will vote for the new law if it provides higher utility. For the median family and all families *further* from the pre-existing law than the median voter,

$$
\begin{aligned}
& u(x^*(L_{new})) > u(x^*(L)) \\
& \Leftrightarrow -\left(C_L b_{i,t} - C_L L_{new}\right)^2 - \alpha\left((1-\gamma C_b)b_{i,t} - (1-\gamma C_b)L_{new}\right)^2 - \beta\sum_j\left((b_{i,t}-L_{new}) - C_b(b_{j,t}-L_{new})\right)^2 \\
& \qquad > -\left(C_L b_{i,t} - C_L L\right)^2 - \alpha\left((1-\gamma C_b)b_{i,t} - (1-\gamma C_b)L\right)^2 - \beta\sum_j\left((b_{i,t}-L) - C_b(b_{j,t}-L)\right)^2 \\
& \Leftrightarrow \left(C_L^2 + \alpha(1-\gamma C_b)^2\right)(b_{i,t}-L_{new})^2 + \beta\sum_j\left((b_{i,t}-L_{new})^2 - 2C_b(b_{i,t}-L_{new})(b_{j,t}-L_{new}) + C_b^2(b_{j,t}-L_{new})^2\right) \\
& \qquad < \left(C_L^2 + \alpha(1-\gamma C_b)^2\right)(b_{i,t}-L)^2 + \beta\sum_j\left((b_{i,t}-L)^2 - 2C_b(b_{i,t}-L)(b_{j,t}-L) + C_b^2(b_{j,t}-L)^2\right) \\
& \Leftrightarrow \left(C_L^2 + \alpha(1-\gamma C_b)^2\right)\left((b_{i,t}-L_{new})^2 - (b_{i,t}-L)^2\right) \\
& \qquad < \beta\sum_j\left(\left((b_{i,t}-L) - C_b(b_{j,t}-L)\right)^2 - \left((b_{i,t}-L_{new}) - C_b(b_{j,t}-L_{new})\right)^2\right) \\
& \Leftrightarrow \alpha > \frac{\beta\sum_j\left(\left((b_{i,t}-L) - C_b(b_{j,t}-L)\right)^2 - \left((b_{i,t}-L_{new}) - C_b(b_{j,t}-L_{new})\right)^2\right)}{\left((b_{i,t}-L_{new})^2 - (b_{i,t}-L)^2\right)(1-\gamma C_b)^2} - \frac{C_L^2}{(1-\gamma C_b)^2}.
\end{aligned}
$$

Note that the inequality is flipped in the last line above because $(b_{i,t} - L_{new})^2 - (b_{i,t} - L)^2 < 0$ for the median family and all families further from the pre-existing law than the median family. Since the above inequality specifies the value of α needed for a given family to vote for the new law, there must exist some value of α satisfying the inequality for the majority of families – i.e., a value of α sufficient for the new law to pass.

A.3 Extension: Endogenized Laws, Backlash

Consider an extension to the baseline model whereby the actions families take influence what the law will be in the next period. Families also obtain disutility from the sheer existence of laws which are far from their own ideological preferences. That is,

$$u_{i,t}(x_{i,t}) = -(x_{i,t} - b_{i,t})^2 - \alpha(b_{i,t+1} - b_{i,t})^2 - \mu(L_{t+1} - b_{i,t})^2$$

Ideological preferences are formed as before, $b_{i,t+1} = \gamma x_{i,t} + (1-\gamma)L_t$, but the law is now determined similarly by a weighted average of the public's actions and the law itself in the preceding period: $L_{t+1} = \pi\bar{x}_t + (1-\pi)L_t$. Note that we could again think of the third term of the utility function as having been present in the baseline version as well. There, however, it would have been a constant since the law was exogenous. Similarly, families might care about the distance of the law from their preferences during the present generation, $(L_t - b_{i,t})^2$, but this too would be a constant and will thus fall out of the function during maximization.

As before, to maximize utility, we differentiate the utility function with respect to $x_{i,t}$.

$$\begin{aligned}
\partial u \big/ \partial x_{i,t} &= -x_{i,t} + b_{i,t} + \alpha\gamma\left(-\gamma x_{i,t} + (1-\gamma)L_t + b_{i,t}\right) + (\mu\pi/N)\left(-\pi\bar{x}_t - (1-\pi)L_t + b_{i,t}\right) = 0 \\
&= b_{i,t}(1+\alpha\gamma+\mu\pi/N) - x_{i,t}(1+\alpha\gamma^2) - L_t\left(\alpha\gamma(1-\gamma) + (\mu\pi/N)(1-\pi)\right) - \bar{x}_t\,\mu\pi^2/N = 0 \\
&= b_{i,t}(1+\alpha\gamma+\mu\pi/N) - x_{i,t}(1+\alpha\gamma^2+\mu\pi^2/N^2) \\
&\quad - L_t\left(\alpha\gamma(1-\gamma) + (\mu\pi/N)(1-\pi)\right) - \sum\nolimits_{j\neq i} x_{j,t}\,\mu\pi^2/N^2 = 0 \\
\Rightarrow x_{i,t} &= \frac{1+\alpha\gamma+\mu\pi/N}{1+\alpha\gamma^2+\mu\pi^2/N^2} b_{i,t} - \frac{\alpha\gamma(1-\gamma)+(\mu\pi/N)(1-\pi)}{1+\alpha\gamma^2+\mu\pi^2/N^2} L_t - \frac{\mu\pi^2}{N^2}\sum\nolimits_{j\neq i} x_{j,t}.
\end{aligned}$$

Observe that this is not yet a closed-form solution – $x_{j,t}$ remains on the right-hand-side.

$$\begin{aligned}
&\frac{\partial x_{i,t}}{\partial x_{k,t}} = -\frac{\mu\pi^2}{N^2} - \frac{\mu\pi^2}{N^2}\sum\nolimits_{j\neq i,k}\frac{\partial x_{j,t}}{\partial x_{k,t}} \\
&\frac{\partial x_{i,t}}{\partial x_{k,t}}\left(1+\mu\pi^2\frac{N+2}{N^2}\right) = -\frac{\mu\pi^2}{N^2} \\
&\frac{\partial x_{i,t}}{\partial x_{k,t}} = \frac{-\mu\pi^2}{\mu\pi^2(N+2)+N^2} < 0.
\end{aligned}$$

Now it is possible to compute $\partial x_{i,t}^* \big/ \partial L_t$ in order to make study the extent of the backlash,

$$\begin{aligned}
&\frac{\partial x_{i,t}}{\partial L_t} = -\frac{\alpha\gamma(1-\gamma)+(\mu\pi/N)(1-\pi)}{1+\alpha\gamma^2+\mu\pi^2/N^2} - \frac{\mu\pi^2}{N^2}\sum\nolimits_{j\neq i}\left(\frac{\partial x_{j,t}}{\partial L_t} + \frac{\partial x_{j,t}}{\partial x_{i,t}}\frac{\partial x_{i,t}}{\partial L_t}\right) \\
&\frac{\partial x_{i,t}}{\partial L_t} = -\frac{\alpha\gamma(1-\gamma)+(\mu\pi/N)(1-\pi)}{1+\alpha\gamma^2+\mu\pi^2/N^2} - \frac{\mu\pi^2(N+1)}{N^2}\left(\frac{\partial x_{j,t}}{\partial L_t} - \frac{\mu\pi^2}{\mu\pi^2(N+2)+N^2}\frac{\partial x_{i,t}}{\partial L_t}\right) \\
&\frac{\partial x_{i,t}}{\partial L_t}\left(\underbrace{1+\frac{\mu\pi^2(N+1)}{N^2}\left(1-\frac{\mu\pi^2}{\mu\pi^2(N+2)+N^2}\right)}_{\equiv B}\right) = -\frac{\alpha\gamma(1-\gamma)+(\mu\pi/N)(1-\pi)}{1+\alpha\gamma^2+\mu\pi^2/N^2} \\
&\frac{\partial x_{i,t}}{\partial L_t} = -\frac{\alpha\gamma(1-\gamma)+(\mu/N)\pi(1-\pi)}{B\left(1+\alpha\gamma^2+\mu\pi^2/N^2\right)} < 0.
\end{aligned}$$

Here we see that the extent of the backlash is increasing in both the extent to which families care about their children's preferences, α, and the extent to which families care about the law being consonant with their own preferences, μ. Importantly, however, it is decreasing in N. Because one individual in a large society can only contribute a small amount to changing the law, the ability to change the law contributes little to the inducement of backlash relative to the ability to influence one's children. For example, consider a case where $\pi = \gamma$ and $\alpha = \mu$ for simplicity. In such a case, in a society of one million, the inducement to backlash provided by the inculcation-of-children channel is *one million times* the inducement provided by the change-the-law channel.

A.4 Additional Robustness Checks

The surpising richness of the 1970s-era American National Election Studies and other contemporaneous survey datasets on women's issues allow for additional exploration of the state ERAs and the backlash they induced. In the first part of this appendix, I conduct additional robustness checks on the main result – backlash in terms of male attitudes. In the second part, I explore additional material outcomes – male and female fertility preferences, marital discontent, and women's economic outcomes – presenting some evidence that the ERA backlash had effects along these margins as well.

