Taxes and Transfers, Growth and Opportunity

by

John S. Olson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2024

Doctoral Committee:

Professor Christopher House, Chair
Professor Ashley Craig, Australian National University
Professor Sarah Miller
Professor Joel Slemrod

John S. Olson
olsonjs@umich.edu
ORCID iD: 0009-0008-1519-8888

© John S. Olson 2024

DEDICATION

To my grandfather Irving, for paving the way
and to my husband Ichal, for lighting it.

ACKNOWLEDGEMENTS

Aside from emails, this is the first piece of non-academic writing I have done in years. Amazing. Before graduate school, I wrote short stories – I also lived, I laughed, I dreamed. A Ph.D. program can suck that out of you. In many ways, it is akin to a brutal process of hazing, battling of egos, and trial by fire. Many of those sick of the environment in academia leave for much more pleasant careers, and existing systems become entrenched. So, we are stuck with a flawed system to acquire the skills necessary to produce and interpret original research, but it is the best one we have – and among the egos, there are angels.

It truly takes a village to navigate this journey. I will do my best to thank everyone who has supported me to this point, though I may inadvertently omit someone. Please know that I am eternally grateful for your role in helping me achieve my goal of earning a Ph.D. in economics, a journey I began nearly a decade ago.

I will start at the beginning: I wish to thank two teachers who had profound impacts on my upbringing during my K-12 education. The first is JoAnne Scott, my third-grade homeroom teacher, who pushed me to broaden my interests and worked to place me in challenging educational programs. The second is Laurie McHenry, my A.P. United States History teacher from high school. Laurie let me eat lunch in her classroom every day and let me rattle on about politics and economics when my peers could not care less. If it was not for her stoking of my passion for improving public policy, I would not be where I am today.

I also owe a debt of gratitude to the professors who took a gamble on a Kansan undergraduate aspiring to go straight into a Ph.D. program, perhaps against his best interests. For their mentorship, I thank Professors Donna Ginter, Jack Porter, Joseph Sicilian, and David Slusky at the University of Kansas. To any undergraduate reading this, when your advisors suggest gaining research assistant experience before pursuing a Ph.D., they are right.

Next for the graduate professors. My sincere thanks to my committee for their advice and support in bringing this ship home. I thank Christopher House, my chair, for his flexibility and willingness to work with students. My thanks to Ashley Craig, for never hesitating to give empathic, indispensable advice when he sees a student struggling. I thank Joel Slemrod for his wit and wisdom that have made me more knowledgeable about the study of taxation than I could have ever hoped to achieve. Although not on my committee, I thank Dmitriy

Stolyarov for his support in my pursuit of a career in line with economics education. My utmost thanks to committee member Sarah Miller, for her help in recusing my graduate career when I was at my lowest and fell off the bandwagon during the COVID-19 pandemic.

Speaking of which, having that in the middle of graduate school was a doozy. It was made all the worse by the University of Michigan's response to the pandemic at that time, forever tarnishing in my mind the reputation of the University and academia in general. In fact, the University's whole approach to interacting with student dissent is a mess. Nevertheless, I do appreciate the opportunity to study and teach at this institution, and for the insight it provided. I still think it necessary to provide this context, as I feel it was a major influence over the course of my graduate career.

Back to the task at hand: giving thanks. I thank my co-authors, without which the third chapter of this dissertation would not exist. Again, I thank Joel Slemrod, along with Laura Kawano and Meng Hsuan Hsieh. I thank Erik Randolph of the Georgia Center for Opportunity for his assistance in securing the data for my first chapter. I also wish to thank all the staff in the economics department, without which the department would cease to function – especially to Julie Hientz, who was a pleasure to work with during placement for the job market.

One of the largest blessings of graduate school was the opportunity to build new friendships with people from all over the world. For their friendship, advice, venting availability, sources of laughter, and so much more, I thank all the friends I made during my time here in Ann Arbor, in no particular order: Hayk Hambardzumyan, Florian Trouvain, Max Huppertz, Shwetha Raghuraman, Hayley Abourezk-Pinkstone, Keshav Garud, Stephanie Karol, Russell and Emilie Morton, Tereza Ranošová, and Oskar Gottlieb. There is also the friend who I started this graduate school journey with, and who supported me throughout (though from afar at a different program): my thanks to Grace Phillips. And to my pre-graduate school friends, Shegufta Huma, Victoria Calderon, Mara McAllister, Hannah Bannister Admussen, Anderson Wall, and Michael and Kelsey Dorian: thank you for your support and the laughs.

I am grateful to my family for their unwavering support. I especially thank my grandfather, Irving Johnson, whose generosity allowed me to pursue education as unburdened as possible, may he rest in peace. My thanks to my sister, Amy Sinclair-Roth, who started her Ph.D. journey before me and was indispensable along the way. And thank you to my father, David Olson, for his amazing ability to remind us all to touch grass.

Finally, words cannot fully describe the thanks I give to my husband, Ichal. Let it be known that there are many struggles that students may go through during their time at graduate school in the midst of a rigid and unforgivable process. For me, one struggle was having a closeted husband, whose full name I cannot use in my acknowledgments as he currently resides in a country halfway around the world with a less-than-stellar human rights record, and we continue to struggle with the United States immigration system. I will forever be grateful to everyone who acknowledged and accommodated this. Sometimes, I think that without Ichal's constant, unyielding support, I would not be alive, let alone finally finishing graduate school. So, thank you, Ichal. This dissertation would never have been finished without you.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

This dissertation aims to explore several features of tax-and-transfer systems that impact economic growth and opportunity.

In Chapter 1, "Benefits Cliffs in the Aggregate: Consequences for Welfare and Business Cycles", I study sudden decreases in public benefits that may occur with a small increase in earnings – which may create incentives that impact upward mobility. Given these concerns, what are the consequences of benefits cliffs for welfare and the response to aggregate shocks in general equilibrium? I find the aggregate implications of benefits cliffs on output are small, but welfare gains from their elimination are large and concentrated. Using the American Community Survey and proprietary data from the Georgia Center for Opportunity, I find that individuals in households approaching benefits cliffs reduce their working hours by about 40 hours annually on average. I then build a business cycle model that matches this result, where my model design allows me to accurately capture the benefits cliffs of the US tax and transfer system. Benefits cliffs create labor supply rigidity and thus attenuate the output response to productivity shocks. In a counterfactual model that smooths over benefit cliffs, output increases about 1.6% more on impact in response to an aggregate productivity shock compared to the baseline model with benefits cliffs, but the welfare gain to formally-constrained households doubles.

Chapter 2, "Place-Based Policy and Optimal Income Transfers in a Federalist Framework with Labor Elasticities in Three Dimensions," I ask what an optimal income transfer system would look like, considering the potential for both federal and state-level programs. I answer this question by building an optimal tax model that accounts for three margins of labor supply: participation, working hours, and mobility across states. I calibrate this model to the United States as a whole, as well as to individual US states. I find that, on average, states find it optimal to tax away federal income transfers, particularly when facing potential inter-state migration, reflecting fiscal constraints and a fear of attracting no or low-income earners. However, states with higher-income populations supplement federal transfers due to increased fiscal space. States with larger pools of no-income earners aim to increase the differences of consumption between those with no and earned income to encourage employment. Backing out implied welfare weights indicates that states must prefer more redistribution than the

federal government in order to rationalize their tax-and-transfer systems, aside from top, mobile incomes.

Finally, in Chapter 3, "A Transparent Look at How Taxes Affect Growth: Evidence from Cross-Country Panel Data" – joint with Meng Hsuan Hsieh, Laura Kawano, and Joel Slemrod – we review the literature that estimates the effect of tax policy on economic growth using cross-country panel data and, in our own analysis, evaluate how different methodological choices affect the conclusions drawn. We find these analyses do not credibly support claims that tax rate changes have a statistically robust medium-term impact on national income. We further assess this literature in light of the recent econometric insights on estimation with staggered treatments. We show why the commonly-used linear projection approach yields biased estimates in this setting, and find that a causal estimate of the effect of tax rate changes again yields no statistically significant effect on economic growth at a five-year horizon.

CHAPTER 1

Benefits Cliffs and Aggregate Fluctuations

1.1 Introduction

Benefits cliffs are pervasive throughout the United States' transfer system. These cliffs occur when increasing one's earned income leads to a sharp reduction in consumption due to higher taxes or a reduction in transfers. In other words, benefits cliffs cause discontinuities - or notches - in the budget constraints of affected individuals. These abrupt phase-outs may discourage families from increasing take-home pay, contributing to a cycle of poverty and reducing upward mobility for precisely those who need it most. For example, Figure 1.1 presents the statutory benefits for a family of 4 living in Texas, with pre-tax/transfer earnings along the x-axis and post-tax/transfer earnings along the y-axis. If this family earned around $5,000 annually, earning an extra dollar would cost them nearly $10,000 in benefits due to the combined loss of Medicaid and cash transfers for adults in the household.[1]

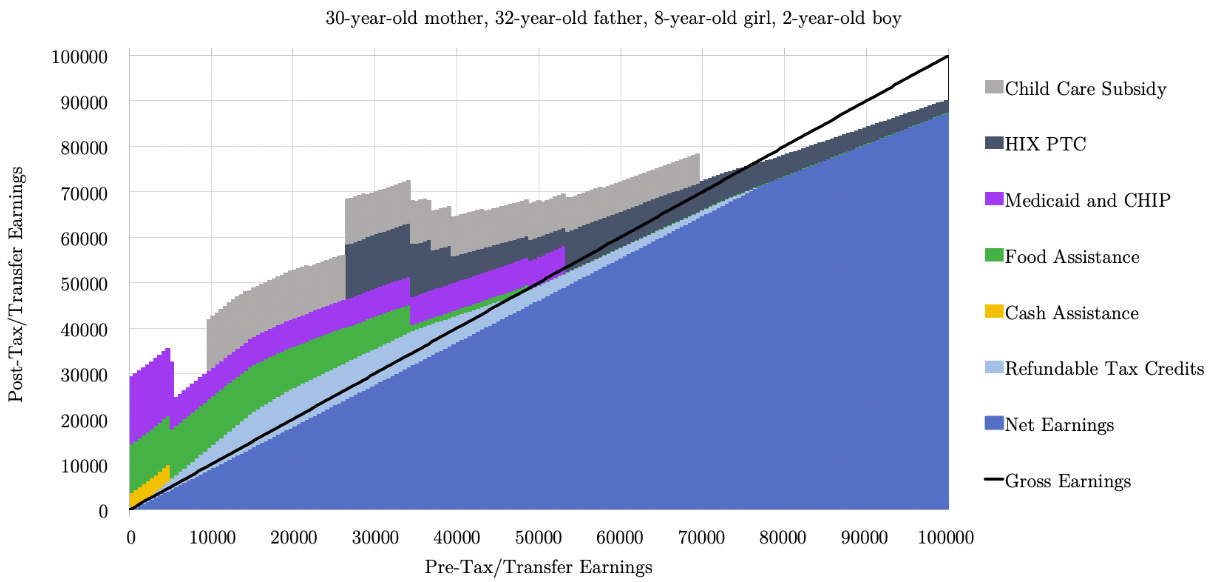
Benefits cliffs may not only directly harm millions of families hitting them as incomes rise, but negatively impact growth in the business cycle. By reducing mobility in labor markets and affecting government revenues, benefits cliffs may impact households both on and off cliffs. Costs for those near cliffs in individual benefit programs in static settings are well-documented,[2] but in addition there are dynamic costs as rising wages cause encounters with cliffs, and costs to governments as reduced labor supply impacts tax collections. Failing to account for these extra costs of cliffs may vastly understate the real costs of leaving these cliffs in place.

In this project, I evaluate how benefits cliffs impact aggregate fluctuations induced by positive productivity shocks and document their welfare implications. I do this by first estimating the effect of benefits cliffs on intensive-margin labor supply across multiple programs

[1] In this project, I currently exclude Section 8 housing subsidies and vouchers due to it being a rationed program with very limited take-up rates. See the data Appendix A.1.1 for a description of how in-kind benefits are valued.

[2] See handbook chapters by Moffitt (2002) and Blank (2002).

Figure 1.1: Pre/Post-Tax-and-Transfer Earnings for a Family of 4 in Texas, 2019



Note: "HIX PTC" stands for health insurance premium tax credits as part of the Affordable Care Act. CHIP is for the Children's Health Insurance Program. The figure excludes Section 8 housing vouchers and subsidies. Values are in 2020 dollars.

simultaneously, and then calibrating a dynamic stochastic general equilibrium (DSGE) model of the economy that mirrors these results. In the data, I find that benefits cliffs reduce working hours of individuals near cliffs by about a full-time week's worth of work compared to similar workers further away from cliffs. I then compare outcomes from the baseline model to a counterfactual that removes benefits cliffs in the tax-and-transfer system and replaces them with a smooth approximation. I find that the impact of benefits cliffs on aggregate output is small, but agents constrained by cliffs miss out on potentially large welfare gains.

I begin by outlining a simple partial-equilibrium model of a worker facing a benefits cliff, which leads into my empirical findings. I illustrate that for rational agents aware of the benefits cliff, wage increases may induce reductions in working hours so that workers can maintain their after-tax/transfer consumption. Moreover, I demonstrate that given uncertainty about income, workers near benefits cliffs will preemptively reduce - or "hedge" - their working hours to reduce their risk of falling over the cliff.

I then see if these theoretical findings bear true in the data, finding persons in households just below cliffs reduce their working hours by about 40 hours annually (or about 2% for a full-time worker), on average. Unlike much of the previous literature, I estimate the effects of benefits cliffs in the universe of transfer programs simultaneously: for county, state, and federal programs in 9 southern US states. I am able to do so with the aid of a novel data set from the Georgia Center for Opportunity, which outputs statute-based benefits depending on family composition and other demographic characteristics.

I next construct a DSGE model, a key contribution of which is an environment that allows for benefits cliffs and nonparametric tax-and-transfer systems. This stands in stark contrast to the majority of DSGE models of fiscal policy, which commonly feature smooth, parametric, monotonic tax-and-transfer systems.[3] I accomplish this using a micro-founded result that agents at cliffs will be inframarginal, obeying the rational response to reduce working hours in response to wage increases in order to maintain their consumption. Combined with discrete - as opposed to a continuum of - heterogeneous agents, this yields a computationally efficient, tractable model best suited for small aggregate shocks that mirrors my empirical findings. This model nonparametricly matches the actual tax-and-transfer systems much better than the smooth functions common in DSGE models. Moreover, the use of heterogeneous agents allows me to observe outcomes throughout the income distribution, including for those away from benefits cliffs.

I then calibrate a counterfactual tax-and-transfer schedule that smooths over benefits cliffs and proscribes the existence of inframarginal agents, which I use to compare outcomes in response to aggregate productivity shocks, paying particular attention to effects on output and consumption-equivalent welfare. The design of the alternative tax-and-transfer system plays an important role, as policymakers face trade-offs between reducing benefits for some to eliminate benefits cliffs for few. In response to an aggregate productivity shock, output on impact improves nearly 1.6% compared to the baseline (e.g. a 3% output gain on impact from a productivity shock in the baseline model grows to 3.12% in the counterfactual), and the improvement in aggregate welfare on impact increases over 9%. Welfare gains in response to productivity shocks are particularly large for formerly-constrained households who were at cliffs in the baseline, as they no longer reduce their hours to avoid falling over cliffs: for the counterfactual considered, the improvement in welfare on impact is 181% (e.g. a 0.75% gain in baseline to a 2.11% gain in the counterfactual).

I conclude with policy implications stemming from these exercises. Although aggregate effects are small, the potential dynamic welfare gains for low-income households merits consideration by lawmakers. Removing benefits cliffs and replacing them with phase-outs involves trade-offs between benefits for some households, effective marginal tax rates for others, and government expenditures. Lawmakers should be cognizant of other programs that phase out over similar ranges of income, as multiple programs phasing out at once can generate undesirably high effective marginal tax rates. Finally, other policy reforms, such as phasing out benefits with time rather than income, may help reduce the impact of benefits cliffs on

[3] See handbook chapters by Krueger et al. (2016) and Kaplan and Violante (2018).

households.

1.1.1 Contribution to Related Literature

The existence of benefits cliffs in general is well-documented, and are caused by government programs that end abruptly above a certain threshold of income or multiple programs phasing out over the same range of income.[4] A simple web search for "benefits cliffs" reveals hundreds of blog posts and white papers from numerous think tanks documenting cliffs in federal and state programs. In their working paper, Altig et al. (2020) incorporate all major federal and state fiscal policies for the United States into a life-cycle model, documenting numerous cases where for a $1,000 increase in earned income, simulated workers face effective marginal tax rates in the thousands of percentage points. Similarly to the white papers, they argue that this locks people into poverty.

This paper contributes to the macroeconomics of public finance by providing a tractable, computationally efficient business cycle model that features benefits cliffs. There has been a recent surge in the use of non-linear tax functions in macroeconomic models. Heathcote et al. (2017) develop a highly tractable non-linear tax function for use in many classes of macroeconomic models that fits survey data, especially well for higher incomes. This function has been implemented in following works, particularly in the literature on heterogenous-agent New Keynesian (HANK) models (McKay and Reis, 2021) and other models of heterogeneous agents (Heer and Rohrbacher, 2021; Brendler, 2023). DeBacker et al. (2019) developed a method to integrate tax rates from a partial-equilibrium microsimulation model into a dynamic general equilibrium model, allowing for nonlinear tax rates in both labor and capital income. However, none of these functional forms allow for effective marginal tax rates exceeding 100%, abstracting away from the presence of benefits cliffs entirely.[5] My model environment instead uses the data to nonparametrically set tax rates and benefits for heterogeneous households, and allows for cliffs in benefit schedules in a business-cycle model.

[4]While not the focus of this paper, reasons for these policy choices are worth highlighting. One is perceived simplicity: lawmakers may find it convenient to cut off a program above a certain income, rather than phase it out. These same lawmakers may not even be unaware of the potential perverse incentives created by benefits cliffs. Moreover, the diverse array of government programs has almost as many definitions and standards, few of which were designed to account for those of other programs (National Conference of State Legislatures, 2019). In other words, in designing these programs, there may be additional costs that are not fully internalized.

[5]Three exceptions to this are Ventura (1999), Altig and Carlstrom (1999), and Moore and Pecoraro (2020b), who use non-smooth tax-and-transfer systems. There is also earlier work featuring non-linear tax systems - see Easterly and Rebelo (1993), Gouveia and Strauss (1994), Benabou (2002), and Li and Sarte (2004) - but these still generate smooth functions. Moreover, all these papers - including those with non-smooth systems - study general equilibrium and growth implications, but do not analyze the impact on business cycles.

This paper also connects to an abundant literature in empirical and theoretical public finance. It adds to the studies of cliffs in individual benefit programs and income tax schedules - examples include Kleven and Waseem (2013); Hamersma (2013); Ruh and Staubli (2019) - which often find that notches induce large behavioral responses from small structural elasticities. This work also touches on simulations of multiple-program participation with non-convex budget sets found in the literature (Blundell et al., 2016; Flood et al., 2004), given my empirical analysis of all benefit programs simultaneously.