Figure A-1 modifies the main dynamic difference-in-differences specification. Instead of pooling all periods more than 10 years after ERA passage into one "long-run" indicator variable, it separates them into a multitude of indicators, the last of which ends 4 decades after ERA passage. This specification is responsive to the finding of Borusyak and Jaravel (2017) that pooling many periods into one "long-run" term – even in a dynamic difference-in-differences specification – may bias the remaining coefficients. In this context, however, I find that the effect size is virtually unchanged when one runs this alternative dynamic specification. The specification also reveals that the backlash is sustained for many decades.

Figure A-2 examines the effects of the ERA on *female* attitudes toward male/female equality. It can be seen that there is no evidence of sharp backlash on the part of women. It is worth noting that, prior to ERA passage, women in ERA-passing states are more sympathetic to the concept of male/female equality than women in non-ERA-passing states, further cementing the observation that, if anything, ERA-passing states were more liberal in their gender attitudes than non-ERA-passing states.

Table A-1 revisits the results using a standardized z-score version of the male/female equality question as the outcome variable, rather than an indicator variable. A higher z-score value represents more positive attitudes toward gender equality. This provides a more continuous outcome measure at the expense of less readily-interpretable coefficients. In any case, the results are fundamentally the same. The introduction of a state ERA leads to a movement in male attitudes toward gender equality by one-third of a standard deviation in the conservative direction.

Table A-2 decomposes the effect into each individual point on the 7-point scale to provide a sense of how the distribution of attitudes toward male/female equality amongst men is changing. That is, are views becoming more polarized or is there a clear movement in one direction? The evidence is that the latter is the case, with views closer to equality becoming less common and views closer to inequality becoming more common. There appears to be an overall rightward shift of the distribution, consistent with the implications of the model.

Column (1) of Table A-3 re-runs the state-level specifications with a linear state time trend included in the regression. This is one way of controlling for the possibility that ERA-adopting states are on a more conservative trend than non-adopting states. Instead, including this time trend simply strengthens the result further, providing evidence that, if anything, ERA adopters are on a more liberal trajectory than non-adopters, which makes intuitive sense. Another robustness check is proposed by Chaisemartin and d'Haultfoeuille (2020), who extend the argument of Borusyak and Jaravel (2017) further and argue that there may be circumstances under which even dynamic difference-in-differences specifications suffer from the same negative-weighting issue that may plague static difference-in-differences specifications. In particular, if the year-*t* dynamic treatment effects are actually heterogeneous across states (for at least some values of *t*), this could drive such a bias. I apply the procedure of Chaisemartin and d'Haultfoeuille using their Stata package `did_multiplegt` and find that my result is robust to it, as seen in column (2) of Table A-3.

Table A-4 revisits the decomposition of the ERA backlash into the campaign effect and the law effect, but with an added twist. Because the law effect is determined by comparing states where the ERA made it onto the ballot but did not pass with states where the ERA made it onto the ballot and did pass, one can restrict the analysis to the closest ERA referenda. As seen in column (4), he effect is robust to restricting to the closest 6 cases – all of which were within a few percentage points of a 50/50 outcome. Indeed, if anything, the effect is stronger in these closest cases, which should represent states where the political climate leading up to the ERA was most similar. Column (2) applies the border-county strategy to the campaign effect regression (since restricting to close elections as a robustness check is impossible in that context), finding that, if anything, the effect of the campaign is to boost stated attitudes toward male/female equality. Once again, the campaign does not appear to be the source of the backlash to the ERA.

A.5 Additional Outcomes

There are many outcomes beyond attitudes and voting patterns that may be affected by backlash. Indeed, the model suggests that any ideologically-coded actions which have the capacity to signal one's ideological positions may manifest backlash. In the context of the ERAs, relationship patterns amongst husbands and wives seem like a particularly relevant outcome.

The National Fertility Survey asked women questions about their *preferred* number of children they'd ideally like to have and about the number of children they *expected* to have, after the joint decision is made by themselves and their husbands. Data from the National Fertility Survey is used in the regressions in Table A-5, and they reveal statistically-significant evidence of divergence. This suggests that, whereas women appear to move in the direction of preferring fewer children, men evidently move toward preferring more or are otherwise exerting more influence over their wives' decision-making.[14]

Given the evidence of divergence between men and women in various dimensions, one might wonder if tensions are increased in marriages as a consequence of the ERA. The GSS has asked questions on self-reported happiness and marital happiness since its inception. Table A-6 shows that, indeed, there is significant evidence of reduced marital happiness and overall happiness for married individuals – but no change in happiness for unmarried individuals. Figure A-3 shows the dynamic specification, which suggests that the effect does not predate ERA passage; rather, it responds sharply afterward.

Turning to the CPS-ASEC data, I now examine whether – given these strong negative effects on male attitudes toward female equality – the state ERAs actually induce negative material consequences for women. Table A-7 provides some evidence in the affirmative. As can be seen in columns (1) through (3), introduction of a state ERA results in a significant

[14] The effect on the gap is statistically significant, but the effect on male and female preferences separately is not.

reduction in incomes for married women but not for unmarried women or for men. This effect, as seen in columns (4) and (5), appears to go through significantly more women remaining homemakers and significantly fewer women making it into management positions. Altogether, these results can be interpreted as a reduction in female empowerment, potentially driven by constraints placed on married women by their backlashing husbands.

A.6 The Broader Women's Movement

The ERA was one of the primary pillars of the women's movement; however, it was not the only one. Large-scale entry of women into the labor force, election of female legislators, and legislation liberalizing access to contraceptives for unmarried women were three of its other biggest facets. Whereas the latter – like the ERA – was imposed in the form of a law, the former two were more bottom-up in nature. This provides the ideal setting for testing whether, indeed, laws play a unique role in generating backlash.

I study the effect of women's entry into the labor market using a shift-share instrument which exploits the fact that, in different industries, female employment has grown at different rates nationally, and prominence of different industries varies from area to area. Consequently, if industry *j* has rapid female employment growth from 1970 to 1990 and it makes up a high share of employment in county *i*, then county *i* will be treated with a large increase in female employment. Formally, the instrument is

$$\Delta S_i^{70,90} = \sum_j \pi_{ij}^{70} \Delta_{-i,j}^{70,90}$$

where π_{ij}^{70} represents the share of industry *j* in total employment of county *i* in 1970 and $\Delta_{-i,j}^{70,90}$ is the national growth of female EPOP (employment-to-population ratio) in industry *j* from 1970 to 1990, computed as a leave-one-out mean. I can then run the first-stage regression

$\Delta E_i^{70,90} = \eta + \gamma \cdot \Delta S_i^{70,90} + u_i$, where $\Delta E_i^{70,90}$ represents the growth of female EPOP in county *i*. This yields a strong first-stage F-statistic of 149, allowing for a valid application of the second-stage regression,

$$\Delta Y_i^{70,90} = \alpha + \beta \cdot \Delta \hat{E}_i^{70,90} + X\gamma + \varepsilon_i$$

where $\Delta Y_i^{70,90}$ denotes the change in some outcome variable of interest in county *i* over the corresponding 1970 to 1990 period[15]. Focusing on the change in attitudes toward male/female equality and the change in Republican vote share as my outcomes of interest, I find no significant evidence of backlash in either domain, as can be seen in Table A-8. With regard to attitudes toward male/female equality, there is no statistically-significant correlation even in the OLS specification. With regard to Republican vote share, to the extent that there is a correlation in the OLS regressions, it is rendered nearly non-significant by the shift-share IV, and in any case, the sign is the *opposite* of backlash, with more female labor-force entry associated with reduced Republican vote shares.

I study the effect of women's election to political office using an electoral RDD on House of Representatives and State Legislature elections. I follow Gyourko and Ferreira (2014), who performed this exercise for mayors, comparing the outcomes generated by male and female mayors subsequent to elections that pitted a male and a female candidate against each other and identifying the effect off of the discontinuity at the 0% victory margin between the male and female candidates. Formally,

$$Y_{it} = \alpha + \beta \cdot FemaleLeg_{it} + f(x_{it}) + \varepsilon_{it} \qquad \forall x_{it} \in (c-h,\, c+h),$$

where Y_{it} is the outcome of interest in district *i* over some defined period subsequent to the

[15] I focus on 1970-1990 both because these were the two decades of most rapid female labor-force entry and because one of the key outcomes of interest – attitudes toward male/female equality – is not available prior to the 1970s.

election year t, x_{it} is the vote share for the female candidate, $FemaleLeg_{it} = \mathbf{1}\{x_{it} > c\}$, and h is the bandwidth around the cutoff c. Again, focusing on attitudes toward male/female equality and the Republican vote share in the subsequent election as my outcomes of interest, I find no significant evidence of backlash, as can be seen in Table A-9. If I instead study effects on female candidates in the subsequent election, I actually find some evidence of *increased* future female vote shares – the opposite of backlash.

I study the liberalization of contraception access to unmarried women using the difference-in-differences framework applied throughout most of this paper. Like the ERA, this is a pillar of the women's movement operationalized through the law. Its effects on fertility patterns and female labor-market decisions were studied in detail by Goldin and Katz (2002). Figure A-4 displays the results of a dynamic specification analagous to the one run in the context of the ERA and reveals that, just like the ERA, this law generated a sharp and significant backlash in male attitudes. Thus there is indeed evidence that laws play a unique role in generating backlash, distinct from the more bottom-up components of the women's movement that were not actualized through legislation.