Using the DSGE model, I also contribute to the literature on the efficiency and welfare costs of notches in tax-and-transfer systems (Blinder and Rosen, 1985; Sallee and Slemrod, 2012; Slemrod, 2013). These papers find that generalizations about the welfare costs of notches cannot be made without considering what alternative schedules are available given a government's tax instruments, an important thread to keep in mind when considering counterfactual simulations. In fact, it is possible that notches in tax-and-transfer systems can *improve* welfare. Blinder and Rosen (1985) provide simulations of cases where a notched system can increase consumption of certain desirable goods at a lower efficiency cost than a linear system. Under very particular assumptions, including the nature of the chosen alternatives, benefits cliffs may perhaps induce a more desirable allocation of leisure. Thus, whether benefits cliffs improve welfare may be an open question. Bringing the analysis into the land of DSGE allows us to quantify any potential welfare gains, as well as providing new insights into the costs and benefits of cliffs. In addition, using such a model yields measures of cliffs' impacts on the business cycle and how counterfactuals may affect those not experiencing such cliffs in the first place.

I study the effects of benefits cliffs on the intensive margin of labor supply in terms of annual hours worked, abstracting away from the extensive margin of employment. I do this because the popular concern surrounding benefits cliffs is almost entirely localized on the intensive margin - a review of the aforementioned blog posts and white papers supports this, as does much of the aforementioned literature concerning notches in individual benefit programs.[6] Moreover, this is where the anecdotal evidence from individuals with lived experiences with benefits cliffs points us - they are very often eager employees but constrained in career advancement.[7]

[6]One example encompassing benefits cliffs across multiple programs is Altig et al. (2020), which documents examples where, due to loss of means-tested public benefits, individuals can be financially worse off in the short-run as careers advance. Most work on the effects of benefits on the extensive margin of labor supply in the United States concerns the Earned Income Tax Credit, which tends to find null to positive effects - see Kleven (2019) and Nichols and Rothstein (2015).

[7]Roll and East (2014) find that 33% of low-income families in Colorado declined opportunities to increase income to avoid losing childcare subsidies. From a US Chamber of Commerce summary of the paper, "Many of these parents knew precisely how much income they could have before they lost their subsidies and estimated that it would take receiving a raise of at least $4 an hour for them to risk losing their childcare

I also focus on the effects of benefits cliffs as opposed to other discontinuities in the budget constraint. An inspection of Figure 1.1 reveals discontinuities where the tax/benefit schedule jumps upwards - so-called "peaks" as opposed to cliffs - largely caused by eligibility for health insurance subsidies at 100% of the poverty line. In theory, peaks create incentives for individuals to *increase* their working hours more than they otherwise would. However, the study of the effects of these peaks is beyond the scope of this paper, as the popular concern revolves around cliffs.[8] Moreover, these peaks are not nearly as prevalent in the 40 states that expanded Medicaid under the Affordable Care Act, which fills in benefits before households become eligible for health insurance subsidies.

The paper proceeds as follows: In Section 1.2, I use a simple model to motivate my empirical investigation. I then briefly describe the data before investigating the effects of benefits cliffs on intensive-margin labor supply. In Section 1.3, I construct a DSGE model that features benefits cliffs and whose calibration matches both the tax-and-transfer system in the data and the main result of the empirical exercise. Section 1.4 presents and discusses a counterfactual policy simulation and model extensions. Section 1.5 considers policy implications and concludes.

1.2 The Relationship Between Benefits Cliffs and Labor Supply

In this section, I aim to uncover the relationship between benefits cliffs and intensive-margin labor supply. In contrast to previous literature, I study responses to the universe of benefit programs simultaneously, rather than any one specific program. I accomplish this through the use of a novel data set from the Georgia Center for Opportunity (GCO), combined with publicly-available data from the American Community Survey (ACS). I first discuss the theory, followed by the data, empirical methodology, and results.

1.2.1 Theory

To motivate my empirical specification and future modeling choices, I start by analyzing labor supply responses on the intensive margin. I begin with several simplifying assumptions:

subsidy... most parents interviewed said they turned down raises or adjusted work hours to avoid losing their subsidy." See Appendix A.1.7 for additional anecdotal evidence.

[8] Peaks, in contrast to cliffs, should encourage labor supply in the steady state. Many of those at these peaks would not be inframarginal in response to positive aggregate shocks, but in response to negative shocks they may wish to increase their labor supply to maintain their benefits. The focus of this paper is on responses to positive aggregate shocks, for which cliffs may be binding.

a homogeneous structural labor elasticity in the population, no optimization frictions, and certainty about current income - the last of which I will soon relax. In contrast to Kleven and Waseem (2013), I will allow wages to change and a more general form of utility that includes income effects, and later allow for uncertainty. Individuals i at time t choose hours $h_{i,t}$ and consumption $c_{i,t}$ given a wage w_t and ability a_i to maximize their utility subject to a budget constraint:

$$u = \frac{c_{i,t}^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \psi \frac{h_{i,t}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \tag{1.1}$$

subject to

$$c_{i,t} = a_i h_{i,t} w_t - T\left(a_i h_{i,t} w_t\right) \tag{1.2}$$

where $T\left(a_i h_{i,t} w_t\right)$ is tax liability at time t, σ is the elasticity of intertemporal substitution, ψ governs the distate for work, and ϵ is the Frisch elasticity of labor supply. If the tax system is linear so that $T(a_i h_{i,t} w_t) = \tau a_i h_{i,t} w_t$, solving the individual's problem yields

$$h_{i,t} = \left(\frac{c_{i,t}^{-\frac{1}{\sigma}}\left(1-\tau\right) a_i w_t}{\psi}\right)^{\epsilon} \tag{1.3}$$

so that agents' hours decisions depend on their ability, the wage, the tax rate, and the Frisch Elasticity.[9]

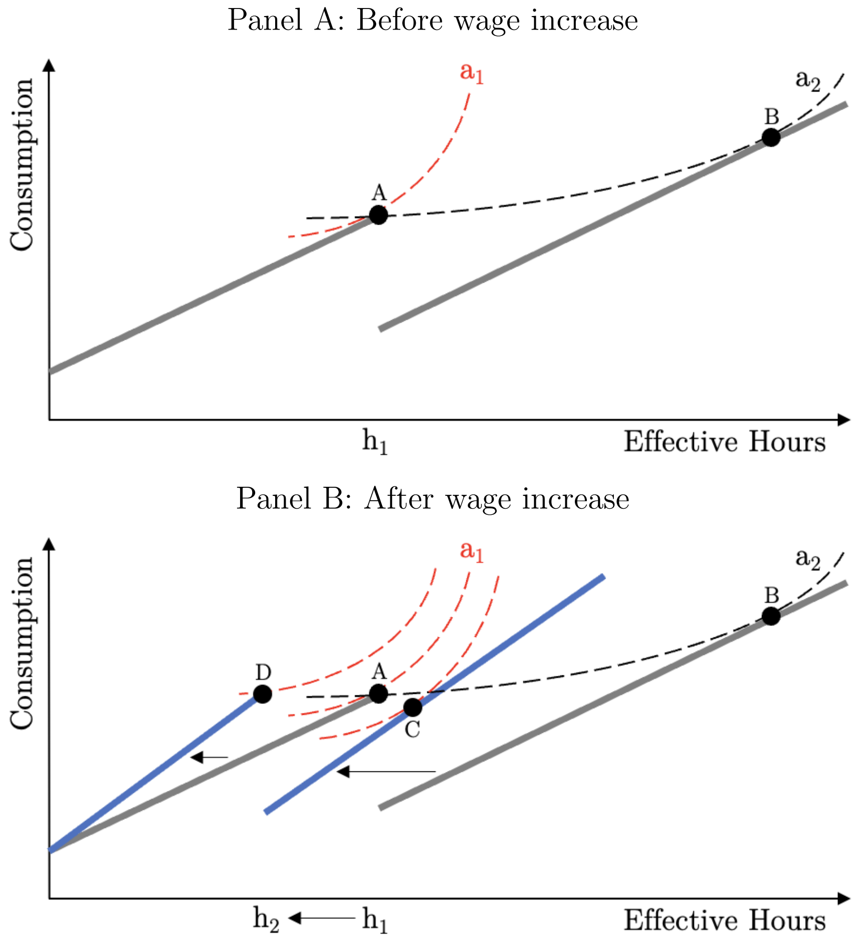
I now introduce a downward notch - a discrete change in after-tax/transfer earnings - of size T at pre-tax/transfer income z_i and analyze how this affects agent locations along the budget constraint. Thus, the tax function is now

$$T(a_i h_{i,t} w_t) = \tau a_i h_{i,t} w_t - T\mathbf{1}(a_i h_{i,t} w_t > z_i) \tag{1.4}$$

where $\mathbf{1}(\cdot)$ is an indicator function for being above the cutoff, so that individuals above the cliff lose access to T benefits. Figure 1.2, Panel A illustrates this scenario in a consumption-earnings indifference curve diagram with a negative notch, where effective hours worked $a_i h_{i,t}$ is along the x-axis and consumption from after-tax/transfer earnings is along the y-axis. For a given wage, the notch introduces a discontinuity in the budget constraint at a certain number of effective hours, labeled in the figure as h_1. The distribution of ability is one factor

[9]Note that this result is not closed-form, given that $c_{i,t}$ is a choice variable. This occurs because of allowing for income effects. Assuming no risk aversion so that $\sigma = \infty$ would cause $c_{i,t}$ to drop out and yield a closed-form solution.

Figure 1.2: Consumption-Earnings Diagrams with a Cliff



Notes: Panel A shows an indifference curve diagram for a tax-and-benefit schedule with a single notch. Panel B shows the same, but with an increase in the wage. Indifference curves are shown for agents with varying levels of ability a. Point A is the notch point before the wage increase, point B is the point at which the agent with the highest productivity that locates at the notch is indifferent to, point C is a hypothetical increase in working hours after an increase in the wage, and point D is the notch point after the wage increase.

that governs the shape of indifference curves: the higher the idiosyncratic productivity a_i, the less averse the household is to work and the flatter the indifference curve. The household with the lowest productivity that locates at the notch, given as a_1, would have located there in the absence of the notch, at point A. The household with the highest productivity that locates at the notch, given as a_2, is exactly indifferent between locating at the notch or beyond the notch point at point B. All households with idiosyncratic productivities between a_1 and a_2 locate at the notch point, as utility at these points dominates the utility of staying in place.

In this environment, a wage increase will cause rational agents at the cliff to reduce their working hours. Consider now an increase in the wage, as in Figure 1.2, Panel B. The budget constraint tilts as the opportunity cost of leisure rises. Equation 1.3 implies that under a smooth tax system, all agents would increase their hours as long as substitution effects dominate income effects. However, in our context, a notch still occurs at the same level of pre-tax/transfer income as before, but now at a lower number of effective hours given the higher wage - in this case, h_2. The household with the lowest productivity that locates at the notch, a_1, would be worse off maintaining or increasing their working hours, for example at point C. Instead, they will maintain their pre-tax/transfer income and decrease their working hours in order to remain precisely on top of the cliff, at point D. Thus, these inframarginal agents set their working hours according to

$$h_{i,t} = \frac{z_i}{a_i w_t} \tag{1.5}$$

where z_i is the pre-tax/transfer earnings location of the notch. ceteris paribus, with an increase in the wage, the labor supply of these households decreases. In fact, all households with abilities $[a_1, a^*)$ will reduce their working hours, where a^* is a critical value of ability $a^* \in (a_1, a_2]$. Agents above this critical value of ability will increase their hours in response to a positive wage shock, "leaping" past the dominated region of consumption. This critical value is trivially important given the few amount of agents that would make this leap, and the desire for the eventual model to provide upper-bound estimates of the detrimental effects of benefits cliffs. I elaborate on these points in Appendix A.2.3.

Uncertainty in this environment also reduces labor supply.[10] Consider a person with household income below the threshold of a single cliff. Suppose that individuals are considering labor supply $h_{i,t}$ before the cliff as given[11] and pre-tax/transfer income $y_{i,t}$ is uncertain, where

$$y_{i,t} = a_i w_t h_{i,t} + \theta_{i,t} \tag{1.6}$$

where $\theta_{i,t} \sim F$ with the probability density function f is an income shock. For simplicity, I assume that this shock is lump-sum. Let $\hat{\theta}_{it}$ be the income necessary to reach the income at the nearest cliff ahead z_i, or

$$\hat{\theta}_{i,t} = z_i - a_i w_t h_{i,t} \tag{1.7}$$

[10]This uncertainty may come from a variety of sources - tips for workers in the hospitality industry, child support payments, unexpected demands from bosses to cover shifts, etc.

[11]This simplifies the problem and allows for a more closed-form solution for $h_{i,t}$, compared to an analysis where agents are entirely uncertain about their hours choice.

which I will also refer to as "distance to the cliff." In Appendix A.2.1, I show that under the simplifying assumption that $\sigma = \infty$ so that agents are risk-neutral about consumption and the marginal utility of consuming is always 1, the hours decision for individuals beneath the cliff in this simplifies to

$$h_{i,t} = \left\{ \frac{a_i w_t}{\psi} \left(1 - \tau - T f(\hat{\theta}_{i,t}) \right) \right\}^{\epsilon} \tag{1.8}$$

where $f(\hat{\theta}_{i,t})$ is the density of income shocks at the income necessary to reach the cliff. This is the normal labor supply condition under constant marginal utility, with the addition of an effect $-Tf(\hat{\theta}_{i,t})$ from the threat of reduced consumption after the cliff. With no cliffs, $T = 0$ and equation 1.8 collapses to the normal labor supply condition. Likewise, if the distance to the cliff is outside the range of income shocks, then $f\left(\hat{\theta}_{i,t}\right) = 0$, the last term of equation 1.8 drops out, and we recover the normal labor supply condition again. Otherwise, there is a negative effect on the hours decision before the cliff induced by consumption risk of falling over the cliff. In other words, individuals "hedge" their working hours to avoid losing benefits. Moreover, there is a testable prediction that as the cliff size T increases, ceteris paribus, the hours chosen decline.

Appendix A.2.2 shows that by assuming a distribution of income shocks, one can log-linearize equation 1.8 to reveal a relationship between the hours choice and distance to the cliff:

$$\ln(h_{i,t}) = \alpha_{i,t} + \epsilon \ln(w_t) + \beta \ln(\hat{\theta}_{i,t}) \tag{1.9}$$

where β, the effect of distance to the cliff on labor supply, is a nonlinear combination of the steady-state distance to the cliff $\hat{\theta}^S_{i,t}$ and parameters including the Frisch elasticity ϵ, the tax rate τ, and those that govern the distribution of income shocks. As the distance to the nearest cliff ahead $\hat{\theta}_{i,t}$ declines, ceteris paribus, the hours choice falls. This motivates my empirical investigation in Section 1.2.3, where I examine the relationship between labor supply and distance to the individual's nearest cliff.

1.2.2 Data

My primary dataset is from the GCO, hereafter referred to as the Benefits Cliffs Workbook (BCW). This is a program that generates tax-and-benefit schedules akin to Figure 1.1. Importantly, it does so for the universe of social programs in every county in nine southern US states: Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Texas. These schedules are generated for the years 2020-2021 based on a plethora of demographic information: whether it is a one- or two-parent household, the number of children, sexes, ages, pregnancy, disability, and childcare statuses (e.g. using a licensed center). Another distinguishing feature is that the BCW bases its schedules on statute: that is, the taxes and benefits are assigned according to the letter of the law, abstracting away from take-up.[12] Thus, I will study people's reactions to statutory benefits cliffs, and address take-up in later sections.

I combine this dataset with every household in the 2015-2019 ACS that resides in the sample states. I generate tax-and-benefit schedules by mapping all available demographic and geographic variables in the ACS to those reported in the BCW. Given that the BCW covers the years 2020-2021, I exclude pandemic legislation from the tax-and-benefit schedules and implicitly assume that - aside from inflation adjustments where applicable - benefits programs in the sample states did not change substantially from 2015-2019.[13]

The constructed dataset contains two important sources of variation that I will use to discern the impacts of benefits cliffs, and I limit the sample to study the intensive margin of labor supply. There is variation across households within the same geographic subunit, and across geographic subunits with the same type of household. Differences in state and county policies create variation in benefit amounts across geography for the exact same type of household and income. Meanwhile, differences in family composition change the size and composition of benefits cliffs within the same geographic unit. Given my focus on the intensive margin of labor supply, I restrict the sample to persons aged 18-64 who report strictly positive wage income and annual working hours.[14] Appendix A.1.1 provides additional details on the construction of the dataset, valuation of benefits, variation, and descriptive statistics.

1.2.3 Methodology

The simple model described in Section 1.2.1 above motivates my main empirical specification, which aims to study the effects on annual hours worked of benefits cliffs, particularly the effects on the labor supply of those near cliffs who may face uncertainty about their income. Equation 1.9 suggests that annual hours worked are a function of distance to the

[12]Even if families are eligible and attempt to take up benefits, there are cases where this still may not be possible - a prime example is the rationed Section 8 housing program. I exclude that program from this analysis.

[13]An important exception to this is Louisiana expanding its Medicaid program in 2016. I currently address this by dropping pre-2016 observations for this state.

[14]While beyond the scope of this paper, I suspect extensive-margin responses would result in a larger effect of cliffs on aggregate fluctuations as people switch in and out of employment rather than adjust working hours. This aspect will be incorporated in future work.

cliff. To observe how hours change relative to the cliff, I bin observations by their distance to the nearest cliff and estimate the following regression:

$$H = \alpha + \sum_{b=1}^{n} \beta_b B_b + \gamma X + \delta_c + \delta_y + \delta_c \times \delta_y + \varepsilon \tag{1.10}$$

where H is a household member's reported annual hours worked, α is the intercept, B_b is the household's income bin relative to the nearest cliff, X is a vector of demographic and tax system controls, including marginal tax rates, wages, age, sex, race, marriage, education, and number of kids. δ_y and δ_y are year and county fixed effects, respectively, and ε is the error term. I follow the methodology of Hamersma (2013) and Haider and Loughran (2008) in that individual households are normalized by their distance to their cliff.[15] Here, β_b are our coefficients of interest, and is the difference in hours worked relative to an omitted bin. For those aware of and behind their nearest cliff, β_b corresponds to β of equation 1.9 in the motivating model of 1.2.1. If the negative effect on hours increases as individuals get closer to the cliff, we should expect negative coefficients on bins approaching the cliff in a regression that omits the bin at the cliff. Bin widths are set to $5,000 and standard errors are clustered at the county level.[16]

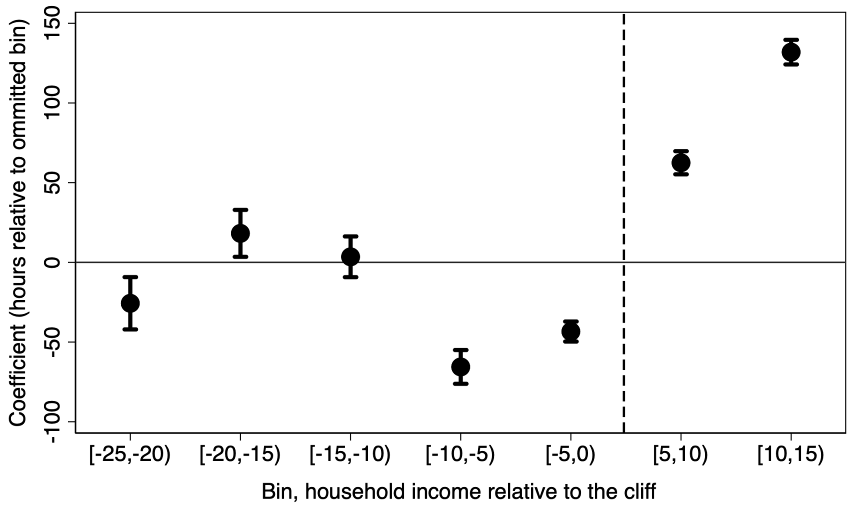
I also explore heterogeneous responses to cliff size for evidence of real behavioral responses. Equation 1.8 suggests that labor supply responses should increase in the size of the cliff. Thus, I will divide the sample into quartiles of benefits cliffs as a percentage of household pre-tax/transfer income, and estimate regressions on each subsample. If persons reduce their annual hours worked more when their household faces larger cliffs, coefficients B_b should be more negative just below the cliff for samples with larger cliffs.