Table A-1: Static Specifications – ERA (Z-Score Outcomes)

Outcome: Gender Role Attitudes (z-score)	(1) State Diff-in-Diff	(2) State Diff-in-Diff	(3) State Diff-in-Diff	(4) Border Dis-continuity	(5) Border Dis-continuity
Sex:	Both	Male	Female	Male	Female
ERA Indicator	-0.096	-0.330***	0.088	-0.291***	0.078
	(0.066)	(0.094)	(0.028)	(0.083)	(0.106)
Year FEs	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes
State-by-Border FEs	No	No	No	Yes	Yes
Individuals in Sample	All	All	All	Border Residents	Border Residents
Years of Data	1972-1988	1972-1988	1972-1988	1972-1988	1972-1988
Clustering	State	State	State	State	State
Observations	15,477	6677	8800	2350	3169

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level
* Denotes significance at the 5% level; † Denotes significance at the 10% level. Regressions in this table are the analogues of regressions in Table 2, albeit with the ANES gender role attitude variable converted into a z-score. A larger value indicates more positive attitudes toward gender equality.

Table A-2: Gender Equality Scale Point-by-Point ERA Regressions

Outcome: Point-by-point indicators for gender equality position	(1) 1	(2) 2	(3) 3	(4) 4	(5) 5	(6) 6	(7) 7
Sex:	Male	Male	Male	Male	Male	Male	Male
ERA Indicator	-0.117*	-0.018	-0.042	0.032	0.032	0.011	0.045***
	(0.053)	(0.017)	(0.039)	(0.033)	(0.019)	(0.030)	(0.016)
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clustering	State	State	State	State	State	State	State
Observations	6677	6677	6677	6677	6677	6677	6677

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Regressions in this table are the analog of the main specification for males – column (2) in Table 2 – but with the regressions run point-by-point along the 7-point gender role attitude variable. An indicator variable is generated for each point; thus the coefficient can be interpreted as the change in the share of men who are at that point in the gender role attitude distribution. "1" corresponds to "men and women should have an equal role in running business, industry, and government." "7" corresponds to "a woman's place is the home."

Table A-3: Additional Robustness

Outcome: Indicator for Positive Attitudes toward Gender Equality	(1) State Time Trend	(2) Chaisemartin – d'Haultfoeuille
Sex:	Male	Male
ERA Indicator	-0.236***	-0.163*
	(0.060)	(0.083)
Year FEs	Yes	Yes
State FEs	Yes	Yes
Individuals in Sample	All	All
Years of Data	1972-1988	1972-1988
Clustering	State	State
Observations	6677	6677

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Column (1) adds state-specific time trends to the baseline specification. Column (2) applies the approach of Chaisemartin and d'Haultfoeuille (2020) using their `did_multiplegt` package in Stata.

Table A-4: Campaign Effects, Law Effects, and Close ERA Referenda

Outcome: Indicator for Positive Attitudes toward Gender Equality	(1) Campaign Effect	(2) Campaign Effect	(3) Law Effect	(4) Law Effect
Sex:	Male	Male	Male	Male
ERA Indicator	-0.030	0.155**	-0.159***	-0.361***
	(0.040)	(0.060)	(0.057)	(0.106)
Year FEs	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes
State-by-Border FEs	No	Yes	No	No
Individuals in Sample	All in ERA-on-Ballot States	Border Residents of ERA-on-Ballot States	All in ERA-on-Ballot States	All in 6 closest ERA-on-Ballot States
Years of Data	1972-1988	1972-1988	1972-1988	1972-1988
Clustering	State	State	State	State
Observations	4,010	991	2,994	990

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Control group for campaign effect regressions is states where the ERA never made it onto the ballot; treatment group is states where the ERA made it onto the ballot but failed. Control group for law effect regressions is states where the ERA made it onto the ballot but failed; treatment group is states where the ERA made it onto the ballot and succeeded.

Table A-5: National Fertility Survey

	(1)	(2)	(3)
Outcome:	Ideal # Children, Self	Expected # Children	Ideal # Children Gap
Sex:	Female	Female	Female
ERA Indicator	-0.134	0.140	0.287***
	(0.087)	(0.111)	(0.092)
Year FEs	Yes	Yes	Yes
State FEs	Yes	Yes	Yes
Years of Data	1965, 1975	1965, 1975	1965, 1975
Clustering	State	State	State
Observations	8,983	9,002	8,967

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Regressions use 1965 and 1975 waves of the National Fertility Survey. 1970 wave lacks publicly-available state codes. The outcome for column (1) is women's responses to a question about the number of children they'd ideally like to have, if it was up to them. The outcome for column (2) is women's responses to a question about the number of children they expect to have, after the joint decision is made by them and their husbands.

Table A-6: Happiness

	(1)	(2)	(3)	(4)
Outcome:	Marital Happiness (z-score)	Happiness (z-score)	Happiness, Married People (z-score)	Happiness, Single People (z-score)
Sex:	Both	Both	Both	Both
ERA Indicator	-0.109***	-0.080***	-0.119***	0.001
	(0.025)	(0.021)	(0.031)	(0.069)
Year FEs	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes
Individuals in Sample	All	All	Married People	Single People
Years of Data	1973-2016	1973-2016	1973-2016	1973-2016
Clustering	State	State	State	State
Observations	13,040	56,104	26,619	11,864

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Regressions use GSS data on self-reported happiness.

Table A-7: Material Economic Outcomes (CPS)

	(1)	(2)	(3)	(4)	(5)
Outcome:	log(Income)	log(Income)	log(Income)	Management Positions	Homemakers
Sub-Population:	Male	Female, Married	Female, Unmarried	Female	Female
ERA Indicator	0.025	-0.046**	0.042	-0.008**	0.013*
	(0.024)	(0.020)	(0.032)	(0.003)	(0.006)
Year FEs	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes
State-by-Border FEs	No	No	No	No	No
Individuals in Sample	All	All	All	All	All
Years of Data	1968-1988	1968-1988	1968-1988	1968-1988	1968-1988
Clustering	State	State	State	State	State
Observations	972,980	609,842	479,162	1,415,832	1,415,832

Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Regressions use CPS-ASEC data. The outcome for columns (1)-(3) is the log of total income. The outcome for column (4) is an indicator for whether the respondent has a management position. The outcome for column (5) is an indicator for whether the respondent indicated that their primary activity was homemaking.

Table A-8: Female Labor-Force Entry Shift-Share

	(1) OLS, Attitude toward Gender Equality	(2) IV, Attitude toward Gender Equality	(3) OLS, Rep Vote Share	(4) IV, Rep Vote Share
Sex:	Male	Male	Both	Both
ΔFemaleEPOP	0.913	3.360	-0.599***	-4.104†
	(0.600)	(2.066)	(0.163)	(2.172)
Year FEs	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes
Data Source	ANES	ANES	Electoral Atlas	Electoral Atlas
Years of Data	1972-1998	1972-1998	1968-1992	1968-1992
Clustering	State	State	State	State
Observations	104	104	3,106	3,106

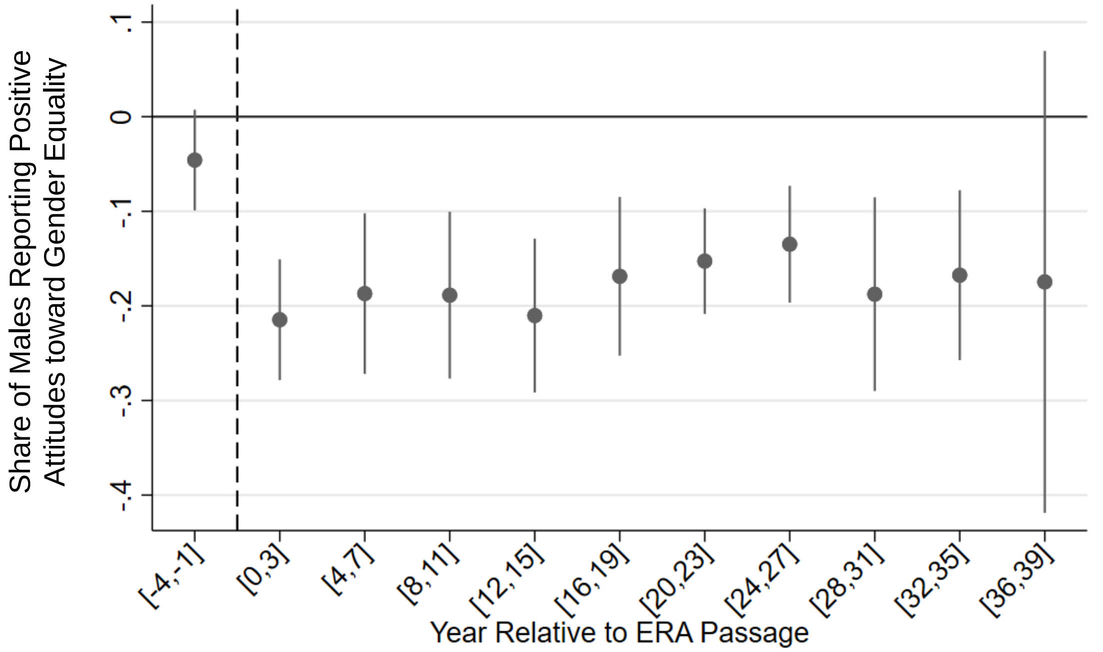
Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Columns (1) and (2) use the indicator for positive attitudes toward gender equality generated from the ANES data as their outcome variable. Columns (3) and (4) use data on official presidential election returns from Dave Leip's electoral atlas as their outcome variable. OLS specifications in columns (1) and (3) regress the outcome directly on the change in the female employment-to-population ratio (EPOP). IV specifications in columns (2) and (4) instrument the latter with the shift-share instrument.