In light of other empirical work about benefits cliffs, this scenario also seems ripe for a bunching exercise, but the results from this are limited for several reasons. One, since I work with survey data, measurements may be imprecise. Two, limitations on awareness and take-up of benefit programs may attenuate any bunching we see before statutory cliffs - while also affecting my estimates, to be addressed below. Finally, many cliffs have relatively flat earnings schedules leading up to them - a sort of "benefits plateau" before the cliff - so that effective marginal tax rates are already elevated in many regions before cliffs. This itself creates disincentive effects for increasing incomes to locate right before the benefits cliff. See

[15] These papers study earnings responses to earnings limits in Medicaid and Social Security, respectively, which also create notches in budget constraints. Note that this normalization is not in terms of absolute value - so that for values of -$5,000 and $5,000, individuals are $5,000 behind or ahead of their nearest cliff, respectively. An alternative specification considers only the nearest cliff ahead, and is described below.

[16] I set bin widths to $5,000 due to many of those surveyed in the ACS rounding their household incomes to the nearest $5,000.

Figure 1.3: Annual Hours Worked Relative to Cliffs



Notes: Results for the binned regression of equation 1.10, which regresses total reported annual hours on bins of household income relative to their nearest cliff. The omitted bin contains the cliff and is represented by the dashed line. 95% confidence intervals are shown. All standard errors are clustered at the county level.

Appendix A.1.6 for bunching results and discussion.

I also gauge the robustness of the results to this binned methodology by employing two additional specifications: a regression-discontinuity design (RDD) and a system of equations estimated by generalized method of moments (GMM). The details of the RDD are in Appendix A.1.3. Motivating the system of equations is the fact that regression 1.10 is a reduced-form that shows the statistical relationship between the hours choice and distance to the cliff. For a more causal effect of distance to the cliff on hours, which may be of interest, I must address the endogeneity between hours decisions $h_{i,t}$ and distance to the nearest cliff ahead $\hat{\theta}_{i,t}$ evident in equation 1.7. Further still, marginal tax rates τ are also endogenous to the hours decision, since income choices affect marginal tax rates. To account for this endogeneity, I jointly estimate a system that includes these interactions between distance to the cliff, the hours decisions, and tax rates. Details and results for this are in Appendix A.1.4.

1.2.4 Results and Discussion

Main Results: I find that there is an apparent reduction in working hours just before benefits cliffs, by about a full-time week's worth of work. The results for the coefficients of interest from specification 1.10 are shown in Figure 1.3, around which are 95% confidence intervals. A person's bin of household income relative to their closest cliff is along the x-axis, and coefficients for each bin are relative to the omitted bin of just after the cliff, represented by the vertical dashed line. Here, being within \$10,000 behind a benefits cliff is associated with an average reduction of about 43-65 hours annually [SEs: 3.2-5.4]. For the average wages in these bins, this translates to just over about \$801-\$1,212 in foregone earnings, or about 3.1-5.4% of household income on average.

Robustness: The RDD and system of equations in appendices A.1.3 and A.1.4, respectively, yield generally conforming results for bins of interest. The RDD implies a reduction of working hours of 12.81-29.73 [SEs: 4.25-3.15] annually just before the cliff. For household incomes within \$10,000 of the cliff, the resulting estimates from the system of equations imply a 52.59-79.75 [SEs: 13.54-16.97] reduction in annual hours worked.

Appendix A.1.5 also explores heterogeneous reactions to cliff sizes to determine if behavioral responses are driving the negative effect on hours, and finds that the larger the cliff size as a percentage of household income, the greater the reduction in hours. Across all specifications, the reduction in working hours for those in the fourth quartile of cliff sizes (about 15% of household income and above) is at least twice that of those in the first quartile of cliff sizes (up to 1.75% of household income), up to over 67 hours annually. I interpret this as evidence of benefits cliffs driving the results.

Accounting for take-up: The results thus far abstract away from the issue of take-up, instead observing the effects of statutory benefits cliffs - a sort of intent-to-treat effect that I now adjust to account for take-up. The above regressions may be interpreted as the net effects of the statutory benefits and take-up. Work by Hernanz et al. (2004) and Ko and Moffitt (2024) show that take-up rates of benefit programs in the United States roughly span the range of 40-80%. I use a back-of-the-envelope calculation to adjust for the effect of lower compliance rates. For compliance rates of 40-80% and an apparent reduction in working hours of 43 hours annually, reductions in working hours span about 54 to 107 hours annually for those who take up benefits programs.

Comparison to other studies: So far, the headline result is an apparent reduction by about one full-time week's worth of work in response to statutory benefits cliffs. Is this result reasonable? To answer this question, I turn to estimates of the elasticity of taxable income (ETI) from the literature. This literature produces estimates for the ETI of top incomes that generally range from 0.12 to 0.4 (Saez et al., 2012), with lower-income households having lower ETIs of about 0.1-0.28 (Gruber and Saez, 2002). To uncover the ETI in my environment, I use the reduced-form approach from Kleven and Waseem (2013), which relates the earnings response to the implicit change in the effective marginal tax rate induced by the cliff. That is,

$$e_R \approx \left(\frac{\Delta z}{z}\right)^2 \frac{1-\tau}{\Delta\tau} \tag{1.11}$$

where e_R is the reduced-form elasticity of taxable income, Δz is changes in earned income z, and $\Delta\tau$ is the change in the effective marginal tax rate τ. The formula treats the cliff as a hypothetical kink that creates a large jump in the effective marginal tax rate, with Kleven and Waseem (2013) arguing that it provides an upper-bound for the true structural earnings elasticity. Given the cross-sectional nature of the data, I will make some back-of-the-envelope adjustments. First, I do not observe pre-marginal-rate-change earnings, only those earnings already in place. However, I do generate an implied reduction in earnings via a reduction in working hours, as detailed above. I can use this implied change in earnings to uncover the original earnings-level z. Changes in the effective marginal tax rate come from the cliff size of those in bins \$10,000-\$0 before the cliff.[17] I restrict estimates of the ETI to those in these income bins, as they appear to be those responding to benefits cliffs. From the estimates from the system of equations, the resulting population-weighted ETIs are 0.114 for those who report as working for wages and 0.264 who report as self-employed, or 0.129 overall. These point estimates are consistent with the estimates of low ETIs for incomes below \$100,000 (Gruber and Saez, 2002), and the larger literature on ETIs between those self-employed and not (Saez et al., 2012).

A note of caution: There may be a temptation to interpret the sum of coefficients in Figure 1.3 as the net effect of benefits cliffs - that is, the apparent increases in working hours after the cliff reduce in aggregate the loss of hours from those just below the cliff - but assuming this to be completely true this would be an inaccurate interpretation. There are many high-income households who work thousands of hours a year that will be naturally placed thousands of dollars ahead of cliffs. Persons in such households may not even consider benefits cliffs that take place thousands of dollars behind their household income levels. This automatically generates strongly positive point estimates substantially beyond the cliff. I illustrate this in Appendix Figure A.4: there is a natural association between hours worked and income, though plotting hours against the nearest cliff yields a dip in annual hours worked.[18]

[17]For example, if their nearest cliff implies earning an extra \$500 in earnings yields a \$500 *decrease* in after-tax-and-transfer income, the implied marginal tax rate is 200%.

[18]The converse is also true: there are many low-income households whose low marginal productivities naturally place them below the cliff. Appendix Figure A.4 shows these households are substantially behind their cliffs. Moreover, there is evidence that increasing the cliff size is associated with larger reductions in working hours - if households just before cliffs were not responding, we would not see this effect.

1.3 Quantitative Model

Next, I develop a business cycle model that can quantify the dynamic effects throughout the income distribution that arise from the negative effects of benefits cliffs on labor hours explored in Section 1.2. The goal of the model is to deliver upper-bound effects of benefits cliffs on macro dynamics: that is, the maximum, yet reasonable detrimental effects, if any, of benefits cliffs on aggregate fluctuations. The same model can be used to quantify impacts on welfare. The model is calibrated to nonparametricaly match the statutory tax-and-transfer systems of states in the sample and the observed drop in hours just before benefits cliffs. The key features of the model are (i) discrete, heterogeneous households, (ii) inframarginal labor supply decisions by some households, and (iii) highly nonlinear, nonparametric tax-and-transfer systems.

1.3.1 Households

The economy is populated by $i = 1, ..., I$ households, of which there are $j = 1, ..., J$ types. A share κ_{ij} of every household type are liquidity-constrained and operate as hand-to-mouth (HTM) consumers, whereas $1 - \kappa_{ij}$ households are Ricardian and have access to a bond market. Every household has its own idiosyncratic productivity, a_i. The type of the household, j, is different for each combination of filing status (e.g. single, married filing jointly, etc.) and number of children. At date $t = 0$, Ricardian households i choose consumption $c_{i,t}$, labor supply $h_{i,t}$, and bond-holdings $b_{i,t+1}$ to maximize expected lifetime utility:

$$\max_{\{c_{i,t}, h_{i,t}, b_{i,t+1}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left(\frac{c_{i,t}^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \psi \frac{h_{i,t}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \right) \tag{1.12}$$

where where β is the discount factor, σ is the elasticity of intertemporal substitution, ψ governs the distaste for work, and ϵ is the Frisch elasticity of labor supply. Households are subject to a budget constraint:

$$c_{i,t} + b_{i,t+1} + \frac{\mu}{2} \left(b_{i,t+1} - \bar{b} \right)^2 = a_i h_{i,t} w_t - T_j \left(a_i h_{i,t} w_t \right) + B_t + (1 + r_t) \, b_{i,t} \; \forall \; t \tag{1.13}$$

where μ and $\bar{b}$ govern debt-holding costs, w_t is the wage, $T_j \left(a_i h_{i,t} w_t \right)$ is a function of taxes and benefits, B_t is a lump-sum transfer from the government, and r_t is the interest rate. HTM households face the same problem as the Ricardian consumers, except with no terms related to bond holdings in their budget constraint.

The left side of the budget constraint reflects household expenditures on consumption and bond holdings, whereas the right reflexts household net income. As part of household expenditures, there is a convex cost $\frac{\mu}{2}\left(b_{i,t+1}-\bar{b}\right)^2$ to holding debt at a level that deviates from $\bar{b}$.[19] However, these costs may be turned off, which I do in the sensitivity analysis of Appendix A.3.2. As part of their income, households earn $a_i h_i w_t$ from supplying labor. Thus, a distribution of labor income arises from the single market-clearing wage. Households also receive their principal and interest from current bond holdings $(1+r_t)\, b_{i,t}$ (unless they are HTM) and a lump-sum transfer from the government B_t. Households pay taxes and receive benefits $T_j\left(a_i h_{i,t} w_t\right)$, which is negative if benefits received exceed tax liability. Note that this tax function depends on the household type j; taxes and benefits not only depend on income but differ based on marital status and number of children.

Households have differing labor supply conditions depending on if they at or away from benefits cliffs. Solving the household problem of utility maximization 1.12 subject to the budget constraint 1.13 yields the consumption-Euler equation and labor supply conditions:

$$c_{i,t}^{-\frac{1}{\sigma}}\left(1+\mu\left(b_{i,t+1}-\bar{b}\right)\right)=\beta E_t\left(c_{i,t+1}^{-\frac{1}{\sigma}}\left(1+r_{t+1}\right)\right) \tag{1.14}$$

$$h_{i,t}=\left(\frac{c_{i,t}^{-\frac{1}{\sigma}}\left(a_i w_t-T_j^{'}\left(a_i h_{i,t} w_t\right)\right)}{\psi}\right)^{\epsilon} \tag{1.15}$$

where $T_j^{'}$ is the effective marginal tax rate, or derivative of the tax-and-transfer schedule with respect to labor. However, there are some households $i\in\{1,...,I\}$ that will have infra-marginal labor supply conditions. These arise from benefits cliffs inducing non-differentiable points in the tax-and-benefits schedule. Households located at these points will instead target their working hours to locate right on the cliff:

$$h_{i,t}=\frac{z_i}{a_i w_t} \tag{1.16}$$

where z_i is the pre-tax/transfer earnings location of the notch. This result follows from the simple theoretical model outlined in Section 1.2.1.

[19]I use bond-holding and debt costs to generate desirable dynamics. As the model abstracts from capital, bonds are a necessary savings instrument for households to be forward-looking. Without debt costs, Ricardian households permanently reallocate their portfolios in response to temporary shocks, behaving according to the permanent income hypothesis and generating unit roots in impulse responses.

1.3.2 Firms and Government

The supply side of the economy is populated by a perfectly-competitive representative firm that takes only labor as an input. In pursuit of an upper-bound effect of the effects of benefits cliffs, I abstract away from capital so output is entirely dependent on hours and total-factor productivity (TFP). Starting at date $t = 0$, the representative firm produces a consumption good and pay wages in order to maximize expected lifetime profits:

$$\max_{\{L_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \left(A_t L_t^{1-\alpha} - w_t L_t \right) \tag{1.17}$$

where α governs returns to scale. A_t is TFP and the source of aggregate shocks, which follow an AR(1) process in the log:

$$\ln(A_t) = \rho \ln(A_{t-1}) + \zeta_t \tag{1.18}$$

where ρ governs the persistence and ζ_t governs the size of the shock. Solving the firm's first-order condition yields the wage:

$$w_t = (1-\alpha) A_t L_t^{-\alpha} \tag{1.19}$$

The government collects taxes and distributes transfers, adjusting the lump-sum payment to households in order to balance the budget in every period:

$$\sum_{i=1}^{I} \omega_i T_j \left(a_i h_{i,t} w_t \right) = \sum_{i=1}^{I} \omega_i B_t \ \forall \ t \tag{1.20}$$

where ω_i are household weights. The baseline model abstracts away from government debt, as most households in the range of incomes sampled do not hold significant portfolios of treasuries. This assumption is not innocuous, though it may be fair under certain assumptions. In general, in models where Ricardian Equivalence does not hold, alternative fiscal closing assumptions (i.e. closing with lump-sum transfers vs changes in government purchases) generate differing quantitative results. Moore and Pecoraro (2020a) find that the qualitative responses of aggregates are similar, but quantitative differences may be large - often greater than 0.1 percentage points if fiscal closure is imposed less than 20 years after a shock, as opposed to 30 or more. Contemporaneous fiscal closure, as outlined here, tends to overstate the impact on output relative to closure in later years. This plays to my interest in deriving an upper-bound for aggregate effects, but I allow for a stock of government debt and alternative fiscal closure rules in a model extension in Appendix A.3.2.

1.3.3 Market Clearing and Equilibrium

A competitive equilibrium is a set of prices (r_t, w_t) and allocations $(c_{i,t}, h_{i,t}, b_{i,t+1}, B_t)$ taking $b_{i,t}$ and A_t as given, so that household optimally conditions for consumption and labor supply - 1.14, 1.15, and 1.16 - hold subject to their budget constraint 1.13. Moreover, firms maximize profits so that their optimality condition 1.19 holds, and the government lump-sum transfer B_t adjusts to maintain budget balance 1.20. Finally, the labor and goods markets clear, and bonds are in zero net supply:

$$L_t = \sum_{i=1}^{I} \omega_i h_{i,t} \tag{1.21}$$

$$Y_t = \sum_{i=1}^{I} \omega_i c_{i,t} \tag{1.22}$$

$$0 = \sum_{i=1}^{I} \omega_i b_{i,t+1} \tag{1.23}$$

Overall, I have constructed a dynamic stochastic general equilibrium model with discrete, heterogenous households, some of which face inframarginal decisions. Precise calibration of the household weights and tax/transfer systems faced by households should deliver an upper bound of the effects of benefits cliffs on the macro economy, while mirroring the key empirical finding in Section 1.2 of hours reductions induced by benefits cliffs. I turn to the calibration now.

1.3.4 Calibration of Household Placement and Tax-and-Benefit Schedules

To keep the model tractable, I simulate a finite number of heterogeneous households. Some of the implications of a continuum of heterogeneous households are explored in Appendix A.2.3. I estimate values of idiosyncratic productivities in the steady state to place households along the tax-benefit schedule. This is best illustrated with an example:

Depending on their placement along the tax-and-benefit schedule, households will have differing labor supply conditions. Consider a pre- and post-tax/transfer diagram as in Figure 1.4. The x-axis is pre-tax/transfer earnings and the y-axis is post-tax/transfer earnings.

Thus, the dashed line represents outcomes with no taxes or benefits. Introducing a tax rate, τ, lowers after-tax income. I then introduce a benefit of amount d that arbitrarily cuts off at some level of pre-tax/transfer income, generating a cliff. I set idiosyncratic productivities so that in the steady state, a household exists on either side of the cliff (represented by points A_1 and A_3) and a household exists right on the cliff (represented by point A_2). Households on either side of the cliff have normal first-order conditions for labor supply, as in equation 1.15. The household right on the cliff is inframarginal and targets their hours according to equation 1.16.

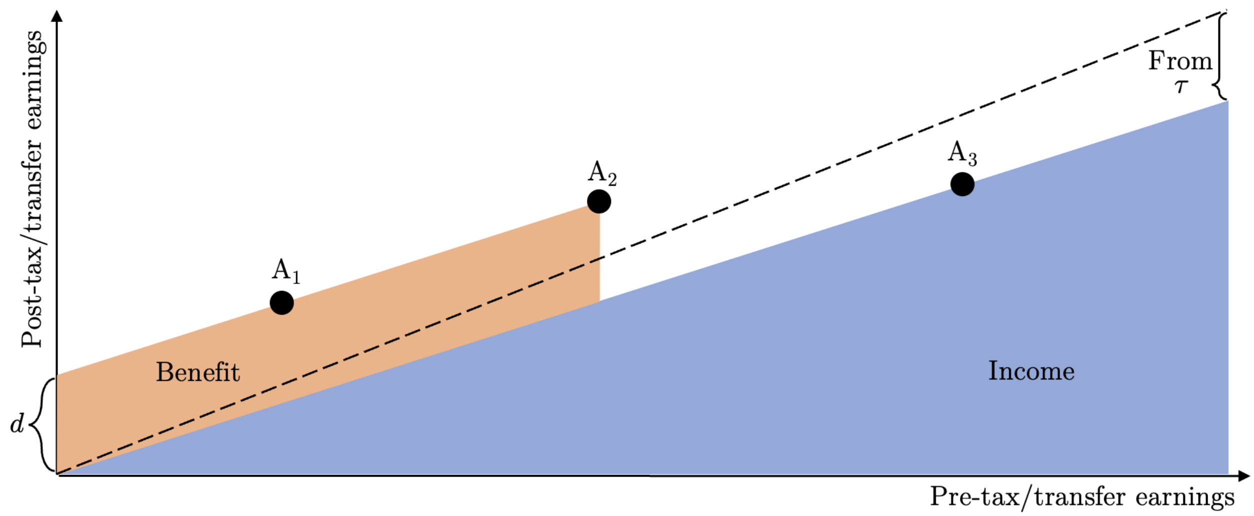
I place households according to this logic for every cliff in their tax-and-transfer schedule, which allows me to nonparametrically set the schedule with a few caveats. Using data from the BCW, I approximate the tax-and-transfer schedule for six types of households: those who are married or not with 0, 1, or 2 kids.[20] This implicitly assumes full fungibility of benefits and does not take into account non-pecuniary benefits or costs of these programs, treating the dollar amount of all benefits the same. For each type of household, I calculate the population-weighted average benefits schedule faced by each type of household across all counties. I then calibrate tax rates and benefits in the model to match these schedules.[21]

A key feature of the model is that any tax-and-transfer schedule, regardless of discontinuities, may be implemented, though there are some trade-offs. In the model's current form, households away from cliffs will *always* be away from cliffs (and households at cliffs will always be at cliffs), so that this model is not suitable for large aggregate shocks. By extension, the model does not explicitly build in the same hedging behavior outlined in and observed in Section 1.2. This matters most for comparisons across steady-states, but less so for the focus of this paper, aggregate fluctuations, and in particular, an upper-bound on the detrimental effects of benefits cliffs to aggregate productivity shocks. Because the steady-state does not reflect hedging, it will understate the potential costs of benefits cliffs. However, dynamically, agents who hedge act identically to those who are right on top of their cliff: an increase in income would reduce working hours. Thus, the effects of households near cliffs in the data will be entirely loaded on households at cliffs in the model.

[20] This covers just over 91% of the sample population.

[21] One may argue that the tax-and-benefit schedules of the sample states are not representative of the United States. However, with the given goal of delivering an upper bound for the impact of benefits cliffs on aggregate fluctuations, the fact that these states feature some of the largest cliffs in the US - due to not expanding their Medicaid programs under the Affordable Care Act - plays in my favor. In future work, I hope to expand the sample to the entire United States with supplemental data.

Figure 1.4: Household Placement Illustration



Note: An example of how households are placed around a single benefits cliff in the tax-and-transfer schedule in the model, with a benefit of size d and a tax rate τ. Point A_1 is placed behind the cliff, point A_2 is placed at the cliff, and point A_3 is placed beyond the cliff.

1.3.5 Calibration of Weights and Other Parameters

Weights: The calibration of household weights ω_i determines the prevalence of each type of household, and is thus done according to the data. Household weights are set according to earnings density data from the ACS. For households in between each cliff, the weights come directly from the earnings density for each household type. For households right on cliffs, weights cannot come directly from the earnings density. This is because the take-up of benefits is often incomplete, thus invalidating the assumption that all households at a statutory benefits cliff behave inframarginally - if a household did not take up a benefits program with a cliff, they would face no cliffs from this program by default. As a baseline, I assume that the take-up rate is 60%. Overall, this leads to household weights where about 5% of households behave inframarginally.

HTM Households: The share of HTM consumers κ_{ij} is calibrated based on data relating the marginal propensity to consume (MPC) out of unexpected transitory income changes across the distribution of cash-on-hand. Using data from the 2010 Italian Survey of Household Income and Wealth, Jappelli and Pistaferri (2014) find that the share of MPCs that equal 1 is nearly 50% in the lowest percentile of cash-on-hand, falling to around 10% for the highest percentile.[22] I set the share of hand-to-mouth households accordingly. In general,

[22] It should be noted that this result is from survey data for Italian consumers, which may not be externally valid for the United States. However, their calculated share of hand-to-mouth consumers is similar to that of the less granular studies in the United States, including Campbell and Mankiw (1989), Aguiar et al. (2020),

Table 1.1: Calibration

Description	Parameter	Value	Source/Target
Preferences			
Discount factor (annual)	β	0.98	Standard value
Intertemporal EIS	σ	1.064	Chetty (2006)
Frisch elasticity of labor supply	ϵ	0.38	Heathcote et al. (2008)
Work distaste	ψ	0.05	Average working hours in US
Target household debt	$\bar{b}$	0	Standard value
Debt-holding costs	μ	0.0002	Unwinding debt over 60 years
Technology			
Returns to scale	α	0	Constant returns
TFP shock size (annual)	ζ	0.03	Pancrazi and Vukotic (2011)
TFP persistence (annual)	ρ	0.552	Pancrazi and Vukotic (2011)
Distribution			
Idiosyncratic productivities	a_i	x	Household incomes (ACS)
Household weights	ω_i	x	Income densities (ACS), take-up rates
Tax rates	$\tau_{i,j}$	x	Tax-and-transfer schedules (BCW)
Benefits	$d_{i,j}$	x	Tax-and-transfer schedules (BCW)
Share of HTM households	$\kappa_{i,j}$	x	Jappelli and Pistaferri (2014)

Note: Values marked with x are household-specific.

households with below-median income have MPCs ranging from 25% to 50%, and households with above-median income have MPCs ranging from 10% to 25%.

Preferences: The discount factor β is set to imply a long-run real annual interest rate of 4% annually, and I set the intertemporal elasticity of substitution σ to 1.064 as a baseline. Chetty (2006) finds that evidence of the effects of wage changes on labor supply imposes an upper bound on the coefficient of relative risk aversion - equivalent to $1/\sigma$ - of about 2, with a central estimate of 1 (log utility). In all but a few cases, this upper bound is 1.25, implying a minimum value of 0.8 for σ. I err on the side of risk aversion to better match the relationship between wages and hours in the data. I set the Frisch elasticity of labor supply ϵ to 0.38, closer to the empirical estimates of the Frisch elasticity from the applied microeconomic literature. This matches the estimate for the elasticity of Frisch labor

and Carroll et al. (2017).

supply with household heterogeneity and measurement error by Heathcote et al. (2008). My baseline calibration gives all households the same value, though I later relax this assumption and perform sensitivity analyses to both σ and ϵ in Appendix A.3.2.

Other parameters: Debt adjustment costs μ and targeted debt-level $\bar{b}$ are set so households try to eliminate their debt around 60 years after the baseline shock. Work distaste ψ is set to match the average number of working hours in the United States. Estimates of TFP shock persistence ρ have trended up and its standard deviation ζ has trended down since 1985, but I use the latest estimates from Pancrazi and Vukotic (2011). Converting these to an annual frequency, I set ρ to 0.522 and ζ to 0.03. Household idiosyncratic productivities a_i are estimated in the steady-state using generalized method of moments to match agents to their incomes along tax-and-benefit schedules as described above. I set $\alpha = 0$ so that the representative firm has constant returns to scale in its factors of production. Table 1.1 presents the parameter values for the baseline model.

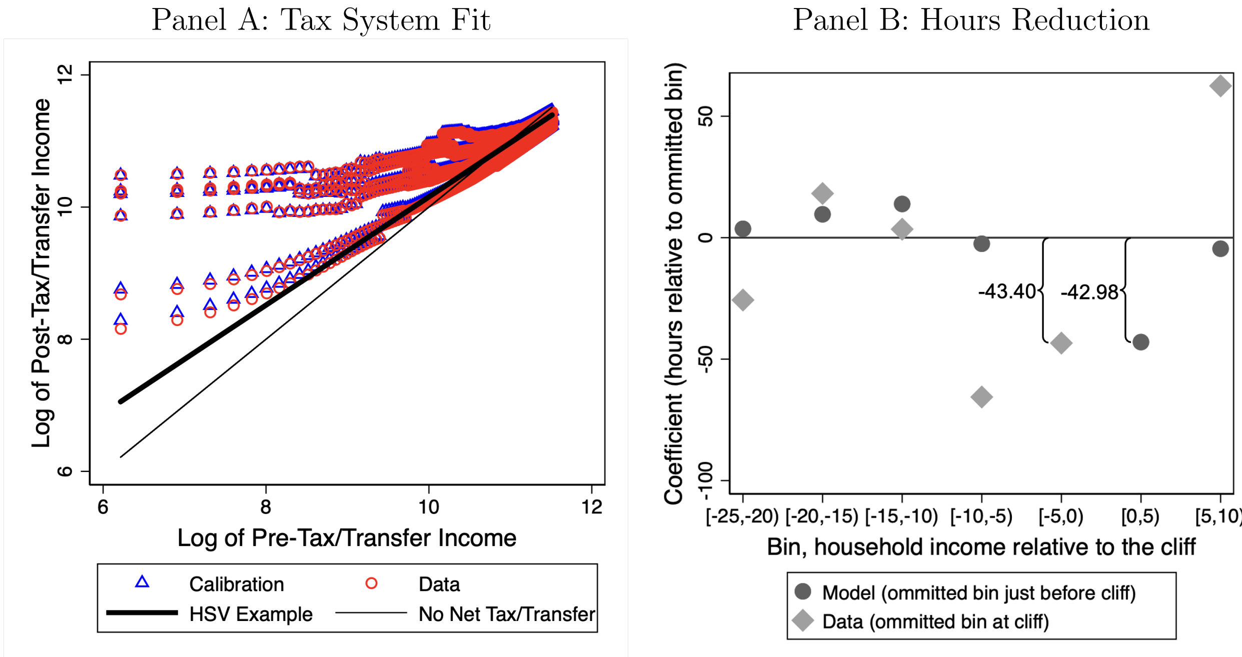
1.3.6 Relationship to the Data

Using inframarginal agents allows for sudden declines in benefit amounts so that the calibrated tax-and-transfer schedule fits the data better than common parametric functions. Figure 1.5, Panel A, shows the fit of the calibrated tax-and-transfer schedule in blue compared to the data in red, with the log of post-tax/transfer amounts from the data along the x-axis and the log of post-tax/transfer amounts from the model along the y-axis. The fit is best on pre-tax/transfer incomes below \$50,000 - at worst, it understates the tax liability of families earning \$100,000 by an average of about \$1,500. For contrast, I also plot outcomes for an example HSV tax function. This tax function has been common in the heterogeneous agent macroeconomic literature since introduced by Heathcote et al. (2017), and takes the form of taxes as a function of pre-tax/transfer income $T(y)$:

$$T(y) = y - \lambda y^{1-\tau} \tag{1.24}$$

where λ shifts the tax function to determine the average level of taxation and τ determines the curvature, or progressivity, of the tax system. In contrast to the assumption of linear income taxes, the HSV allows for progressive rate structures, including negative average tax rates at the lower end of the income distribution. However, taxation under this function is smooth, monotonic, and does not allow for cliffs. I plot the HSV tax function as calibrated in Heathcote et al. (2017) as the thick black line in Figure 1.5, Panel A. The tax function effectively captures most of the US income tax system, particularly for higher incomes, but underestimates benefits for the lowest incomes - a result already highlighted by Heathcote

Figure 1.5: Model Performance



Note: Panel A plots the log of pre-tax/transfer income vs. the log of post-tax/transfer income, as in Figure 1A of Heathcote et al. (2017), for the six different tax schedules. The data from BCW is in red circles, the model calibration is in blue triangles, and the HSV calibration in thin black lines. The thick black line indicates no net taxes/transfers. Panel B plots the point estimates of the of total annual hours relative to the cliff (from the data) or relative to the bin just behind the cliff (from the model).

et al. (2017). Moreover, the HSV particularly fails to match the data in regions with discontinuities, instead smoothly increasing monotonically throughout. My calibration improves upon the HSV function in both these regards.

In contrast to the empirical setting and for computational tractability, the model features individuals between and at cliffs, and not near them, though it captures the same implications. As mentioned above in Section 1.3.4, the lack of agents just before cliffs matter most for steady-state comparisons, but not for dynamics as agents near cliffs act identically to agents at cliffs, reducing their hours in response to a wage increase. Thus, the negative effects of benefits cliffs will be loaded entirely on the inframarginal agents. The empirical results suggest a decline of about 40 hours annually due to benefits cliffs. Given a reasonable calibration, the model should match this outcome for those at cliffs. I simulate the model for 1,000 periods, retrieving incomes and hours from the simulation and treating the data as a cross-section as in the empirics. The average difference between hours of households at the notch and away from the notch is 42.98, conforming to the empirical results and illustrated by Figure 1.5, Panel B. By construction of estimating abilities in the steady-state, the distribution of incomes and working hours also match the data from the ACS sample.

1.4 Counterfactual Policy Simulation

In this section, I compare responses to aggregate productivity shocks across the baseline model and a counterfactual that smooths over benefits cliffs. Additional counterfactuals and their results can be found in Appendix A.3.1. I start by describing the counterfactual simulation, and highlight the results later in this section.

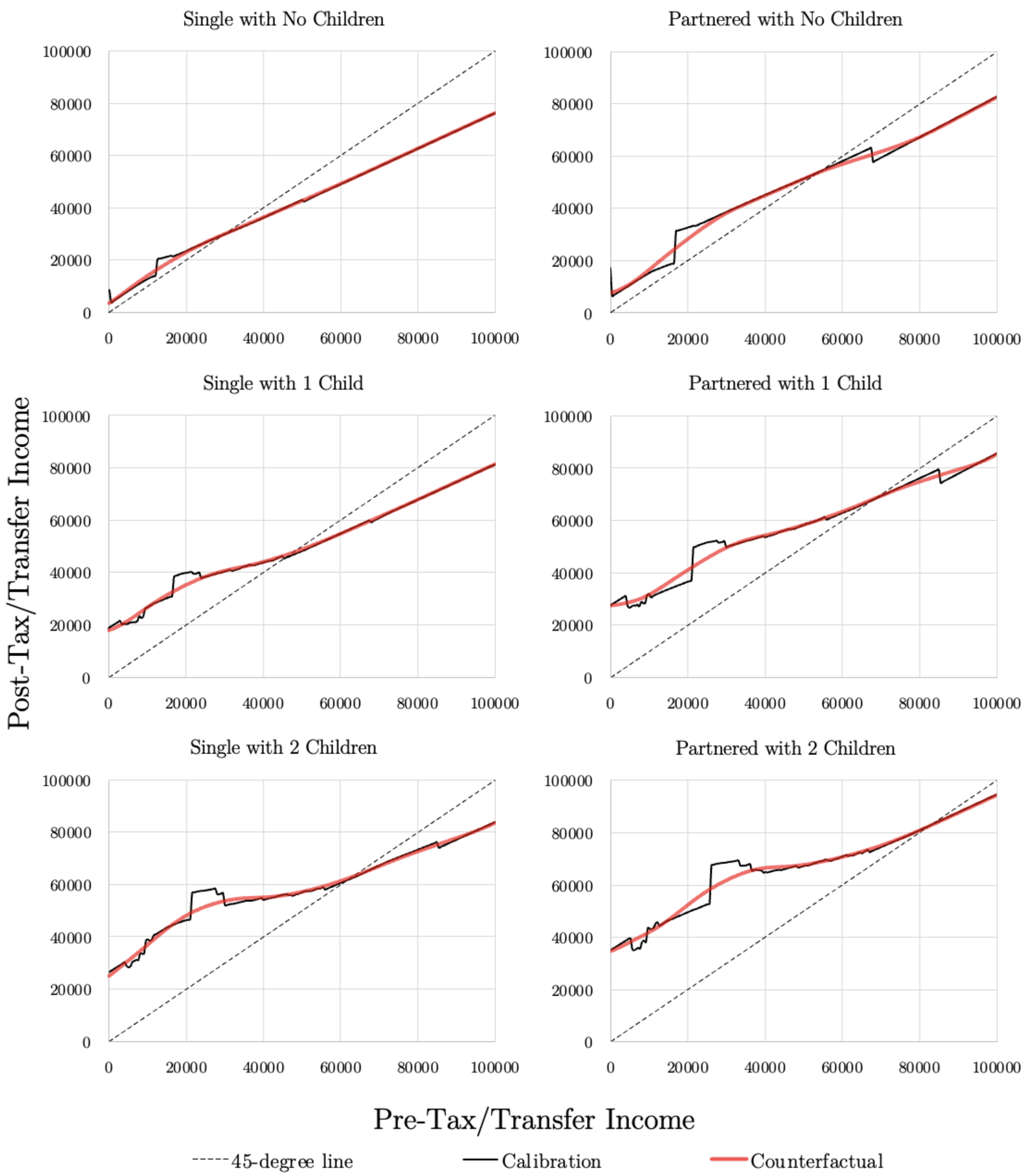
The main counterfactual I consider replaces the tax-and-benefits schedule for each household type with a smoothed version. This allows the tax-and-transfer system to remain as close to the baseline as possible, but without cliffs. Specifically, it keeps all tax rates the same but adjusts benefits using a locally-weighted regression to "round off" cliffs. Benefits are set so that for pre-tax/transfer income z_{ij}, a smoothed post-tax/transfer income y^s_{ij} is calculated using a subset of post-tax/transfer earnings y_{ij}, with indices $i^-_{ij} = \max(1, ij - k_{ij})$ and $i^+_{ij} = \min(ij + k_ij, N)$. Here, N is the total number of observations - I have 200 for each tax/transfer schedule, in \$500 increments of pre-tax/transfer income from \$0 to \$100,000. k_{ij} is a range of income excluded from calculating the smoothed post-tax/transfer income y^s_{ij}, where $k_{ij} = (N \times b - 0.5)/2$. b is a chosen parameter related to the bandwidth. Smoothed values y^s_{ij} are then a weighted regression prediction for each z_{ij}, where the weights are given by the tricube weighting function:

$$w_{ij} = \left(1 - \left(\frac{|z_{ij} - x_{lj}|}{\Delta}\right)^3\right)^3 \tag{1.25}$$

where $\Delta = 1.0001 \max(z^+_{ij} - z_{ij}, z_{ij} - z^-_{ij})$. I set b to 0.41, which is the lowest value that allows for a smoothed schedule and ensures that effective marginal tax rates (EMTRs) are below 95%. Values lower than this overfit the baseline schedule, generating EMTRs over 95%, effectively not ridding the tax/benefit schedules of cliffs. I plot the calibrated and counterfactual tax-and-transfer schedules in Figure 1.6. The appendix considers counterfactuals of replacing the entire tax-and-transfer schedule with a linear income tax and demogrant, as well as replacing the tax-and-transfer schedule with a version of the upper envelope of the schedules that do not feature marginal tax rates exceeding 95%.