Table A-9: Female Legislators RDD

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Female Vote Share, HoR	Female Vote Share, State Leg.	Female Vote Share, HoR (Non-Incumbent)	Female Vote Share, State Leg. (Non-Incumbent)	Republican Vote Share, HoR	Republican Vote Share, State Leg.	Attitude toward Gender Equality, HoR
Sex:	Both	Both	Both	Both	Both	Both	Male
FemaleVictory	0.107	0.024***	0.096	0.022*	-0.027	0.007	-0.048
	(0.083)	(0.008)	(0.106)	(0.011)	(0.059)	(0.014)	(0.101)
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Data Source	MIT Election Lab	Klarner et al.	MIT Election Lab	Klarner et al.	MIT Election Lab	Klarner et al.	ANES
Years of Data	1976-2018	1967-2010	1976-2018	1967-2010	1976-2018	1967-2010	1972-2008
Bandwidth	5%	5%	5%	5%	5%	5%	5%
Clustering	District	District	District	District	District	District	District
Observations	2,124	18,057	1,116	13,069	5,717	17,823	5,626

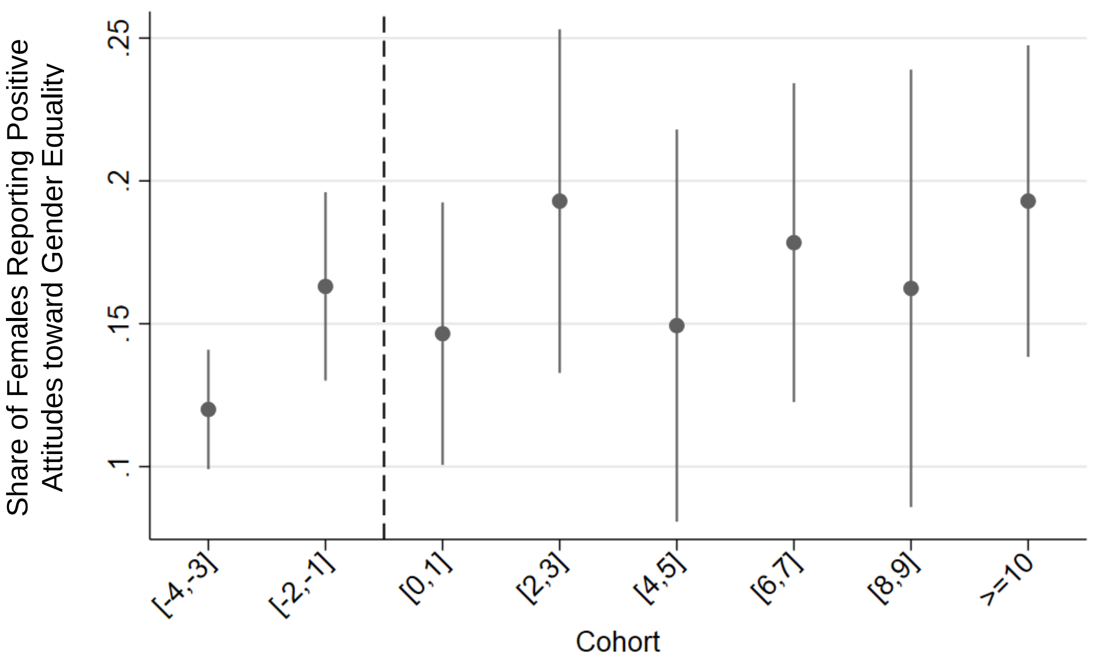
Note: *** Denotes significance at the 1% level; ** Denotes significance at the 2.5% level; * Denotes significance at the 5% level; † Denotes significance at the 10% level. Columns (1) and (2) use the vote share of female candidates as their outcome. Columns (3) and (4) use the vote share of non-incumbent female candidates as their outcome. Columns (5) and (6) use Republican vote share as an outcome. Column (7) uses the indicator for positive attitudes toward gender equality generated from the ANES data as their outcome variable. The odd-numbered columns study close House of Representatives elections using data from the MIT election lab. The even-numbered columns study close state legislative elections using data from Klarner et al. (2013). The fact that the ANES does not contain state legislative district geocodes makes it infeasible to run a specification studying the effect of close state legislative elections on ANES gender role attitudes.

Figure A-1: Dynamic Difference-in-Differences – ERA Effects, Long Horizon



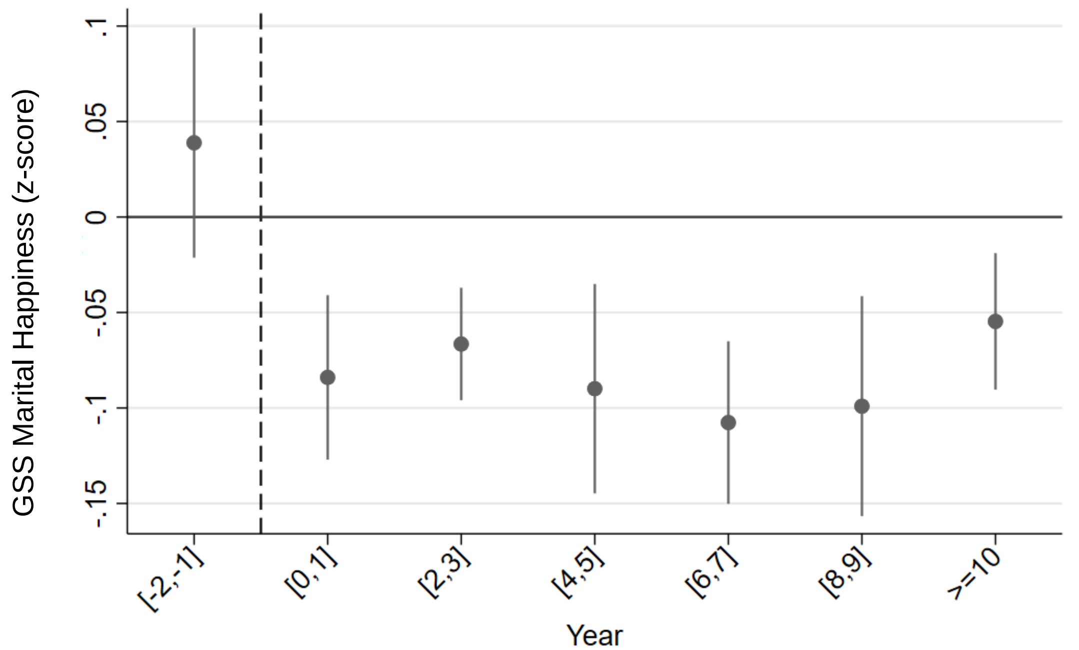
Note: Year 0 corresponds to the year the state ERA takes effect.

Figure A-2: Dynamic Difference-in-Differences – ERA Effects on Female Attitudes



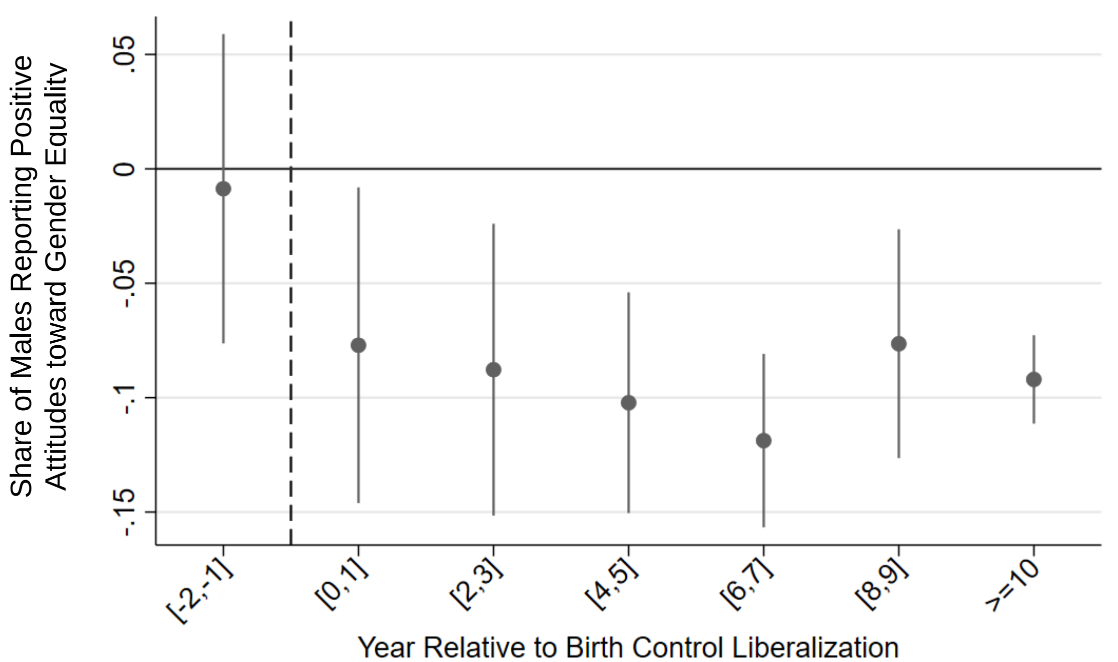
Note: Year 0 corresponds to the year the state ERA takes effect.