While I show one potential counterfactual here, note that any counterfactual that eliminates benefits cliffs will involve trade-offs between benefits and marginal tax rates, which will in turn affect welfare. There is no way to eliminate benefits cliffs entirely without some individuals having their benefits decreased and/or without excessively high marginal tax

Figure 1.6: Counterfactual Tax-and-Transfer System Compared to Baseline



Note: The dashed lines (labeled for 45 degrees) is earned income with no taxes and benefits. The black lines are for the calibrated baseline tax-and-transfer system, which features benefits cliffs. The smoothed counterfactual tax system is shown in red, which is calibrated from the baseline to smooth over cliffs.

rates, unless there are truly heroic assumptions made about government debt. Further still, simply eliminating benefits cliffs may not be a desirable goal. For example, I could eliminate all benefits cliffs above \$35,000 for a family of 4 by replacing the schedule with a horizontal line at that point, but this would yield EMTRs of 100% and disincentive many from earning higher incomes. Slemrod (2013) emphasizes that no general statements about the welfare changes from eliminating tax-and-transfer notches can be made without considering alternatives, such as the nature of the optimal income tax and the tax instruments available to governments. Here, I compare changes in welfare from having cliffs to the smooth alternative described above, which assumes that the government can have a completely flexible income tax system as in Mirrlees (1971).

1.4.1 Results

Aside from comparing the effects on output, I am also interested in the effects on welfare. Here, I chose a consumption-equivalent definition of welfare. Specifically, for each household, I find η for

$$E[U(c_{counter}(1+\eta), h)] \equiv E[U(c_{baseline}, h)] \tag{1.26}$$

where $U(.)$ is utility over consumption and hours. Here, η gives the percent of consumption that an agent would be willing to forgo to return to the baseline model from their counterfactual. Aggregate welfare changes are implicitly utilitarian, as if giving every agent the same weight in a social welfare function. Future work will explore alternative social welfare functions, such as Rawlsian.

Table 1.2 summarizes my results, to be discussed below. Panel A presents the percent change in steady-state output for the smooth counterfactual relative to the baseline model. However, these headline numbers obscure heterogeneous changes in output across the income distribution, which I highlight in the same table. In Panel B, I show the percent-changes in welfare in the new steady-state, disaggregating between households formerly subjugated to cliffs and otherwise for each of the counterfactual models. Finally, in Panel C of Table 1.2, I show differences between the baseline and smooth counterfactual models when subject to a 1% TFP shock. The values here are the percent difference between baseline and counterfactual responses. For example, output in the baseline model increases on impact in response to a positive TFP shock but increases 1.6% *more* in the counterfactual model with smoothed benefits - in other words, a 3% output gain on impact from a productivity shock in the baseline model grows to 3.12% in the counterfactual).

Steady-state: Although not the focus of this project, changes in steady-state variables may

Table 1.2: Smooth Counterfactual Results Relative to the Baseline with Cliffs

Panel A: Steady-State Output

Aggregate	By income, thousands of USD				
	\$0-\$20	\$20-\$40	\$40-\$60	\$60-\$80	\$80-\$100
1.448%	3.787%	15.368%	-4.002%	-1.312%	-2.832%

Panel B: Steady-State Welfare

Aggregate	Formerly on cliffs	Never on cliffs	*Addendum:* %Δ in Lump-Sum Payment
-4.582%	-2.179%	-4.628%	-8.8%

Panel C: Response to TFP shock

Improvement in Output	Improvement in Welfare		
	Aggregate	Formerly on cliffs	Never on cliffs
1.582%	9.352%	181.809%	6.057%

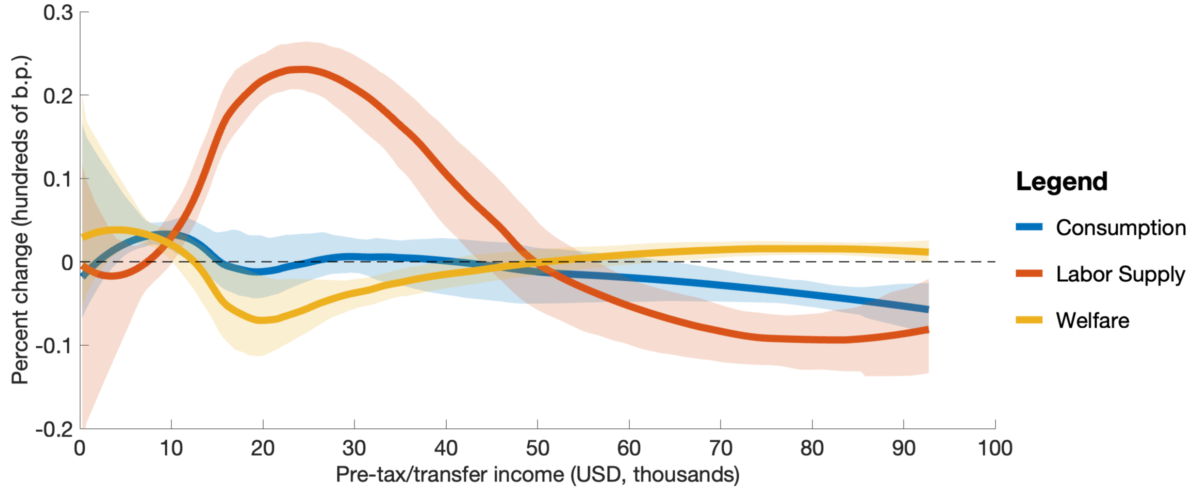
Note: Values are in percent changes from the baseline model with benefits cliffs to the smooth counterfactual without benefits cliffs. This includes Panel C, which is the percent improvement of the response of output under the counterfactual model compared to the same response in the baseline. Panel A shows changes in output in aggregate and across incomes, and panel B shows changes in consumption-equivalent welfare.

still be of interest and inform policy implications. Under the smooth counterfactual, output increases about 1.5% in the steady-state, with increases mostly stemming from increased production from those with incomes below \$40,000. This increase in output is not costless, as increased labor supply causes average welfare to fall by about 4.6% in this counterfactual steady-state - though over two points less for those formerly at cliffs compared to their always-unconstrained counterparts. Given that the model does not explicitly feature hedging, these steady-state comparisons are lower bounds - the output gain would be larger and the welfare loss would be smaller if the baseline steady-state featured hedging before cliffs. Ultimately, compared to this specific alternative benefits system, benefits cliffs reduce steady-state output by at least 1.448%, though welfare losses in the steady-state cannot be ruled out.

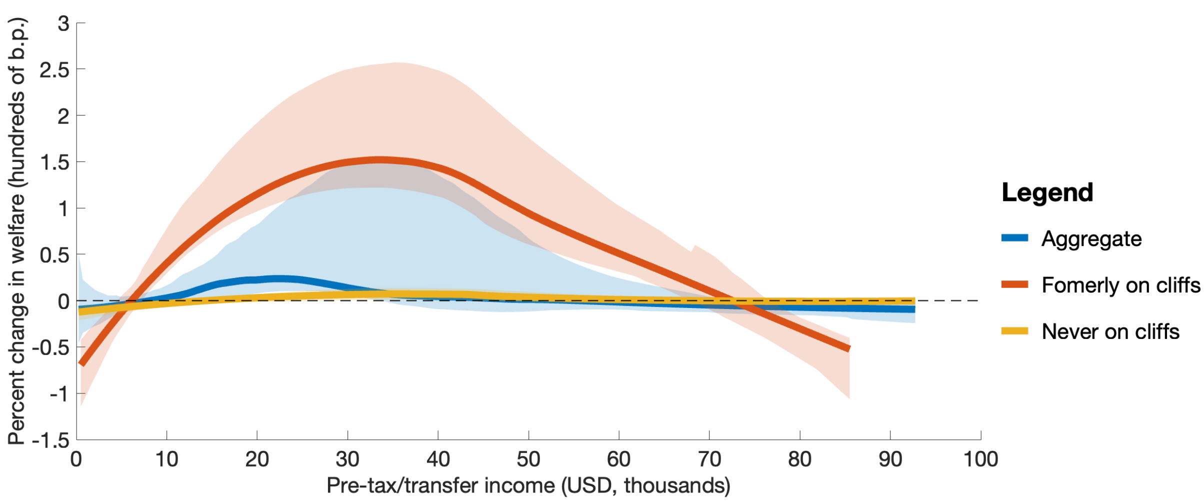
Changes in steady-state welfare are also highly heterogeneous, and are largely a consequence of the chosen counterfactual. Figure 1.7, Panel A plots percent changes in steady-state welfare as measured by consumption equivalence across the income distribution in yellow. The x-axis is an agent's income in USD, while the y-axis is percent change from

Figure 1.7: Smooth counterfactual results relative to the baseline with cliffs, across incomes

Panel A: Steady-State Outcomes



Panel B: Response to TFP Shock



Note: Lines are the result of fitting a LOWESS function to the outcomes for individual households. Panel A shows the percentage change in steady-state outcomes compared to the baseline model with benefits cliffs to the smooth counterfactual model without benefits cliffs. Outcomes are for consumption-equivalent welfare (yellow), consumption (blue), and labor supply (red) across the income distribution. Panel B shows the percentage change in the welfare responses to a total factor productivity shock from the baseline model with benefits cliffs to the smooth counterfactual without benefits cliffs across the income distribution. Outcomes are for everyone (in blue), those formerly on cliffs (in red), and those never on cliffs (in yellow).

its steady-state value. The values are unweighted, so that changes in outcomes for individual households may be more easily observed. For the smooth counterfactual, welfare in the steady state increases the most for those with incomes below \$12,000 - a peak of about 4.5%. Driving these effects are changes in steady-state consumption and labor supply. The blue and red lines of Figure 1.7, Panel A show the percentage change in these values, respectively, across the income distribution. Again, outcomes are unweighted to better observe the responses of individual households. Increases in welfare are highly, positively correlated with increases in consumption, and negatively associated with increases in labor supply. Given the reduction of benefits in the counterfactual - which is needed in over to smooth over resulting cliffs - primarily around the introduction of health insurance subsidies for families, income effects induce households in the roughly the $10,000-$50,000 range to increase their labor supply the most relative to other households - by up to over 20% - leading to the greatest drop in steady-state welfare for those who earn around $20,000.

Aggregate Productivity Shock: Compared to the baseline model with cliffs, the smooth counterfactual delivers large, concentrated welfare gains and smaller aggregate implications for output in response to productivity shocks. Output increases nearly 1.6% more on impact in this counterfactual model relative to the baseline. In other words, a productivity gain that would lead to a 3.00% output gain in the baseline model yields a 3.12% output gain in the smooth counterfactual - or about an additional $27 billion using Q3 2023 real GDP for the United States.[23] This occurs as agents formerly constrained to reduce their working hours in order to maintain their benefits are freely able to adjust their hours under the smooth counterfactual.

While the impact on output may be small relative to the size of the baseline economy, impacts on welfare are greater and highly concentrated, and also affect households away from cliffs. Aggregate welfare increases 9.35% more in response to a productivity shock, rising from a 1.07% gain in the baseline model to a 1.17% gain in the smooth counterfactual. For many formerly-constrained households, the welfare response to a shock more than doubles, averaging a 181% gain for formerly-constrained agents - from a 0.75% gain in the baseline model to a 2.11% gain in the smooth counterfactual. In other words, for formerly-constrained agents, for every $1 of consumption-equivalent welfare gain to a productivity shock in the baseline model with cliffs, there is a $2.81 gain in the smooth counterfactual.[24] These additional gains from a productivity shock are not confined to formerly-constrained households: always-unconstrained households' welfare increases by about 6%, from 1.06% in the baseline to 1.12% in the smooth counterfactual. Figure 1.7, Panel B plots these relative changes in welfare across the income distribution on average in blue, and disaggregating between households formerly subjugated to cliffs (in red) and not (in yellow). Like the results for

[23]For comparison, this is less than half of the spending on the Earned Income Tax Credit or roughly a fifth of the expenditure on SNAP (otherwise known as food stamps).

[24]With the assumption of perfect fungibility between consumption goods and benefits, I can translate the gains into dollar amounts since welfare is given in consumption-equivalent terms. For the average formerly-constrained household, welfare gains to an aggregate productivity shock improve from $263.52 to $741.37, an increase of nearly $500.

steady-state welfare, improvements to welfare in response to a shock are heterogeneous, and concentrated on those formerly on cliffs in the \$20,000 - \$50,000 range.

1.4.2 Discussion

Almost all of the above results are stable for alternative calibrations of parameters. Appendix A.3.2 considers alternative assumptions on parameters, including allowing for a stock of government debt, abstracting from bond-holding costs, a more flexible calibration of Frisch elasticities, and varying parameters by $\pm 25\%$ of their baseline values. Across all specifications, the impact on the response of output to a TFP shock is relatively steady, and never exceeds a 4% gain relative to the baseline. The most sensitive outcome is the improvement in welfare for those formerly on cliffs, which ranges from 97% to 361.76% depending on parameter assumptions. Hence while aggregates are less sensitive to parameter assumptions, welfare is not, although it at least nearly always doubles for those formerly on cliffs.

What do we learn from this beyond what is already gleamed from the partial equilibrium exercises of Section 1.2, or similar studies? For one, we can see that the impact of the elimination of cliffs goes well beyond those at the cliffs themselves. Consider the smooth counterfactual: marginal tax rates and benefits barely budge compared to the baseline model, particularly for incomes above \$60,000. Still, we see declines in labor supply approaching -10% for some agents above this income compared to the model with cliffs, as the increased labor supply of lower-income agents supplants them and allows reductions in the lump-sum payment necessary to balance the budget. Two, we can examine the implications of benefits cliffs for aggregate shocks on aggregate variables. Here, we find that they are small: the response of output improves only 1.58% in the smooth counterfactual on impact. This may arguably have already been surmised if one properly reckoned that a small percentage of the population is subject to benefits cliffs, but impacts on welfare would still be obscure, given that removing cliffs may not only increase consumption, but increase working hours. In the smooth counterfactual, I find that welfare not only improves considerably in response to a positive shock for those on cliffs, but it increases over 6% for those never at cliffs to begin with.

Furthermore, this general equilibrium model implies much larger swings in outcomes from eliminating cliffs than implied in the partial equilibrium model, at least partially driven by the changes in schedules necessary to eliminate cliffs. Assume that every single person in the affected bins reduces their working hours by 79.75 hours annually - the largest empirical point estimate - solely because of benefits cliffs. Adding back these lost hours increases labor supply in the data sample by 0.961% overall. In contrast, the increase in output (and therefore labor supply in this model) in the smooth counterfactual is nearly 65% larger. One may argue that this may not be a fair comparison, as benefits and effective marginal tax rates have changed in this counterfactual, and thus labor supply is being affected by factors other than benefits cliffs. However, these reflect the real trade-offs that policymakers must face: the elimination of benefits cliffs necessitates changes in transfers and effective marginal tax rates, particularly if such reforms are to be revenue-neutral.

Although a business-cycle model, the results here carry longer-term implications. If total factor productivity is more likely to rise than otherwise, the averse effects of cliffs on output and welfare are likely to accumulate over time. A back-of-the-envelope exercise yields once again that the aggregate implications are small relative to large potential welfare gains for those constrained by cliffs. I feed the estimated series of changes in utilization-adjusted total factor productivity since 1947 from Fernald (2014) into the baseline and counterfactual models and compare outcomes.[25] This exercise implies today's economy would be up to 1.4% larger were it not for benefits cliffs. It also implies a cumulative welfare loss of missed gains for households constrained due to benefits cliffs of 143% of their current average welfare. Even if the steady-state welfare falls in this counterfactual with smoothed benefits cliffs, welfare gains over time may potentially more than offset this loss.

1.5 Policy Implications and Conclusion

Although the aggregate effects are small, the fact that such large welfare gains in response to improvements to productivity are theoretically possible - particularly for lower-income households who could use upward mobility the most and may have fewer opportunities to advocate for themselves - merits consideration by policymakers. In this section, I briefly describe implications for policy and conclude.

1.5.1 Policy Implications

Lawmakers should avoid creating benefits cliffs in the first place - once a cliff exists, removing it will necessitate a trade-off between benefits, effective marginal tax rates, and revenues. benefits cliffs are not phenomenon from policies long since passed, and are still being implemented by policymakers. For example, the Inflation Reduction Act (Pub. L. 117–169), while eliminating some benefits cliffs, also generated new ones: tax credits for electric vehicles with values up to $7,500 are cut off at incomes above $150,000 ($300,000 if

[25] This approximation is admittedly crude - many of these benefit programs did not exist in 1947.

filing jointly).

Fortunately, given that the existence of benefits cliffs is well-documented, as are potential solutions. First and foremost, benefits should be gradually phased out or capped with income, rather than be subject to sudden losses. The Inflation Reduction Act does this for health insurance subsidies: rather than be eliminated above 400% of the poverty line, the target premium rises to and is capped at 8.5% of income. Again, trade-offs will be present: phasing out a program without lowering benefits for some necessitates increases in spending, with the cost decreasing with the speed of the phase-out. However, phasing out a program too quickly risks creating high effective marginal tax rates that discourage labor supply. Moreover, policymakers should be cognizant of other benefits that may phase out over the same region, as multiple programs phasing out simultaneously can generate high marginal effective tax rates that discourage labor supply.

If phasing out with income is not possible nor desirable, policymakers may consider phase-outs along other lines, such as time. For example, under current policy in the state of Florida, parents lose 100 percent of childcare subsidies when their earnings reach 85 percent of state median income. The proposed Families' Ascent to Economic Security (FATES) program in Florida would enable parents who would otherwise lose the childcare subsidy to instead pay a steadily increasing share over three years. Other proposals include providing additional monetary incentives for continued employment - essentially programs to help paper over the dominated regions of earnings induced by benefits cliffs.

Finally, longer-term proposals include expanding access to educational funding as well as employer-funded programs that invest in the skills development of entry-level workers. With programs of this nature, the hope is that workers can enter the labor force with incomes that place them beyond the most egregious benefits cliffs (National Conference of State Legislatures, 2019).

1.5.2 Conclusion

Benefits cliffs theoretically create disincentives for improving labor supply, which inhibit the response of output and welfare to productivity shocks. I uncover evidence that these disincentives are indeed present: in a sample covering the universe of benefit programs in nine southern US states, persons in households just before benefits cliffs reduce their hours worked by about 40 hours annually. However, I find that the impact of benefits cliffs on aggregate fluctuations is small. At most, GDP increases 4% more on impact in response to a positive TFP shock in counterfactuals that remove cliffs, and 1.58% in the main counterfactual considered in this paper, which smooths over cliffs. Nevertheless, the impacts on welfare may be quite large depending on the household. In all cases, the elimination of cliffs at least nearly doubles the welfare impact from a positive TFP shock for many formerly at cliffs. This doubling of the welfare response is robust to a range of critical parameter values. Moreover, it is possible for these gains to flow through to agents never at cliffs in the first place. To ensure that welfare gains from productivity improvements flow through to all households, policymakers should avoid creating benefits cliffs in their design of welfare systems.

CHAPTER 2

Place-Based Policy and Optimal Income Transfers in a Federalist Framework

2.1 Introduction

From the Beveridge Report of the United Kingdom in 1942 to the advent of the "Great Society" programs of the Johnson administration in the 1960s, governments have established income-support programs throughout the 20th century with the goal of promoting the economic and social well-being of their citizens. Starting in 1962, Milton Friedman popularized the idea of a negative income tax (NIT) - first proposed by Juliet Rhys-Williams in the Beveridge Report - leading to the Family Assistance Plan proposal by President Nixon in 1969, an effort to establish such a program in the United States. Concerns revolving around payments made to non-working households and incentivizing nonemployment eventually led to the enactment of the earned income tax credit (EITC) in 1975, which provides income subsidies but no guaranteed minimum income (GMI). Still, programs such as Temporary Assistance for Needy Families (TANF) and the Supplemental Nutrition Assistance Program (SNAP, otherwise known as food stamps), among others, provide some sort of GMI - although not universally - in the United States. Indeed, work subsidies have been a staple of the United States' social safety net, with over 23 million families receiving $53 billion from the EITC alone as of December 2023 (Internal Revenue Service, 2024). Explicit GMI programs were piloted in the United States during the 1960s and 1970s, generally documenting moderate reductions in work effort as reviewed in Robins (1985). However, with increasing attention towards income inequality combined with the rise of automation, advocates and economists alike have renewed interest in GMI programs. A GMI, such as a NIT or universal basic income (UBI), may address concerns of seemingly slow wage growth and potentially increased unemployment.[1]

[1]For an overview of these concerns about the effects of automation, as well as commentary on basic incomes and work subsidies as a potential treatment, see Manyika et al. (2017).

In this project, I study what an optimal income transfer system in a federalist society such as the United States would look like. Like over 40% of the world population and nearly 50% of the world's economy, the United States uses a mode of government that combines a general government with regional governments in a single political system known as federalism.[2] This allows not only the national (federal) government to set income transfer programs, but also regional (state) governments as well. For example, 29 states plus the District of Colombia have their own EITCs in addition to the federal credit. With this, subfederal governments may not only be concerned about balancing work incentive and equity concerns, but also have distributional and labor concerns unique to their own populations. They may also fear attracting low or no-income earners from other jurisdictions and losing high-income individuals and their tax dollars.

I address the question of what would an optimal income transfer system look like - balancing concerns for distribution and incentives to work subject to meeting revenue requirements - including for both federal and state-level programs, in three main parts. First, I construct an optimal tax model that takes into account three margins of employment: decisions to work or not with the extensive margin, decisions of whether to work more with the intensive margin, and finally a margin of mobility across states. I then calibrate the model for the United States as a whole, as well as each of the 50 US states plus the District of Columbia, and perform numerical simulations to yield the optimal income transfer systems for each of the jurisdictions. Finally, I conclude with a comparative exercise of the optimal income transfer programs of two disparate states that illustrates how key sub-national considerations - such as the pool of unemployed workers and state-specific labor elasticities - drives differences in the optimal tax-and-transfer schedule across states.

I begin by constructing a federalist model of income transfers and taxation that accounts for three distinct elasticities of labor supply. The framework is a discrete model of occupational choice in which the government can only tax and/or supplement labor income. Governments set taxes and transfers in order to maximize a weighted sum of individual utilities subject to raising a given amount of revenue. Individuals are able to adjust their labor supply among three margins: an extensive margin choice where individuals can choose to work or not, an intensive margin choice where individuals may switch to higher-paying occupations, and what I call a "mobility-margin" choice where individuals may choose the same occupation across alternative sub-national jurisdictions. The model yields an optimal tax-and-transfer formula in terms of elasticities for each of the aforementioned margins and

[2]Author's calculations. The list of countries organized as federations is from the SBS World Guide, and populations and PPP GDP figures are from the International Monetary Fund's World Economic Outlook Databases.

welfare weights. Relative to a model where interstate mobility is not possible, this generally leads states to offer less redistributive tax-and-transfer schedules.

I then calibrate and numerically simulate the model for both federal and state governments and find that states should substantially adjust taxes and transfers in response to concerns about mobility across states. The model assumes a range of extensive, intensive, and mobility elasticities across income, in which individuals' extensive labor elasticity falls with income, whereas their intensive and mobility elasticities rise. This allows for greater precision in numerical simulations than assuming uniform elasticities across (subsets of) earnings, in light of the literature that demonstrates heterogeneity in labor supply elasticities across incomes. Numerical simulations indicate that, on average, states find it optimal to tax away federal income transfers, particularly when facing potential inter-state migration, reflecting fiscal constraints and a fear of attracting no or low-income earners. However, all results indicate that all jurisdictions find it optimal to offer some amount of guaranteed minimum income combined with a more generous work subsidy.

I next perform a comparative exercise between two disparate states - Michigan and Massachusetts, which differ in the key dimensions of fiscal space, pools of non-employed, distribution of incomes, and extensive labor elasticities - to illustrate mechanisms in how states set their optimal transfer programs. I find that wealthier states, such as Massachusetts, should desire to supplement federal transfers due to increased fiscal space. States with larger pools of no-income earners, such as Michigan, should aim to increase the difference in consumption between those with no and earned income to encourage employment. I finally use the model to back out implied welfare weights for both federal and state US tax-and-transfer systems. That is, assuming governments are following this optimal tax-and-transfer model, what does that imply about their distributional preferences across income?

2.1.1 Contribution to Related Literature

This paper contributes to the literature on optimal income taxation by allowing for sub-national jurisdictions to set their own income transfers and taxes. In his seminal work, Mirrlees (1971) derives formulas for nonlinear income taxes focusing on the intensive margin of employment. Saez (2001) demonstrates that there is a simple link between optimal tax formulas and intensive elasticities of earnings familiar to empirical studies. Going on, Saez (2002) extends this framework to incorporate the extensive margin of employment. Gordon and Cullen (2012) use a similar framework to study income redistribution in a system of federal and state governments, again focusing on the intensive margin. This paper unites all three concepts: income redistribution in a system of federal and state governments, accounting for the extensive and intensive margins of employment, as well as potential migration across states. It is important to account for the extensive margin in this federalist system - and not just the intensive and migration responses - due to this paper's focus on taxes and transfers for low incomes. As indicated by McClelland and Mok (2012), the literature generally agrees that the extensive margin of labor supply plays a larger role for low incomes. It is thus crucial to incorporate the extensive margin to more accurately derive optimal taxes for these individuals.

This work also touches on studies of fiscal federalism and finds that states are particularly concerned about the extensive margin of employment and migration. Moreover, as expressed in Gordon (1983), a common finding is that a centralized federal authority is best suited to carry out income distribution, as externalities between states lead to under-provision. One such channel is interstate migration, which may have a significant influence on states in their setting of benefits. If it does, then states should be given less responsibility for income redistribution vis-à-vis the federal government (Gramlich, 1997). This paper confirms state states under-provide relative to the federal government on average, but also finds that states with more fiscal space may wish to increase transfers relative to a national program.

Next, the emergence of place-based policies as a tool to combat the geographically disparate outcomes of employment has given rise to the notion that central authorities should target wage subsidies to more disadvantaged states.[3] Given that some states may exhibit larger extensive-margin labor elasticities among their low-income earners as found in Benjamin and Summers (2018), this also touches on optimal fiscal federalism under heterogeneous preferences. Gordon (2023) creates an optimal federalist income tax model with migration and differing desires for insurance for income risk, in which individuals can mitigate risk by sorting into jurisdictions with differing income tax systems. Like that paper, I assume state governments set their tax-and-transfer systems best-suited for their residents at that time. In my comparative exercise, a key behavioral difference between states is extensive margin labor elasticities, rather than current income risk and risk aversion. If higher extensive margin labor elasticities translate to less risk aversion, this paper confirms that optimal taxes and transfers for these states are less progressive, as in Gordon (2023). I find that states with higher extensive elasticities indeed should wish to generate transfers that yield large differences between no and low income, but this often comes with a smaller guaranteed minimum income.

A final strand of the literature this work contributes to is underlying social welfare pref-

[3] Benjamin and Summers (2018) argue that "pro-employment policies, such as a ramped-up Earned Income Tax Credit, that are targeted toward regions with more elastic employment responses, however financed, could plausibly reduce suffering and materially improve economic performance."

erences. Given that an optimal tax-and-transfer system depends on welfare weights given to different segments of the population, imposing existing tax systems and behavioral assumptions allows for backing out implied welfare weights. This is most similar Hendren (2020) and Embree (2023), who back out the marginal social welfare weights implied by the federal and state tax systems, respectively, in the United States. Embree (2023) does so based off assumptions of the elasticity of taxable income and marginal propensities to consume for state income and consumption taxes combined. In this paper, I do so based on assumptions of extensive, intensive, and mobility elasticities for approximations of state income tax-and-transfer schedules. Like Hendren (2020), I find governments tend to place more weight on those with low incomes compared to higher incomes. Like Embree (2023), I also find that the implied welfare weights do not decrease monotonically with income as might be expected with common assumptions on social welfare.

This paper proceeds as follows: in Section 2.2, I construct an optimal tax model for a federalist system. In Section 2.3, I conduct my numerical simulations: optimal income transfers by the federal government acting in solitude, optimal income transfers for an average US state, and interactions between the two. I also highlight optimal income transfers for two specific states, Michigan and Massachusetts, and demonstrate the mechanisms driving their choices of optimal transfers. Section 2.4 outlines backing out implied welfare weights from existing state tax and transfer systems and results. Section 2.5 considers policy implications and concludes.

2.2 Model

In this section, I derive the optimal tax and transfer schedule for a federalist system. The first subsection describes the government's problem and parametization of social welfare weights to allow for tractable solutions. The second subsection defines each of the key margins of adjustment for individuals' labor supplies. The final subsection derives and interprets the optimal tax formula.

2.2.1 Setup and Welfare Weights

The setup of the model is an extension of the optimal tax framework first described in Saez (2001) and extended to transfers with particular concern for the extensive labor elasticity in Saez (2002). The main contribution of the model is allowing for multiple jurisdictions to set tax and transfer policy. Consider just one of them and call it state K. I assume that government K's welfare-maximization problem concerns only those within their jurisdiction, and that it chooses its taxes taking the actions of the other jurisdictions as given. In the spirit of Saez (2002), I develop a discrete model of occupational choice: there are $I + 1$ occupations in each state K, with the salary for the nonemployed being $w_{0K} = 0$ and jobs $i = 1, ..., I$.[4] I assume that there is perfect substitution of labor types in the supply-side of the economy, and thus that salaries w_{iK} for occupation i in state K are fixed.[5] Thus, government K's problem of choosing transfers to maximize the utility of its constituents across the income distribution subject to a budget constraint is given by:

$$\max_{T_{iK}} \sum_{i=0}^{I} \mu_{iK} U_{iK} \left(w_{iK} - T_{iK}\right) \tag{2.1}$$

$$\text{subject to } \sum_{i=0}^{I} h_{iK} T_{iK} = H_K \tag{2.2}$$

where the sub-indexes indicate occupation i in state K, μ_{iK} is an individual's welfare weight, U_{iK} are individual's utility functions, w_{iK} is before-tax income, T_{iK} is a tax (or transfer, if negative), h_{iK} is the density of people with occupation i in state K, and H_K is the exogenous revenue requirement for state K.[6]

In effect, the government chooses after-tax income by choosing T_{iK}, which is equivalent to choosing consumption c_{iK} since wages are exogenously given and $c_{iK} = w_{iK} - T_{iK}$. Thus, using p_K as the Lagrange multiplier of the budget constraint for jurisdiction K, the first-order condition of this government's problem is given by:

$$T_{iK} : -\sum_{i=0}^{I} \mu_{iK} \frac{\partial U_{iK} \left(w_{iK} - T_{iK}\right)}{\partial c_{iK}} + p_K \left[h_{iK} - \sum_{j=1}^{I} \frac{\partial h_{iK}}{\partial c_{iK}} T_{jK} \right] = 0 \tag{2.3}$$

where p_K is the marginal value of public funds. The government trades off decreases in welfare from decreased consumption induced by marginally increasing taxes in the first term, with the benefits of marginally relaxing the budget constraint with additional revenue in the second term.

[4]Here, I am borrowing the term "occupation" from Saez (2002)'s setup of the model. It is best to think of occupations not as differing jobs, but as differing income levels from producing a single consumption good as a result of a distribution of idiosyncratic productivities. Differing occupations allow for there to be a distribution of income, where individuals may select into higher pre-tax-and-transfer income jobs. The government is able to observe only income levels and thus can condition taxation only on income instead of, say, skills.

[5]In this paper, I focus on wage income. Low-income individuals are less likely to hold significant assets, and if they do, they are likely to be retirees. This paper is primarily concerned with the employment decisions of low-wealth, low-income, prime-age wage earners.

[6]Like Saez (2002), I consider income taxation and transfers only on the individual level. For a treatment that extends Saez's model to households with children, see Bart and Horton (2024).

This condition can be simplified by directly calibrating marginal social welfare weights. Similar to Saez (2002), I define the marginal social welfare weight for occupation i in jurisdiction K as:

$$g_{iK} = \frac{1}{p_K h_{iK}} \sum_{i=0}^{I} \mu_{iK} \frac{\partial U_{iK}\left(w_{iK} - T_{iK}\right)}{\partial c_{iK}} \tag{2.4}$$

or the dollar-equivalent value for the government of distributing an extra dollar uniformly to individuals working in occupation iK where p_K is the marginal value of public funds in jurisdiction K. Marginal social welfare weights g_{iK} depend on the current tax schedule through c_{iK}, weights μ_{iK}, occupation densities h_{iK}, and utility functions U_{iK}. Directly calibrating g_{iK} still encompasses the classic welfare approach of maximizing a weighted sum of individual utilities (as these weights μ_{iK} can be chosen to yield the marginal weights g_{iK}) while not requiring explicit specifications of individuals' utility functions. Thus, with some rearranging, the first-order condition for the government's problem becomes:

$$\left(1 - g_{iK}\right) h_{iK} = \sum_{i=1}^{I} \frac{\partial h_{iK}}{\partial c_{iK}} T_{jK} \tag{2.5}$$

As in Saez (2002), I posit that the welfare weights g_{iK} are an inverse exponential function of after-tax income. Specifically, the welfare weights can be expressed in a discrete form as

$$g_{iK} = \frac{1}{p_K c_{iK}^{\upsilon}} \tag{2.6}$$

$$\text{subject to } \sum_{i=0}^{I} g_{iK} h_{iK} = 1 \tag{2.7}$$

where υ is a parameter governing redistributive tastes and the second equation provides a normalization of the welfare weights g_{iK}. This parametization allows for simple calibration of a whole range of government redistributive tastes: starting from no redistributive tastes (pure utilitarianism) when $\upsilon = 0$, the government valuing N times less marginal consumption when disposable income is multiplied by N when $\upsilon = 1$, and to the Rawlsian criterion in the limit as υ approaches $+\infty$.

2.2.2 Elasticities

Before the derivation of the optimal tax formula, I first define three labor elasticities that will be key parameters in the optimal tax formula. A key term in these elasticities is the density of workers in occupation i in state K, or h_{iK}. As in Saez (2002), this derivation assumes away income effects, so that all these elasticities are considered compensated.[7]

$$\textit{Extensive Elasticity: } \eta_{iK} = \frac{c_{iK} - c_{0K}}{h_{iK}} \frac{\partial h_{iK}}{\partial \left(c_{iK} - c_{0K}\right)} \tag{2.8}$$

The extensive labor elasticity governs the decision of whether an individual works or not. Specifically, for occupation i in state K, it depends on the relative consumption levels (or disposable income) of the occupation c_{iK} and consumption when unemployed c_{0K}. The elasticity η_{iK} measures the percentage of employed workers in occupation i within state K who decide to leave employment when the difference between disposable incomes for occupation i and unemployment decreases by 1 percent.

$$\textit{Intensive Elasticity: } \zeta_{iK} = \frac{c_{iK} - c_{(i-1)K}}{h_i} \frac{\partial h_{iK}}{\partial \left(c_{iK} - c_{(i-1)K}\right)} \tag{2.9}$$

The intensive labor elasticity governs the decision of whether to select into a higher or lower-paying occupations. For an occupation in state K, this depends on the consumption level associated with occupation i and the nearby occupation $i-1$. Traditionally, intensive elasticities are expressed in the empirical literature as with respect to the wage w_{iK}. Call these elasticities with respect to the wage ε_{iK}. Then, as in Saez (2002), I convert the empirical estimates of intensive elasticities to the elasticity here via $\zeta_{iK} = \varepsilon_{iK} w_{iK}/(w_{iK} - w_{(i-1)K})$. To summarize, this elasticity measures the percentage of employed workers in job $i-1$ who move into job i when the difference between consumption in job i and job $i-1$ increases by 1 percent.

$$\textit{Mobility Elasticity: } \psi_{iK} = \frac{c_{iK} - \bar{c}_{iK}}{h_{iK}} \frac{\partial h_{iK}}{\partial (c_{iK} - \bar{c}_{iK})} \tag{2.10}$$

The mobility labor elasticity governs the decision to switch between the same occupation across states. For occupation i in state K, this depends on the consumption level associated with said occupation and the consumption level associated with the same occupation in different states, indicated by $\bar{c}$. For my purposes, it measures the percentage of employed workers in job i in state K that leave the state within ten years when the difference between "domestic" consumption for job i and consumption in other states for job i increases by 1 percent.[8]

[7]Most studies find that income effects are small relative to substitution effects - see e.g. Blundell and MaCurdy (1999). If income effects were to be included, optimal marginal tax rates tend to be higher, as the loss of income encourages more labor supply.

[8]The choice for ten years comes from evidence that suggests that the effect of income on individual migration decisions is gradual but largely complete over a decade (Kennan and Walker, 2011). Gramlich (1997) estimate the process can take even longer when it came to the Aid to Families with Dependent

Note that the intensive and mobility elasticities implicitly assume an exogenous ordering. Individuals who move between states are assumed to move to the same occupation (that is, the same pre-tax/transder income level) in a different state. Individuals consider only adjacent occupations when determining the extent of their labor supply. This is similar to the construction of behavioral responses in Saez (2001, 2002) under the Mirrlees (1971) assumption that there is a uni-diminsional skill parameter that characterizes each taxpayer and gives rise to a distribution of incomes. Individuals are marginal in their decision-making, and hence only need consider nearby occupations when deciding their labor supply, given that it is positive.