Figure A-3: Dynamic Difference-in-Differences –
ERA Effects on Marital Happiness



Note: Year 0 corresponds to the year the state ERA takes effect.

Figure A-4: Dynamic Difference-in-Differences –
Birth Control Pill Legislation Effects on Male Attitudes



Note: Year 0 corresponds to the year the female birth control access was liberalized.

Appendix B: Appendix to Chapter 2

B.1 mTurk Observational Survey Questionnaire

Page 1

Please verify:

I'm not a robot

reCAPTCHA
Privacy - Terms

Please answer the following problem: 5 + 9 = ?

Please answer the following problem: 2 + FOUR + 3 + SIX = ? (Ignore the numerals, only sum the words)

Page 2

What is your sex?

- Male
- Female

What is your age?

What is your race?

- White, non-Hispanic
- Black, non-Hispanic
- Asian, non-Hispanic
- Hispanic/Latino
- Mixed Race
- Other

Page 2 (cont'd)

What is your approximate household income?

- $0 - 24,999
- $25,000 - 49,999
- $50,000 - 74,999
- $75,000 - 99,999
- $100,000 - 149,999
- $150,000 - 199,999
- Above $200,000

What is your education level?

- Less than High School
- High school graduate
- Some college
- 2 year degree
- 4 year degree
- Master's degree
- Professional degree
- Doctorate

In which state do you currently live?

In which city or town do you currently live?

Page 3

For which presidential candidate did you vote in 2016?

- Hillary Clinton
- Donald Trump
- Gary Johnson
- Jill Stein
- Other/Did Not Vote/Prefer Not to Say

Where would you place yourself on the political spectrum?

- Very Conservative
- Conservative
- Slightly Conservative
- Moderate
- Slightly Liberal
- Liberal
- Very Liberal

What is your party identification?

- Democratic
- Republican
- Independent/Other

Page 4

Do you approve or disapprove of the way your state governor is handling the coronavirus (COVID-19) outbreak?

- Strongly disapprove
- Disapprove
- Neither approve nor disapprove
- Approve
- Strongly approve

Page 5

Different states have had different numbers of coronavirus deaths per million population. We are going to ask you a series of questions about which states you believe have had higher deaths per million population.

For each pair of states we will ask which state you think has had higher deaths per million population from coronavirus. We will then ask a follow-up question about how much higher you think deaths per million population were.

Your responses to these questions are incentivized. That is, you will receive a higher payment the closer your answers are to the truth.

For example, suppose State A has a population of 1,000,000 and 200 deaths from coronavirus. This means it has had deaths per million population of 200 per million.

Suppose State B has a population of 2,000,000 and 100 deaths from coronavirus. This means it has had a deaths per million population of 50 per million.

This means that State A has a higher deaths per million population, and specifically, deaths per million population in State B are 25% [= 50/200] as high as deaths per million population in State A.

Comprehension Check

Suppose State C has a population of 1,000,000 and 150 deaths from coronavirus. And suppose State D has a population of 2,000,000 and 600 deaths from coronavirus.

Which state has higher deaths per million population?

- State C
- State D

Page 6

Consider the following two states below. As of October 12th, which of these states do you think had a higher number of deaths per million population from coronavirus?

California

Florida

California

Please do not leave the page while this survey is in progress

Page 7

You said that Florida has higher deaths per million population than California. For every 100 deaths per million population in Florida, how many deaths per million population do you think there are in California? Please select your answer using the slider below.

(Note: If you have no idea at all which state has more deaths per million population and you randomly guessed on the previous page, the best thing to do here is to guess 100.)

0 10 20 30 40 50 60 70 80 90 100

California 65

Please do not leave the page while this survey is in progress

[Pages 6 and 7 repeated 10 times for randomly-selected state pairs]

Page 8

Next we are going to ask you questions about some of the factors that cause states to be affected differently by coronavirus.

For example, you might believe that it is due to state governments having policy and competence differences.

You might also believe that states would have higher deaths per million population because of non-political factors:
- They have an older population on average
- They are more densely-populated
- They have higher public transit usage
- They have more international travelers
- And any other factors you can think of

We are going to ask you how big a difference in deaths per million population you believe these factors would produce.

Page 9

Consider the following two states below. Supposing the governments in the two states were equally competent, which state do you believe **would** have had higher deaths per million population from coronavirus due to the other aforementioned factors?

Michigan

Rhode Island

Michigan

Page 10

You just said that Michigan would have higher deaths per million population from coronavirus due to the aforementioned non-governmental factors than Rhode Island.

For every 100 deaths per million population that Michigan would have due to these non-governmental factors, how many deaths per million population would Rhode Island have due to these non-governmental factors? Please select your answer using the slider below.

0 10 20 30 40 50 60 70 80 90 100

Rhode Island 65

[Pages 9 and 10 repeated 10 times for randomly-selected state pairs]

Page 11

Thank you for participating!

Your completion code is: 20522

To receive compensation for participating, please enter this code into the box on your Mechanical Turk window.

B.2 mTurk Survey Experiment Questionnaire

[Survey begins with same demographic questions as in Appendix B.1]

Belief Elicitation

Next we are going to ask you to estimate how many coronavirus (COVID-19) deaths your home state has had per million people who live in the state.

For example, suppose State A has a population of 2 million people and has had 1600 deaths from coronavirus (COVID-19), while State B has a population of 10 million people and has had 8000 deaths from coronavirus (COVID-19).

Then:

- State A has had 1600/2 = 800 deaths per million population
- State B has had 8000/10 = 800 deaths per million population.

How many deaths per million population from coronavirus (COVID-19) has each of these hypothetical states had?

State A = 1600, State B = 8000

State A = 800, State B = 800

State A = 100, State B = 200

State A = 1500, State B = 600

Your response to this question is incentivized. That is, you will receive a higher payment the closer your answers are to the truth.

As of December 19th, the average US state has had 959 deaths per million population from coronavirus (COVID-19). *

How many deaths per million population do you think Massachusetts has had from coronavirus (COVID-19)?

Belief Elicitation (cont'd)

Next we would like you to consider some of the factors that cause states to be affected differently by coronavirus (COVID-19).

For example, you might believe that it is due to state governments having policy and competence differences.

You might also believe that states would have higher deaths per million population because of non-political factors:
- They have an older population on average
- They are more densely-populated
- They have higher public transit usage
- They have more international travelers
- And any other factors you can think of

We are going to ask you how big a difference in deaths per million population you believe these factors would produce.

Recall that, as of December 19th, the average US state has had 959 deaths per million population from coronavirus (COVID-19). Suppose the state government of Massachusetts was equally competent to the average US state government.

In that case, how many deaths per million population do you think Massachusetts would have had from coronavirus (COVID-19), taking into account the aforementioned factors?

[After belief elicitation, randomize over Control, Treatment, Hypothetical Arms]

Control Arm

Do you approve or disapprove of the way your state governor is handling the coronavirus (COVID-19) outbreak?

- Strongly approve
- Approve
- Neither approve nor disapprove
- Disapprove
- Strongly disapprove

Information Treatment Arm

Previously you estimated that, as of December 19th, Massachusetts has had 959 deaths per million population from coronavirus (COVID-19).

According to official tallies, Massachusetts actually had 1675 deaths per million population from coronavirus (COVID-19).

Do you approve or disapprove of the way your state governor is handling the coronavirus (COVID-19) outbreak?

- Strongly approve
- Approve
- Neither approve nor disapprove
- Disapprove
- Strongly disapprove

Hypothetical Arm

Previously you estimated that, as of December 19th, Massachusetts has had 959 deaths per million population from coronavirus (COVID-19).

Suppose you learned that, according to official tallies, Massachusetts actually had 1070 deaths per million population from coronavirus (COVID-19). In that case, would you approve or disapprove of the way your state governor is handling the coronavirus (COVID-19) outbreak?

- Strongly approve
- Approve
- Neither approve nor disapprove
- Disapprove
- Strongly disapprove

Appendix C: Appendix to Chapter 3

C.1 Proof of Proposition

Observe that the individual can spend no more than his endowment ω – plus any investment income – over the course of his life. Thus, in a setting without taxation, his budget constraint would be by $(1 + R)c_1 + c_2 = (1 + R)\omega$. However, in the present setting, the individual must pay income tax on his interest income.

By definition, the individual saves and invests $\omega - c_1$ at the end of the first period. At the start of the first period, he receives interest income of $R(\omega - c_1)$, on which he must pay total tax $\tau(R(\omega - c_1)) * R(\omega - c_1) = [\alpha + \beta R(\omega - c_1)]R(\omega - c_1)$. Consequently, his budget constraint is

$$(1 + R)c_1 + c_2 = (1 + R)\omega - \alpha R(\omega - c_1) - \beta R^2(\omega - c_1)^2 .$$

Letting the individual have a discount factor of δ, the Lagrangian for the relevant intertemporal utility maximization problem is as follows:

$$L(c_1, c_2, \mu) = u(c_1) + \delta u(c_2) - \mu\left[(1+R)c_1 + c_2 + \alpha R(\omega - c_1) + \beta R^2(\omega - c_1)^2\right].$$

Differentiating the Lagrangian with respect to c_1 and c_2 and setting these expressions equal to zero in order to obtain a maximum,

$$\partial L / \partial c_1 = u'(c_1) - \mu\left[(1+R) - \alpha R - 2\beta R^2(\omega - c_1)\right] = 0 ,$$

$$\partial L / \partial c_2 = \delta u'(c_2) - \mu = 0 .$$

This leads to the following expression:

$$g(c_1; \alpha, \beta) \equiv -u'(c_1) + \left[1 + (1-\alpha)R - 2\beta(\omega - c_1)R^2\right] \cdot \delta u'\left((\omega - c_1)\left(1 + (1-\alpha)R - \beta(\omega - c_1)R^2\right)\right) = 0 .$$

Applying the Implicit Function Theorem, we have

$$\partial c_1^* \Big/ \partial \alpha = \frac{\delta u'(c_2)R + \left[1+(1-\alpha)R - 2\beta(\omega - c_1)R^2\right]\delta u''(c_2)(\omega - c_1)R}{-u''(c_1) + 2\beta\delta u'(c_2)R^2 - \left[1+(1-\alpha)R - 2\beta(\omega - c_1)R^2\right]^2 \delta u''(c_2)},$$

Note that the denominator of this expression is unambiguously positive provided $u' > 0$ and $u'' < 0$. The numerator, however, is ambiguous. The first term is positive (for sufficiently low β), whereas the second is negative. For logarithmic utility, these two effects – the income and substitution effects – cancel each other out exactly. For linear utility, the first term dominates and consumption increases/investment decreases in response to an increased base tax rate. More precisely, investment decreases if

$$\delta u'(c_2)R > -\left[1+(1-\alpha)R - 2\beta(\omega - c_1)R^2\right]\delta u''(c_2)(\omega - c_1)R$$

$$\Leftrightarrow -\frac{c_2 u''(c_2)}{u'(c_2)} < \frac{c_2}{(\omega - c_1)\left[1+(1-\alpha)R - 2\beta(\omega - c_1)R^2\right]} = 1.$$

Now, again applying the Implicit Function Theorem,

$$\partial c_1^* \Big/ \partial \beta = \frac{2\delta u'(c_2)(\omega - c_1)R^2 + \left[1+(1-\alpha)R - 2\beta(\omega - c_1)R^2\right]\delta u''(c_2)(\omega - c_1)^2 R^2}{-u''(c_1) + 2\beta\delta u'(c_2)R^2 - \left[1+(1-\alpha)R - 2\beta(\omega - c_1)R^2\right]^2 \delta u''(c_2)}.$$

Note that $\partial c_1^* / \partial \beta$ is *not* the comparative static of interest when it comes to determining how consumption changes as tax progressivity changes while holding overall taxation constant. Increasing β without modifying α will increase the average tax rate at any (positive) level of income. If it was found that increasing β alone increased consumption and thus decreased saving, this would scarcely be a breakthrough, as it may simply be going through the channel of an increased average tax rate rather than the channel of progressivity in itself.

So, consider an increase in β from β_1 to β_2. In order to keep an individual's overall tax rate constant, how must α change? For a given income level y_0,

$$\alpha_1 + \beta_1 y_0 = \alpha_2 + \beta_2 y_0 \Rightarrow \alpha_2 = \alpha_1 + (\beta_1 - \beta_2) y_0 .$$

Thus, if β increments by Δ, α must decrement by Δy_0. Consequently, in order to prove the proposition, it is necessary to show that the directional derivative

$$\partial c_1^* \big/ \partial \beta - (\omega - c_1) R \, \partial c_1^* \big/ \partial \alpha > 0 .$$

Substituting in the above expressions, this yields

$$\frac{\delta u'(c_2)(\omega - c_1)R^2}{-u''(c_1) + 2\beta\delta u'(c_2)R^2 - \left[1 + (1-\alpha)R - 2\beta(\omega - c_1)R^2\right]^2 \delta u''(c_2)} > 0 .$$

Note that, with $u' > 0$ and $u'' < 0$, this expression is unambiguously positive. Because investment $s = \omega - c_1$, this means that

$$\partial s^* \big/ \partial \beta - (\omega - c_1) R \, \partial s^* \big/ \partial \alpha < 0 ,$$

which proves the proposition.

C.2 Constructing Measures of Fiscal Size

Constructing a measure of the size of the legislated tax changes that occurred with the flat tax introduction in these countries is not an entirely straightforward process. It requires information on how the incidence of an income tax change falls on various income groups. Evidence suggests that 99% of the income distribution is well-approximated by a lognormal distribution (Clementi and Gallegati 2005). So, I begin with the data on share of total income by population decile from the WIID. These data can be fit to a lognormal distribution, with mean

$$\mu = \frac{\sum_{j=1}^{10} \ln y_j}{10}, \text{ where } y_j = \frac{s_j \cdot GNI}{POP/10}$$

and variance

$$\sigma^2 = \frac{\sum_{j=1}^{10} (y_j - \mu)^2}{10 - 1},$$

where s_j is the share of income accruing to population decile j (ordered by income), GNI is gross national income, and POP is national population. y_j, then, is average gross income in population decile j.

With a lognormal income distribution, the share, $S(y)$, of income accruing to people with income below y can then be calculated by integrating the income distribution up to incomes y and dividing this by the integral of said distribution over all incomes.

$$S(y)=\frac{\int_0^y {}^{1}\!\!\diagup\!_{\sigma\sqrt{2\pi}}\cdot\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)dx}{\int_0^\infty {}^{1}\!\!\diagup\!_{\sigma\sqrt{2\pi}}\cdot\exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)dx}$$

Denoting by $P(y)$ the proportion of the population with incomes below y, it is now possible to compute the share of total income originating from each bracket—that is, the fraction exposed to each marginal income tax rate. To that end, define

$$y^i\equiv\frac{D+max\_inc^0}{1-\tau_{ep}},$$

where D denotes the personal deduction, max_inc_i denotes the maximum income included in tax bracket i, and τ_{ep} denotes the payroll tax paid by employees. (Note that the denominator of y^i is instead 1 if the employee's contribution to payroll tax is not deductible from income tax in the country in question.) max_inc^0 is defined to be 0. Thus y^i is a measure of adjusted gross income. Next, define

$$\begin{aligned}M(y^i)&\equiv\left(S(y^i)-S(y^{i-1})\right)\cdot GNI-\left(P(y^i)-P(y^{i-1})\right)\cdot POP\cdot max\_inc^i\\U(y^i)&\equiv\left(max\_inc^i-max\_inc^{i-1}\right)\cdot\left(1-P(y^i)\right)\cdot POP\end{aligned}$$

$M(y^i)$ represents the total amount of (adjusted gross) income in the range [$max\_inc^{i\text{-}1}$, max_inc^i] made by individuals whose total AGI is within said bracket. This is computed by subtracting the total income in the range [0, max_inc^{i-1}] of these people from their overall total income. $U(y^i)$ represents the total amount of income in the range [max_inc^{i-1}, max_inc^i] made by individuals whose gross income is *above* said bracket. This is computed simply by multiplying the number of individuals whose gross income is above said bracket by the width of the bracket.

$$\Psi(y^i) = \frac{M(y^i) + U(y^i)}{GNI}$$

The two possible sources of income in bracket [max_inc^{i-1}, max_inc^i] are then divided by total national income, yielding $\Psi(y^i)$, which measures the share of total adjusted gross income originating from bracket *i*.

As such, the total fiscal size of an income tax change can be computed as

$$\Delta T_{Inc} = \frac{\sum_{i=1}^{N} (\tau_i^{after} - \tau_i^{before}) \Psi(y^i)}{\sum_{j=1}^{N} \tau_j^{before} \Psi(y^j)} \cdot T_{Inc},$$

where T_{Inc} is the total income tax revenue in the economy (as a percentage of GDP), τ_i^{before} is the marginal tax rate on individuals in bracket *i* before the reform, and τ_i^{after} is the marginal rate on those individuals after the reform. The total fiscal size of a payroll tax/social contribution change, $\Delta T_{Payroll}$, can be computed analogously.

Much more simply, the total fiscal size of a corporate tax change and a VAT change are calculated, respectively, as

$$\Delta T_{Corp} = \frac{\tau_{Corp}^{after} - \tau_{Corp}^{before}}{\tau_{Corp}^{before}} \cdot T_{Corp}$$

$$\Delta T_{VAT} = \frac{\tau_{VAT}^{after} - \tau_{VAT}^{before}}{\tau_{VAT}^{before}} \cdot T_{VAT}$$

Note that T_{Inc}, $T_{Payroll}$, T_{Corp}, and T_{VAT} are obtained from the IMF Government Finance Statistics dataset. Combining these components, this leads to a measure of the overall fiscal size of the tax change,

$$\Delta T = \Delta T_{Inc} + \Delta T_{Payroll} + \Delta T_{Corp} + \Delta T_{VAT} .$$

C.3 Measuring Progressivity

The Average Marginal Tax Rate (AMTR) is a measure that has been much-used in the macro-public finance literature. It dates back to Barro and Sahasakul (1983, 1986), providing a macro-level measure of the marginal tax rates faced by a typical unit of income in the relevant country. In particular, it is calculated as follows:

$$AMTR = \sum_b ShareIncome_b \cdot MTR_b$$

That is, the AMTR is a weighted average of the individual marginal tax rates (MTRs) in a country's tax schedule, where each MTR_b is weighted by the share of total income *for which the earner is in* the corresponding tax bracket, *b*. For instance, consider a country with two tax brackets, 20% below 1000 units of currency and 30% above 1000 units of currency. If half the population makes 800 units of currency in a year and the other half makes 1200 units, the average marginal tax rate is 25%. Even though the bulk of income in this hypothetical economy was taxed at the 20% rate, half of the populace faces the 30% on *any marginal income* that they earn. This is what the AMTR measures. As discussed by Barro and Sahasakul (and a multitude of more recent papers), because individuals respond to marginal rates rather than average rates in a whole range of economic decision-making, the AMTR is a more useful concept for macro-level examination of the response of investment, labor supply, etc. to various incentive changes.