2.2.3 Derivation

I follow the perturbation method of Saez (2001) to derive the optimal transfers and back out the optimal tax rates.[9] That is, consider a perturbation of a slight increase in taxes dT in state K for jobs iK, $(i+1)\,K, \ldots, IK$ given by $dT_{iK} = dT_{(i+1)K} = \ldots = dT_{IK}$. Through a mechanical effect, this tax raises revenue from all occupations and associated densities h_{iK} and above:

$$[h_{iK} + h_{(i+1)K} + \ldots + h_{IK}]dT \tag{2.11}$$

which is the increase in revenue from the tax pertubation dT. Given the marginal welfare weights, this tax revenue is valued by the government as

$$[(1-g_{iK})h_{iK} + (1-g_{(i+1)K})h_{(i+1)K} + \ldots + (1-g_{IK})h_{IK}]dT \tag{2.12}$$

In particular, this tax change alters the difference in consumption given by $c_{iK} - c_{(i-1)K}$ by dT and causes all other differences of $c_{jK} - c_{(j-1)K}$ for occupations $i \neq j$ to remain the same. This causes three separate movements in labor supply and corresponding changes in tax revenue to occur.

One, the difference between consumption while employed and unemployed, or $c_{jK} - c_{0K}$, changes for all occupations $j \geq i$. Via the extensive elasticity, $h_{iK}\eta_{iK}dT/(c_{iK} - c_{0K})$ individuals in each job $j \geq i$ switch to being unemployed as the tax lowers the after-tax benefit

Children program - over 45 years.

[9]For brevity I do not include the graphical illustration here, though I encourage the reader to observe the informative graphs accompanying Saez (2001)'s explanation if they are unfamiliar with his approach.

of employment. This reduces revenue by

$$-dT\sum_{i=j}^{I}h_{jK}\eta_{jK}\frac{T_{jK}-T_{0K}}{c_{jK}-c_{0K}} \tag{2.13}$$

or the sum of all individuals in occupations $j\geq i$ who switch to unemployment and the difference between taxes/transfers for their original occupation iK and non-employment $0K$.

Two, via the intensive elasticity, $h_{iK}\zeta_{iK}dT/(c_{iK}-c_{(i-1)K})$ individuals in job i switch to job $i-1$ because the tax change discourages labor supply at the margin, which causes a change in tax revenue of

$$-h_{iK}\zeta_{iK}dT\frac{T_{iK}-T_{(i-1)K}}{c_{iK}-c_{(i-1)K}} \tag{2.14}$$

or the amount of individuals who switch to occupation $i-1$ and the difference between taxes/transfers for their original occupation iK and new occupation $(i-1)K$.

Three, via the mobility elasticity, $h_{iK}\psi_{iK}dT/(c_{jK}-\bar{c}_{jK})$ individuals in each job $j\geq i$ move to a different jurisdiction because after-tax income for the same job in a different state becomes more lucrative due to the domestic tax increase. For the domestic government, this causes a loss of tax revenue of

$$-dT\sum_{i=j}^{I}h_{jK}\psi_{jK}\frac{T_{jK}}{c_{jK}-\bar{c}_{jK}} \tag{2.15}$$

or the sum of all individuals in occupations $j\geq i$ that switch to occupations in other states times their tax/transfer payments. The domestic state receives no payments from those who leave, nor does it value their utility.

At the optimum, the sum of all of these effects must be zero for state K, as seen in the first-order condition 2.3 above. That is, welfare-weighted utilities across the income distribution and changes in government revenue induced by the behavioral elasticities must balance out. Replacing the first term in equation 2.3 with equation 2.11 and the second term in equation 2.3 with formulas 2.12-2.14 yields:

$$\begin{gathered}\sum_{i=j}^{I}dT\left(1-g_{jK}\right)h_{jK}-dT\sum_{i=j}^{I}h_{jK}\eta_{jK}\frac{T_{jK}-T_{0K}}{c_{jK}-c_{0K}}-\\ h_{iK}\zeta_{iK}dT\frac{T_{iK}-T_{(i-1)K}}{c_{iK}-c_{(i-1)K}}-dT\sum_{i=j}^{I}h_{jK}\psi_{jK}\frac{T_{jK}}{c_{jK}-\bar{c}_{jK}}=0\end{gathered} \tag{2.16}$$

First dividing through by dT, the formula for optimal taxes follows from rearranging:

$$\frac{T_{iK} - T_{(i-1)K}}{c_{iK} - c_{(i-1)K}} = \frac{1}{h_{iK}\zeta_{iK}} \sum_{i=j}^{I} h_{jK} \left(1 - g_{jK} - \eta_{jK} \frac{T_{jK} - T_{0K}}{c_{jK} - c_{0K}} - \psi_{jK} \frac{T_{jK}}{c_{jK} - \bar{c}_{jK}} \right) \tag{2.17}$$

The formula for optimal taxation given by equation 2.17 reveals that the optimal tax for a given job or income level is decreasing in the weight placed on their welfare g_{jK}. The optimal tax decreases the higher the intensive elasticities ζ_{iK}. It also is decreasing in extensive elasticities η_{jK} (combined with how generous after-tax income for unemployment is) and mobility elasticities ψ_{jK} (combined with after-tax income from other jurisdictions). For the optimal federal social planner there is simply one jurisdiction K (the nation as a whole), and barring international movements, the mobility elasticity is always zero, so the last term drops out and the model considers all populations across states at once.

Note that the formula for optimal taxes in 2.17 is technically a formula for transfers, but the implicit marginal tax rate τ_{iK} with respect to earning an extra dollar of pre-tax/transfer income is given by:

$$\tau_{iK} = \frac{T_{iK} - T_{(i-1)K}}{w_{iK} - w_{(i-1)K}} \tag{2.18}$$

where w_{iK} are wages for occupation i in state K. Also note that, given that density weights will be endogenous to the tax-and-transfer system, I use a simplification from Saez (2002) that abstracts from the intensive-margin behavioral responses from h_{iK} and focuses on the extensive margin.[10] The density weights are given by

$$h_{iK} = h_{iK}^{0K} \left(\frac{c_{iK} - c_{0K}}{c_{iK}^{0K} - c_{0K}^{0K})} \right)^{\eta_{iK}} \tag{2.19}$$

where the superscript-0 variables indicate the actual variables of their counterparts above (that is, the current after-tax incomes from tax schedules in reality) and will be calibrated below.

Overall, the system consists $I + 2$ simultaneous equations: the budget constraint 2.2, the welfare weight density constraint 2.7, and the equations for optimal taxes given by 2.17 for

[10]The density weights are endogenous because the distribution of earnings and the unemployment level are affected by taxes and transfers. In principle, the functional form of the weights should be chosen so as to be compatible with the structure of all three behavioral elasticities, as well as coinciding with the empirical weights when the simulated tax schedule is identical to the actual tax schedule. However, it is incredibly computationally intensive (or, as put in Saez (2002), "impossible") to find functions for density weights that are compatible with the range of all three labor elasticities and all possible values of consumption. Focusing on just the extensive labor elasticity simplifies the problem, as well as accounting for the largest swings in employment (zero to non-zero) within a given state.

Table 2.1: National Calibration

Pre-tax Income	Density (h_{iK})	Extensive Elasticity (η_{iK})	Intensive Elasticity (ε_{iK})
\$ 0	16.897%	0.750	0.1000
\$ 1,000	1.097%	0.747	0.1006
\$ 2,000	0.797%	0.744	0.1012
⋮	⋮	⋮	⋮
\$240,000	0.078%	0.030	0.2440
\$245,000	0.067%	0.015	0.2470
>\$250,000	1.319%	0.000	0.2500
		Revenue Requirement (H_K):	\$10,430
Default GMI (c_{0K}^{0K}):	\$6,662	Default Tax Rate:	37%

Notes: Dollar values are expressed in March 2018 dollars.

$i = 1, \ldots, I$. There are $I + 2$ unknowns: taxes/transfers T_{iK} for $i = 0, 1, \ldots, I$ and the marginal value of public funds p. I then calibrate the model and solve the simultaneous equations numerically, bringing us to the results of the numerical simulations.

2.3 Numerical Simulations

2.3.1 Optimal Federal Transfers

I begin by finding the optimal income transfers on a national level as a baseline. For the federal calibration, I calibrate the density of incomes according to the March 2018 CPS Annual Social and Economic Supplement (see Table 2.1 for a truncated version). To focus on the adult-age population eligible for work, I restrict the sample to nonstudents ages 18 through 60. For the intensive and extensive labor elasticities, I use the findings in the literature review compiled by the Congressional Budget Office. Specifically, McClelland and Mok (2012) review estimates for extensive and intensive labor elasticities. Conforming to the given definitions for elasticities, extensive labor elasticities range from 0.3 to 1.2 for low incomes, and intensive labor elasticities are low, around 0.1 for low incomes and 0.25 for high incomes.

I present the baseline federal calibration in Table 2.1, though a few additional notes are necessary and I test the sensitivity of my assumptions below. I do not assume uniform labor elasticities. I set the bottom extensive labor elasticity to 0.75 (the average of 0.3 and 1.2) and phase it out linearly to zero by the income of \$250,000 to reflect the empirical finding that the elasticity falls with income. Similarly, I set the bottom intensive elasticity to 0.1 and linearly phase it up to 0.25 for the income of \$250,000, reflecting the empirical evidence that intensive elasticities rise with income.[11] Given that this initial calibration is for a nation-wide program, I set the mobility elasticity to zero. I set the redistributional preference parameter υ equal to 1, equivalent to the government valuing marginal consumption N times less when disposal income increases N times. Next, I approximate the actual tax schedule linearly by calibrating the GMI according to an empirical measure of federal transfers (see Appendix B.1 for details) to \$6,662 and setting the default marginal tax rate is the top rate of 0.37. The revenue requirement equals the total revenue raised by the individual income tax for the federal government in fiscal year 2018, divided by the working-age population.

The calibration of the labor elasticities and distributional preferences each affect the shape of the optimal tax-and-transfer schedule. To illustrate the sensitivity of the results to each of these parameters, I vary each of them by 50% above and below their baseline values.[12] To observe extremes, I also combine these adjustments to generate the "most progressive" and "least progressive" tax-and-transfer schedules - for example, 50% lower behavioral elasticities combined with a 50% higher parameter for redistributive preferences will generate the most progressive schedule for the values considered.

I show the results for the federal calibration in Figure 2.1 and Table 2.2. For the baseline calibration, I find that the optimal income transfer system includes a guaranteed minimum income of just over 60% of the 2018 federal poverty threshold, or about \$8,250 dollars.[13] The work credit then rapidly phases in, matching earned income nearly dollar-for-dollar at the lowest incomes, with the maximum credit pushing incomes to just over 90% of the federal poverty threshold when combined with earned income, or over \$12,000. The transfers are then phased out fairly rapidly, but at a rate just below the peak marginal rate for high incomes (52% compared to the top high-income marginal rate of 57%), with the credit phasing out entirely when income is almost 170% of the federal poverty threshold, or \$22,000.

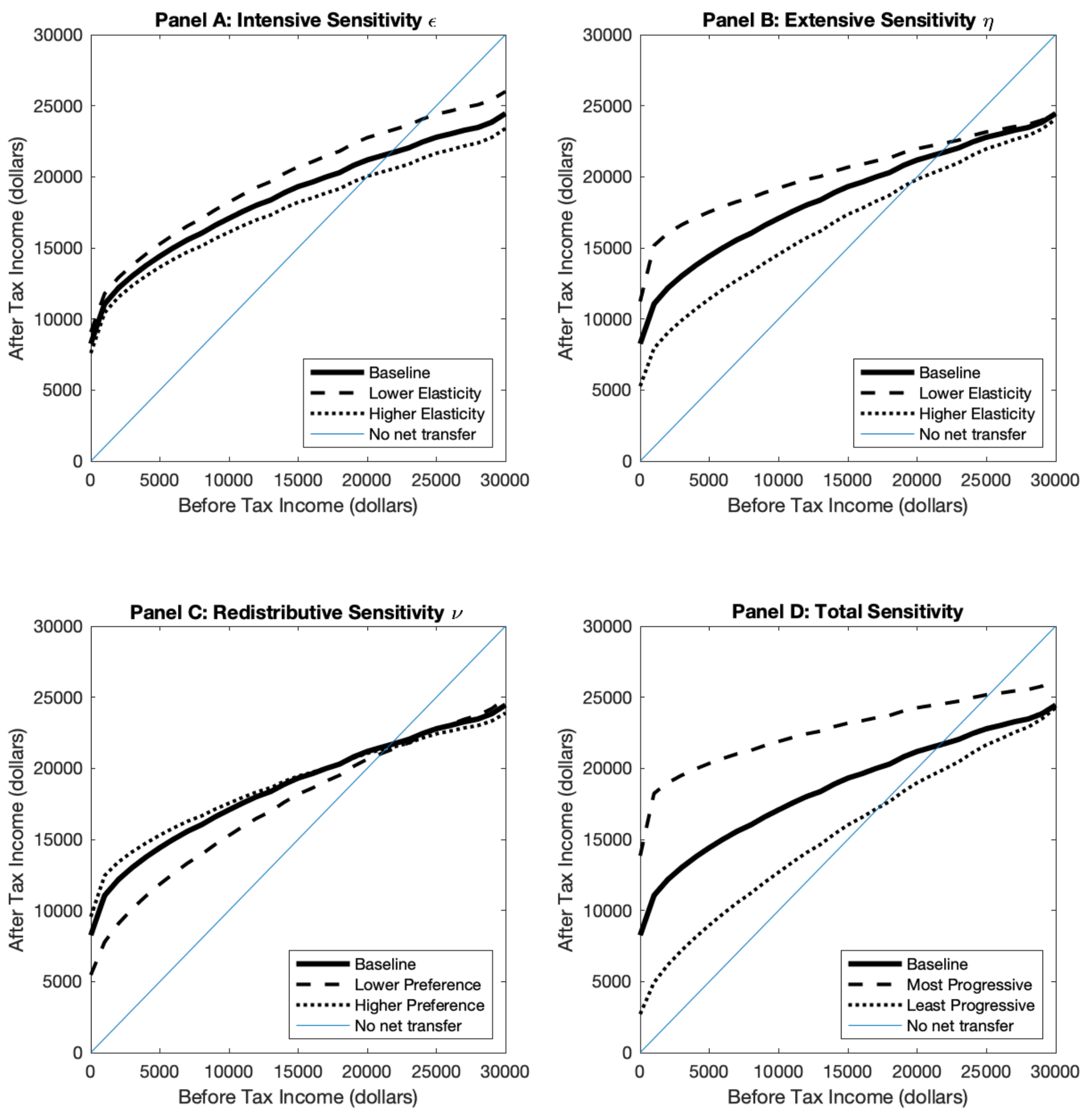
Panels A-C of Figure 2.1 and Table 2.2 illustrate how changing the intensive elasticity, extensive elasticity, and redistributive preference parameter affect the optimal tax-and-transfer

[11] I recognize that elasticities need not be linear with income, and will test the sensitivity of the results to this assumption in future work. The choice of \$250,000 is due to top-coding of the CPS, although this is not far from the \$100,000 and above bracket (in 1992 dollars) used in Gruber and Saez (2002)

[12] Except for the extensive elasticity - given that its baseline value is zero for incomes above \$250,000, its "low" calibration for top incomes remains zero and its "high" calibration for top incomes becomes 0.5.

[13] The 2018 poverty threshold for a single person under age 65 was \$13,064.

Figure 2.1: National Optimal Income Tax Schedules



Notes: The optimal federal income tax schedule, given in March 2018 dollars. The thin blue line represents no net transfers or taxes, when before- and after-tax incomes are identical. The thick black line is the optimal baseline federal income tax schedule. Dashed lines are alternative calibrations

schedule. In general, the taxation of higher incomes is sensitive to the intensive elasticity, the guaranteed minimum income and work subsidy is sensitive to the extensive elasticity, and both aspects of the tax schedule are sensitive to redistributive preferences. Combining all three yields the "least progressive" schedule (high elasticities and low redistribution

Table 2.2: National Optimal Income Tax Schedules

Tax Feature	Baseline	Panel A: Intensive Elasticity		Panel B: Extensive Elasticity	
		Low $\varepsilon \sim [0.05, 0.125]$	High $\varepsilon \sim [0.15, 0.375]$	Low $\eta \sim [0.375, 0]$	High $\eta \sim [1.125, 0.5]$
GMI	\$8,257	\$9,074	\$7,599	\$11,230	\$5,268
Phase-in MTR	97%	-92%	-97%	294%	88%
Peak EITC	\$10,180	\$10,914	\$9,540	\$14,170	\$7,034
Pre-tax Income at Peak*	\$2,000	\$2,000	\$2,000	\$1,000	\$2,000
Phase-out MTR	52%	46%	58%	66%	40%
Break-even point*	\$22,000	\$24,000	\$20,000	\$22,000	\$20,000
High-Income MTR:	57%	70%	48%	56%	46%

Tax Feature	Baseline	Panel C: Redistributive Parameter		Panel D: Total Sensitivity	
		Low $\nu = 0.5$	High $\nu = 1.5$	Most progressive	Least progressive
GMI	\$8,257	\$5,470	\$9,559	\$13,852	\$2,708
Phase-in MTR	97%	55%	191%	339%	50%
Peak EITC	\$10,180	\$7,120	\$11,465	\$17,240	\$4,201
Pre-tax Income at Peak*	\$2,000	\$3,000	\$1,000	\$1,000	\$3,000
Phase-out MTR	52%	39%	56%	71%	29%
Break-even point*	\$22,000	\$21,000	\$21,000	\$25,000	\$17,000
High-Income MTR:	57%	43%	59%	72%	34%

Notes: Dollar values are expressed in March 2018 dollars. Noted tax features include the GMI (guaranteed minimum income), the phase-in MTR (the rate at which the maximum income subsidy phases in with income), the peak EITC (the maximum value of the work subsidy), the pre-tax income at peak (the amount of income at which the peak EITC is earned), the phase-out MTR (the rate at which the subsidy phases out with income), the break-even point (at which the individual receives no net transfers or taxes), and the high-income MTR (the marginal income tax rate for top incomes). Panel D combines adjustments to generate the most and least progressive tax-and-transfer schedules for considered parameter values. Baseline parameter values are given in Table 1. * indicates within \$1000 of income.

preferences) and the "most progressive" schedule (low elasticities and high redistribution preferences) as shown in Panel D of Figure 2.1 and Table 2.2. Depending on the calibration, the optimal GMI may range from \$2,708 to \$13,852 with the maximum work subsidy reaching \$4,201 to \$17,240. The more progressive the calibration, the greater the phase-in rate of the work subsidy, the phase-out rate, and the marginal tax rate on top incomes.

For the baseline calibration, the optimal federal tax-and-transfer system is more progressive than the linear approximation of current reality. The GMI increases by nearly \$1,600 and the top marginal tax rate on income rises by twenty percentage points. This is not out of line with much of the literature on optimal taxation - indeed, Saez (2001); Gruber and Saez (2002), and Saez (2002) all find optimal top marginal income tax rates are almost always above 50%.