The standard deviation of the marginal tax rate (SDMTR) is a useful extension of this concept that is amenable to measuring tax progressivity.

$$SDMTR = \sum_b ShareIncome_b \cdot (MTR_b - AMTR)^2$$

Consider, for example, a pure flat tax system wherein every individual pays 20% on all income. In this case, because the MTR is equal to the AMTR throughout the tax schedule, the SDMTR will be precisely zero. The greater the commonality of deviations in the MTR from the AMTR (i.e., the higher the progressivity), the higher the value of the SDMTR. Note that, for the purposes of my empirical work, I compute the $ShareIncome_b$ in each bracket b using the income distributions derived in Appendix C.2.

Appendix D: Appendix to Chapter 4

D.1 Existence of Steady State

Steady state in our model requires constant real variables. We show that there exists a steady state with a constant growth rate of the money supply. This steady state requires that the exogenous minimum wages in each states grow at the same rate as the money supply. We start with the intertemporal Euler equation, which says

$$\dot{c}_t = \frac{1}{\gamma}\left(\frac{R_{s,t}}{P_{s,t}} - (\delta + \rho)\right).$$

For all s and for all t in steady state, it must therefore hold that

$$\frac{R_{s,t}}{P_{s,t}} = \delta + \rho.$$

Thus, in each state s, R and P must change at the same rate in steady state. The intratemporal Euler equation is

$$U'(C_t)\frac{W_{s,t}}{P_{s,t}} = V'\left(H_{s,t}\right)$$

This tells us that in each state s, W and P must change at the same rate in steady state. Note now that the consumer substitution equation in steady state,

$$\dot{y}_{s,t}^{NT} - \dot{y}_{s,t}^{T} = 0 = -\sigma_{NT,T}\left(\dot{p}_{s,t}^{NT} - \dot{p}_t^{T}\right),$$

tells us that all non-tradable prices must grow at the same rate as the national tradable price in steady state. This in turn tells us that the state price indices grow at the same rate in all states, since

$$\dot{p}_{s,t} = \frac{GDP_s^{NT}}{GDP_s}\,\dot{p}_{s,t}^{NT} + \frac{GDP_s^{T}}{GDP_s}\,\dot{p}_t^{T}.$$

Our above analysis then yields that R and W also grow at the same rate in all states, the same rate as the state price indices. This rate of price growth is given by the rate of money growth, since

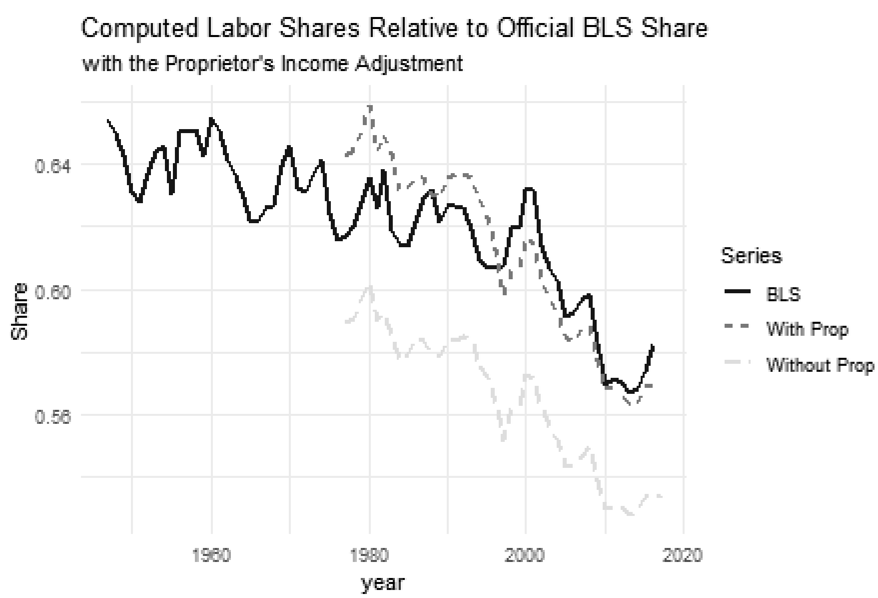
$$\dot{m}_t = \frac{\boldsymbol{GDP}_s'}{GDP_t}\,\dot{\boldsymbol{p}}_t.$$

Now the profit maximization equations of the tradable and non-tradable sectors tell us that it must be that minimum wages grow at the same rate in all states, a rate that is given by the rate of money growth.

D.2 Additional Plots for Calibrations

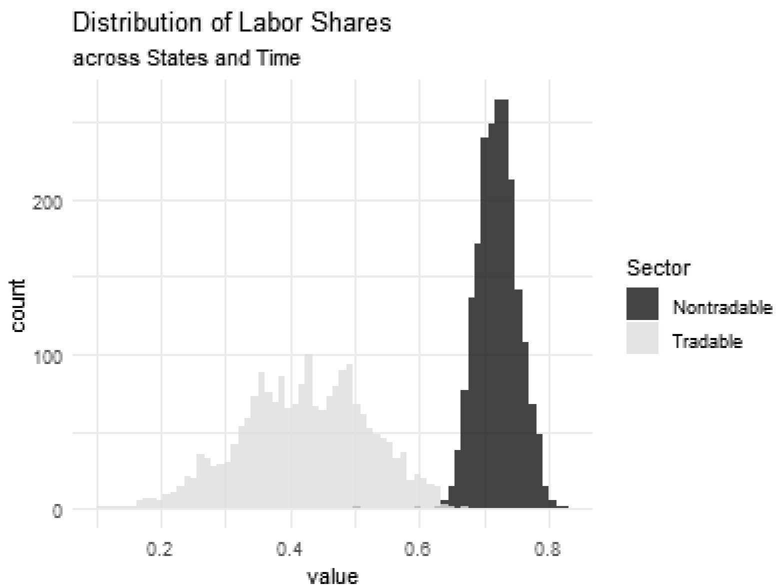
Below we show how our computed labor share at the national level compares to the BLS share. Adjusting for proprietor's income as the BLS suggests brings our calculation much closer to theirs.

Figure D-1: Labor Share Comparison



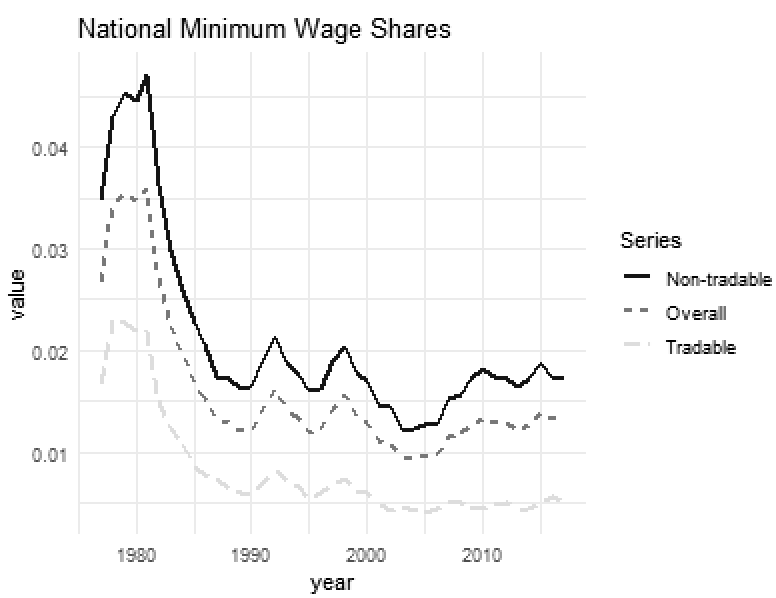
Below, we compute our labor shares using the BLS methodology at the state level, separately for tradables and non-tradables. Shares are much lower in the tradable sector.

Figure D-2: Distribution of Labor Shares



Below we show the minimum wage cost shares at the national level. These are substantially lower than the share of total employment at the minimum wage, since these shares are weighed down both by the fact that minimum wage workers earn less than other workers and by the labor share.

Figure D-3: Tradable/Non-Tradable Minimum Wage Shares over Time

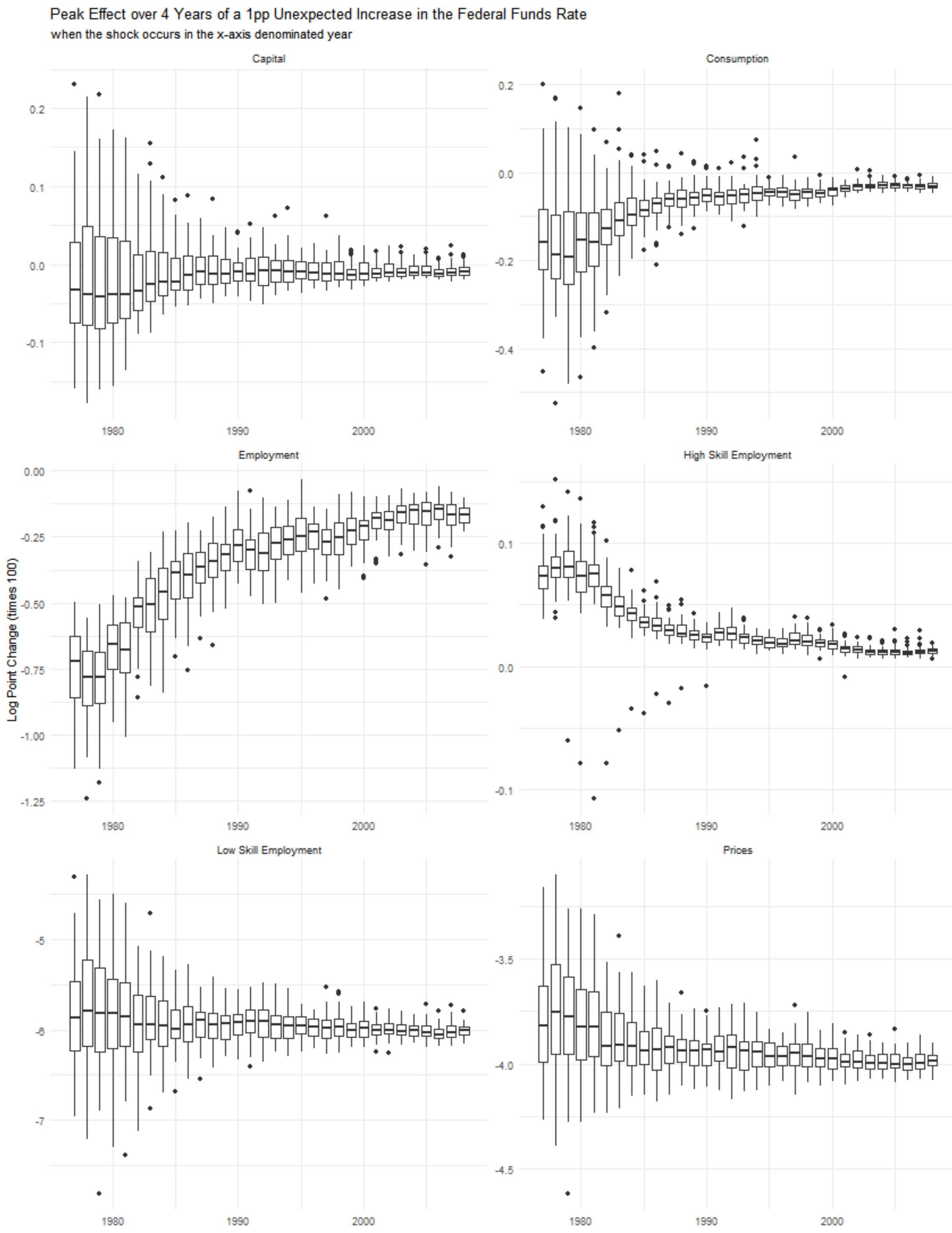


D.3 Additional Model Outcomes

Below, we reproduce all model plots in the paper for an alternative specification, $\sigma_s = .1$:

Figure D-4: Model Outcomes (Alternative Calibration)

Panel 1



Panel 2

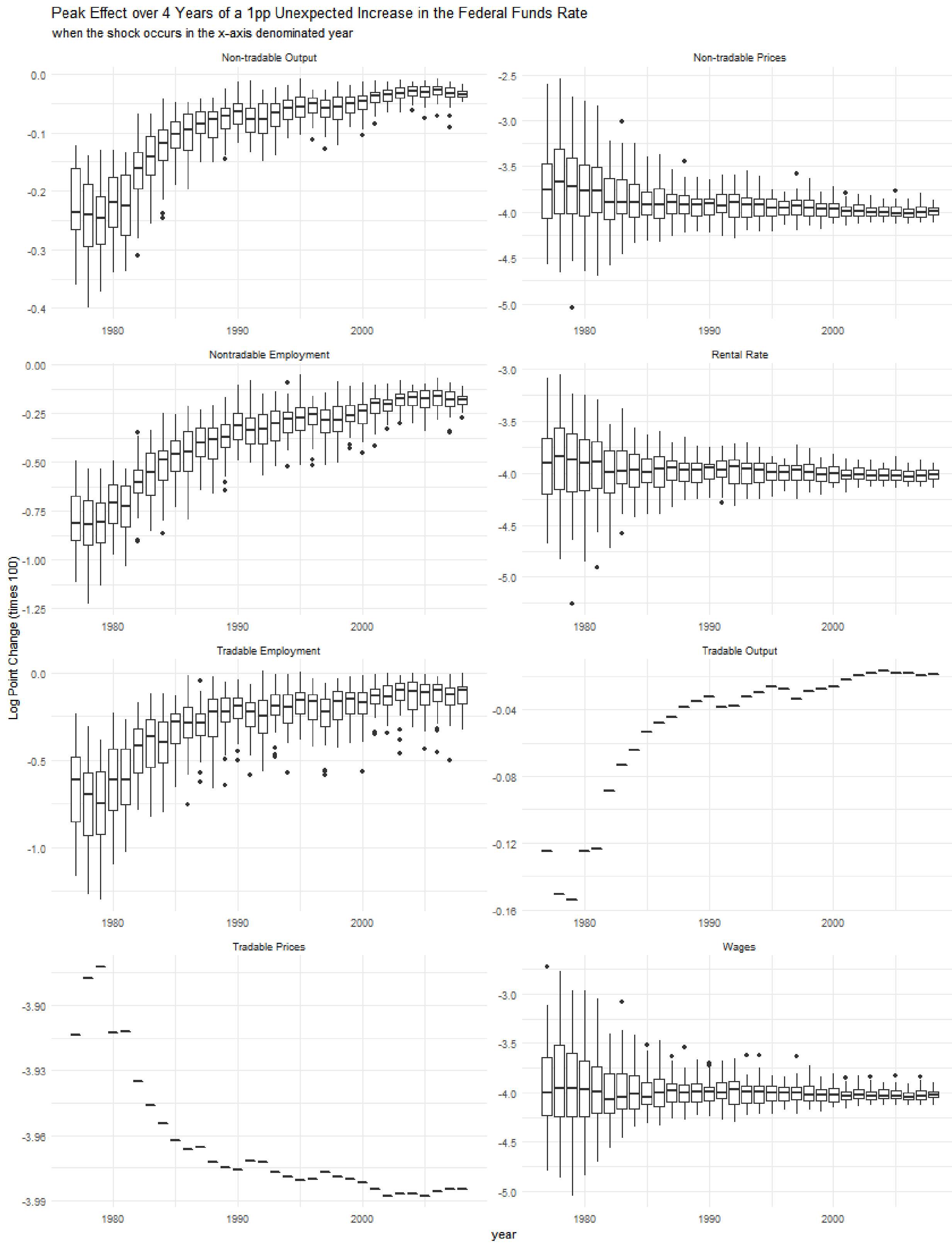
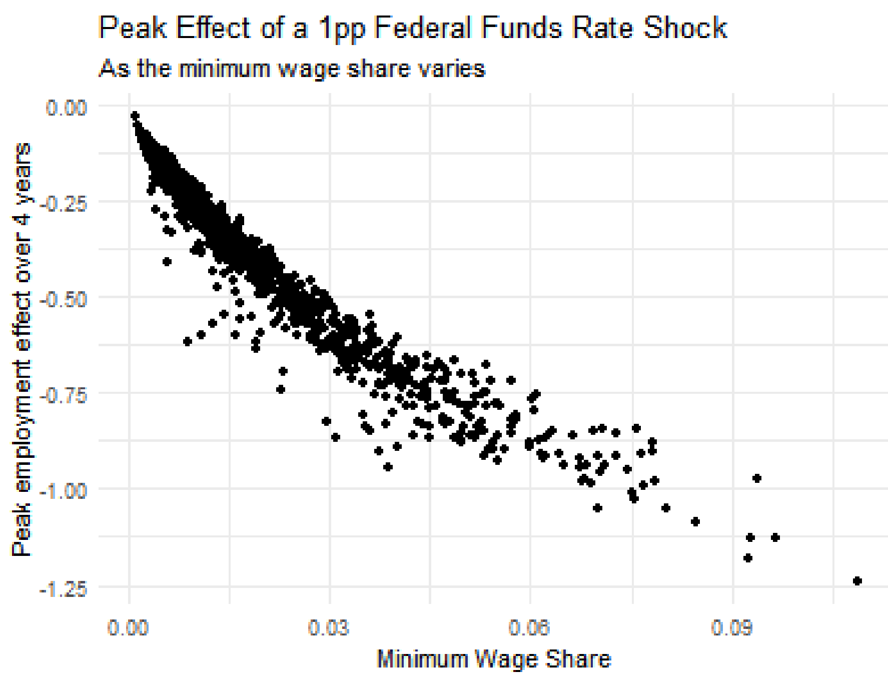

Peak Effect over 4 Years of a 1pp Unexpected Increase in the Federal Funds Rate
when the shock occurs in the x-axis denominated year
Non-tradable Output
Non-tradable Prices
Nontradable Employment
Rental Rate
Tradable Employment
Tradable Output
Tradable Prices
Wages
Log Point Change (times 100)
year

Figure D-5: How Employment Effects Vary with the Minimum Wage Share (Alternative Calibration)

Panel 1



Panel 2