2.3.2 State Transfers and Mobility

I now turn to the optimal income transfers from the perspective of states. States start out taking a linear approximation of the federal tax-and-transfer system as given. That is, states assume that the federal government sets the GMI and tax rate to its values in Table 2.1. The federal government's tax-and-transfer schedules matters to states, as their maximization problem depends on after-tax consumption of individuals.[14] Of primary interest here is how states react when taking into account all three labor elasticities: extensive, intensive, and mobility. For this calibration, I consider this scenario to be that of an "average state." That is, for the default income schedule, I created a database of transfer programs, whose generosity varies across states, to calibrate the GMIs (see Appendix B.1), from which I use their average of \$2,227. The default MTR is an average of the states' MTRs (6.405%), and the revenue requirement is an average of the revenue raised by each state's individual income tax divided by the adult population.[15] As a comparative exercise, I first derive the results for when states do not account for mobility of the working-age population – that is, I first set the mobility elasticity ψ_{jK} to zero for each state.

Next, I derive the optimal income transfer system for when states take into account the mobility elasticity. The rest of the calibration is the same as for "non-mobile" states, except for the introduction of a granular mobility elasticity that varies across incomes. I calibrate the elasticity of mobility for bottom incomes to 0.5, as found in Kennan and Walker (2011), and phase it up to about 19 for incomes of \$160,000 before phasing it down to about 10 for incomes of \$250,000 and above, as found in Gordon and Cullen (2012).[16] I set the other state income schedules that this average state reacts to, $\bar{c}_{iK}$, to the default state averages of their GMI and linear approximation of taxes. Hence if the "average state" used in this scenario

[14] I assume that approximating the real-world federal schedule is most informative for policy relevance, but I present an alternative calibration for when states believe the federal government is acting according to their optimal income transfer schedule in Appendix B.2. An alternative specification, rather than states taking after-federal-tax incomes as given, would be a simultaneous game between federal and state governments. However, most evidence points towards states reacting to federal fiscal policy rather than vice-versa (Czerwinski and McCool, 2011).

[15] As an example of the variation across states, before accounting for the fraction financed by the federal government, New Hampshire has a GMI of \$15,234 and Mississippi has a GMI of \$4,581. There are 9 states without individual income taxes. For these, I set their MTR and revenue requirements to the average state MTR and revenue requirements from when they are taken out of the sample.

[16] The high-income mobility elasticities seem incredibly high - an elasticity of 19 here implies that over 10 years, a 1% decrease in after-tax consumption for an occupation relative to other states causes a 19% decrease in that occupation's population density. These high elasticities are necessary to generate realistic top marginal state income tax rates, as in Gordon and Cullen (2012). This is a potential weakness - Kleven et al. (2020) note that mobility elasticities are not exogenous, structural parameters, and that they can vary greatly depending on the population being analyzed, the size of the tax jurisdiction, the extent of tax policy coordination, and a range of non-tax policies. Future work should endogenize these mobility elasticities, particularly for higher incomes.

Table 2.3: Average State Optimal Income Tax Schedule

Tax Feature	Without Mobility			With Mobility		
	Baseline	Most progressive	Least progressive	Baseline	Most progressive	Least progressive
GMI	\$6,539	\$12,022	\$8,143	\$4,413	\$8,513	\$3,484
Phase-in MTR	129%	121%	74%	324%	435%	39%
Peak EITC	\$9,116	\$15,903	\$3,423	\$7,656	\$12,860	\$4,660
Pre-tax Income at Peak*	\$2,000	\$2,000	\$2,000	\$1,000	\$1,000	\$3,000
Phase-out MTR	52%	72%	27%	41%	66%	32%
Break-even point*	\$20,000	\$23,000	\$15,000	\$20,000	\$20,000	\$18,000
High-Income MTR	20.0%	35.0%	15.0%	5.0%	9.0%	-5.0%

Notes: Dollar values are expressed in March 2018 dollars. Noted tax features include the GMI (guaranteed minimum income), the phase-in MTR (the rate at which the maximum income subsidy phases in with income), the peak EITC (the maximum value of the work subsidy), the pre-tax income at peak (the amount of income at which the peak EITC is earned), the phase-out MTR (the rate at which the subsidy phases out with income), the break-even point (at which the individual receives no net transfers or taxes), and the high-income MTR (the marginal income tax rate for top incomes). "Most" and "least progressive" combine considered parameter values to generate alternative tax-and-transfer-schedules. Baseline parameter values are given in Table 3. * indicates within \$1000 of income.

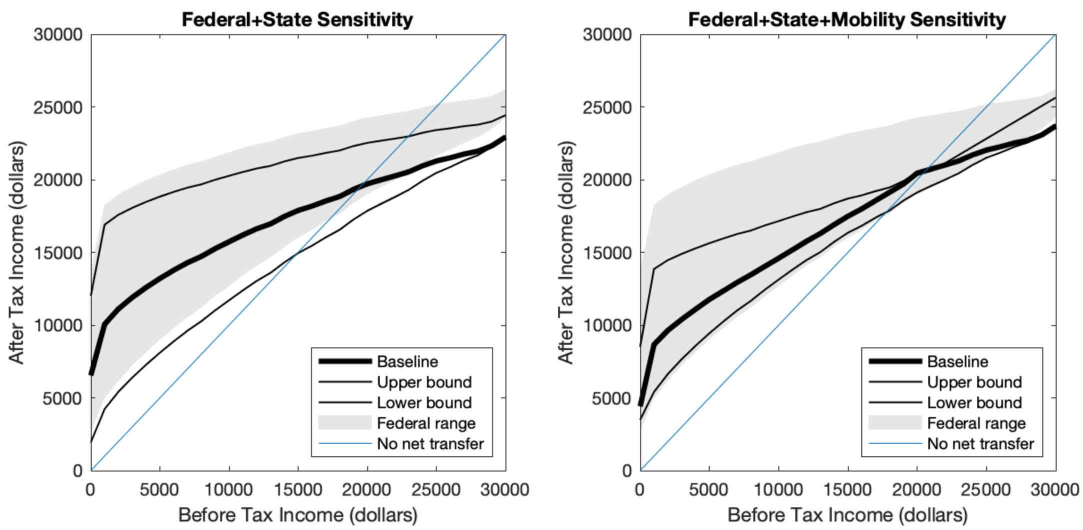
changed nothing, then nobody will move. However, if this average state finds it optimal to change its tax schedule, then people may move to the default average state according to their mobility elasticity.[17] I test the sensitivity of the optimal schedules to my assumptions below.

The results for the "mobile" and "non-mobile" states are presented in Table 2.3 and Figure 2.2. There are several noteworthy results from the state calibrations, particularly when the states set the transfer system taking into account all three labor elasticities. First, on average and even without the mobility elasticity, states find it optimal to set less generous tax-and-transfer schedules than just the federal government. This lends credence to Gordon (1983)'s finding that income distribution from state governments rather than a central authority may lead to under-provision. Fiscal constraints primarily drive this phenomenon. That is, states face tax rates on high incomes by the federal government that are already sizable and have GMIs that are already small compared to the federal GMI. Thus, on average, they reduce the total income transfer in order to raise their necessary revenue without significant tax increases on higher incomes, beyond that already imposed by the federal government.

The results are even more stark when states respond to the potential movement of individuals across state lines. On average, states substantially reduce the total GMI by nearly

[17] I also calculated both the non-mobile and mobile results by numerically solving for the optimal income transfer system for each state individually. The average of those results is hardly different from the results using the "state averages" values here, but is much more computationally intensive.

Figure 2.2: Average State Optimal Income Tax Schedule



Notes: Optimal income tax schedules, given in March 2018 dollars. The thin blue line represents no net transfers or taxes, when before- and after-tax incomes are identical. The solid black lines represent optimal federal + state income tax schedules, without and with accounting for mobility, respectively. The thin black lines represent "most progressive" and "least progressive" combinations of considered parameter values. The shaded grey area is the range of most to least progressive calibrations for the federal government only, for comparison.

50% compared to when the federal government took sole responsibility, from about \$8,250 to about \$4,400. This is because states are very averse to attracting no-income people from other states despite high welfare weights, given that they are solely a drain on fiscal space from the state government's perspective. However, the phase-in for the EITC is substantial, at over \$3 for every dollar of earned income, and the phase-out rate falls to 41% from the phase-out of 52% for the optimal federal program. This reflects that states are less fearful of supplementing income as opposed to providing incomes to those with none. It also reflects the optimality of increasing the relative difference of potential consumption between no-income and low-income individuals, which will be elaborated on below. Finally, given the high mobility of top incomes, states react by quartering their top marginal rates on high-income earners, from 20% to 5%. This also puts downward pressure on positive income transfers given the need to raise a certain amount of revenue.

For the baseline calibration, the optimal tax-and-transfer system set by states while accounting for mobility is less progressive than in reality. States reduce the total GMI by over 50% from the default total of \$6,662 to \$3,167. The average top state marginal income tax rate rises only slightly, by just over 2.5 percentage point. This suggests that current arrangements already correct to some extent for a perceived under-provision of transfers that would prevail if states were left entirely to their own devices. To test the sensitivity of the results, I adjust each behavioral elasticity - including the mobility elasticity - and the redistributive preference parameter to 50% above and below their baseline values, as in Section 2.3.1. The combination of low elasticities and high redistributive preferences generates the "most progressive" tax-and-transfer schedule, and high elasticities and low redistributive preferences the "least progressive." In all state simulations, the range of tax-and-transfer schedules from least to most progressive lies underneath the same range for federally-set schedules. The inclusion of mobility reduces the overlap between state-set and federally-set schedules from about 80% to about 50%.

2.3.3 A Comparative Exercise: Michigan and Massachusetts

In their paper for the Brookings Institution, Benjamin and Summers (2018) find "local labor demand has more impact on the not-working rate in places where non-employment is high than in places that are already near full employment" and argue that "this heterogeneity is crucial in justifying spatially heterogeneous policies that encourage employment more in some areas than in others." Their model "predicts that the ratio of consumption of not-working to employed... should indeed be lower in areas with high not-working rates." With these ideas in mind, I now put them under further scrutiny with an exercise in solving for the optimal income transfers for Michigan and Massachusetts.

As seen under the calibration in Table 2.4, Michigan's pool of no-income adults is just over 35% larger than Massachusetts', which has considerably more high-income earners as well. In addition to calibrating the income densities, I also change the base amount from which the extensive elasticities phase out to zero. That is, following the evidence in Benjamin and Summers (2018) that extensive elasticizes are greater in states with high non-employment, I calibrate the low-income elasticity to 1.1 and 0.4 for Michigan and Massachusetts, respectively. All other elasticities remain the same, the default GMIs and revenue requirements are calculated as above but for each state specifically, and the default MTRs correspond to their top marginal rates. Also note that I continue to set the other state income schedules that states react to, $\bar{c}_{iK}$, to the default average across states. I numerically solve for two cases in each state: when they do not account for the mobility elasticity, and when they do. I present the results in Table 2.5 and Figure 2.3.

What stands out most starkly is how much more generous Massachusetts' optimal income transfer system is compared to Michigan's, even going as far as to supplement the federal transfer. However, the impact of this finding is muted – although not entirely – if one keeps in

Table 2.4: Michigan and Massachusetts Calibrations

	Michigan		Massachusetts	
Income	Density (h_{iK})	Ext. Elast. (η_{iK})	Density (h_{iK})	Ext. Elast. (η_{iK})
\$ 0	16.633%	1.100	12.218%	0.400
\$1,000	0.862%	1.096	1.338%	0.398
\$2,000	0.822%	1.091	0.732%	0.397
⋮	⋮	⋮	⋮	⋮
\$240,000	0.046%	0.044	0.263%	0.016
\$245,000	0.053%	0.022	0.164%	0.008
\$250,000	0.875%	0.000	2.246%	0.000
	Default GMI (c_{0K}^{0K}):	\$1,536	Default GMI (c_{0K}^{0K}):	\$5,577
	Default Tax Rate:	4.25%	Default Tax Rate:	5.10%
	Revenue Req. (H_K):	\$1,816	Revenue Req. (H_K):	\$4,098

Notes: Dollar values are expressed in March 2018 dollars.

mind that the 2018 price level is nearly 16% higher in Massachusetts compared to Michigan. Two primary factors drive the rest of the difference, as discussed in turn.

The first is the income densities. Massachusetts has substantially more high-income residents compared to Michigan, and a lower number of no-income residents. This allows them to raise more revenue and make income transfers more generous. However, given the high mobility elasticity of high-income individuals, Massachusetts loses much of its ability to redistribute when taking into account potential citizen movement, and to a much greater degree than Michigan. For example, Massachusetts' peak EITC falls by over 20% from \$17,182 to \$13,574 when taking into account mobility, but for Michigan the peak EITC falls by almost 12.5% from \$7,117 to \$6,230.

The second is the difference in extensive elasticities. Michigan's higher extensive elasticities make its residents more prone to dropping out of the labor force, thus making lower benefits at zero income more optimal to incentive employment and avoid paying large transfers to many no-income earners. That being said, the model confirms Benjamin and Summers (2018) finding that relative consumption between the working and non-working should be greater in states with higher non-employment. With the mobility elasticity switched on, Massachusetts grows the gap between the transfer for no-income to those earning at least \$1000 of income by almost 35%, but for Michigan this gap grows nearly 93%. For compar-

Table 2.5: Comparative State Optimal Income Tax Schedule

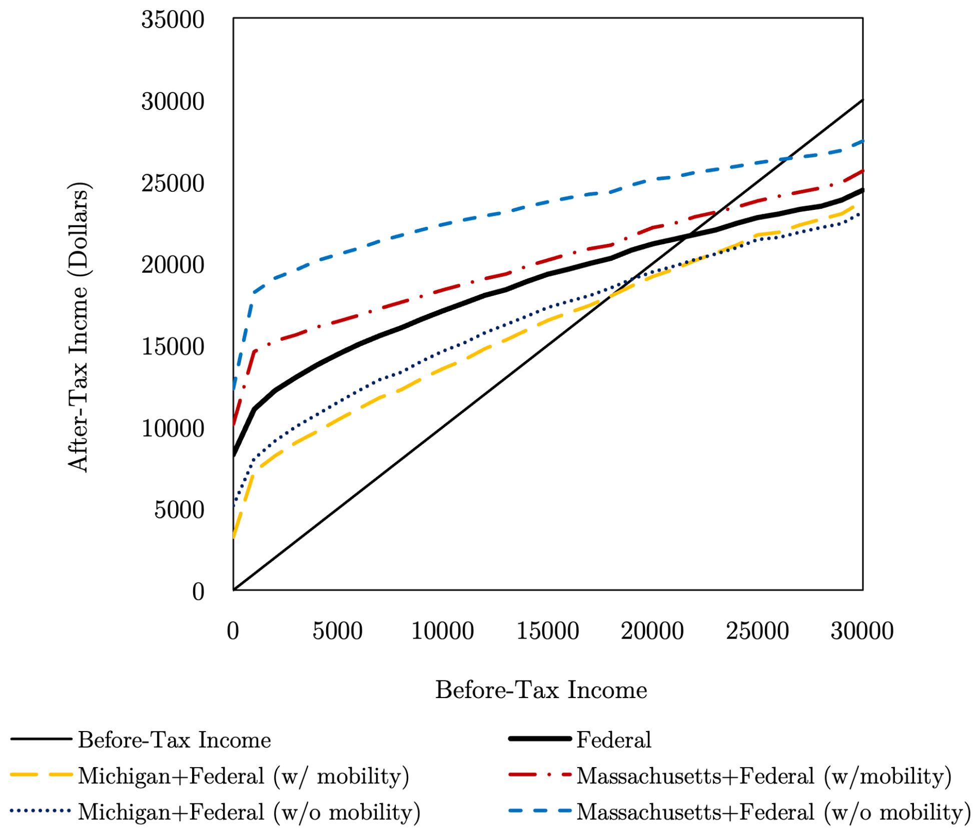
State	Without Mobility		With Mobility	
MI	GMI:	$5,160	GMI:	$3,234
	MTR for phase-in:	98%	Phase-in MTR:	300%
	Peak EITC:	$7,117	Peak EITC:	$6,230
	Pre-tax Income at Peak*:	$2,000	Pre-tax Income at Peak*:	$1,000
	Phase-out MTR:	42%	MTR for phase-out:	37%
	Break-even point*:	$19,000	Break-even point*:	$18,000
	Top State MTR:	10%	Top State MTR:	7%
MA	GMI:	$12,323	GMI:	$10,122
	MTR for phase-in:	486%	Phase-in MTR:	345%
	Peak EITC:	$17,182	Peak EITC:	$13,574
	Pre-tax Income at Peak*:	$1,000	Pre-tax Income at Peak*:	$1,000
	Phase-out MTR:	67%	MTR for phase-out:	61%
	Break-even point*:	$26,000	Break-even point*:	$23,000
	High-Income MTR:	17%	High-Income MTR:	8%

Note: Dollar values are expressed in March 2018 dollars. Noted tax features include the GMI (guaranteed minimum income), the phase-in MTR (the rate at which the maximum income subsidy phases in with income), the peak EITC (the maximum value of the work subsidy), the pre-tax income at peak (the amount of income at which the peak EITC is earned), the phase-out MTR (the rate at which the subsidy phases out with income), the break-even point (at which the individual receives no net transfers or taxes), and the high-income MTR (the marginal state income tax rate for top incomes).

ison, the average state in Table 2.3 and Figure 2.2 grows the transfer from no to earned incomes by nearly 74% when taking into account mobility.

Both states exhibit aversion to guaranteed minimum incomes, in particular with the mobility elasticity in effect. This is especially true for Michigan, whose optimal total GMI falls by over 37% from $5,160 to $3,234 when going from "non-mobile" to "mobile," whereas Massachusetts' falls by almost 18% from $12,323 to $10,122. However, the optimal phase-in rate nearly triples for Michigan, and actually declines for Massachusetts, reflecting that Michigan has a greater ability and incentive – given high extensive labor elasticities and

Figure 2.3: Michigan and Massachusetts Optimal Income Tax Schedules



Note: Optimal income tax schedules, given in March 2018 dollars. The thin black line represents no net transfers or taxes, when before- and after-tax incomes are identical. The solid black line represents the optimal federal income tax schedule. Dashed lines combine the default federal income tax schedule with the optimal state income tax schedule for Michigan and Massachusetts, with and without allowing for mobility across states.

higher unemployment – to encourage employment compared to Massachusetts.

In Appendix B.3, I plot the optimal tax and transfer schedule for every US state and DC following the criteria above, but keeping the bottom extensive elasticity at 0.75 for all states, as in the national calibration. For comparison, I also plot the default linear approximation of the jurisdictions' tax-and-transfer system. Almost all jurisdictions, when left to their own devices, provide less generous transfers to no-or-low-incomes than the total of federal and state transfers in the approximated reality, again suggesting that current arrangements already somewhat correct for perceived under-provision. The guaranteed minimum income can vary considerably from state to state, with most states' ranging from \$1,000 to \$10,000. However, all states generate differences between transfers received when earning no income to at least some income: the difference between transfers received at \$0 of income and the maximum income subsidy for each state range from \$1,117 to \$3,486.

2.4 Implied Welfare Weights

I now investigate what social preferences that would be necessary for actual state tax-and-transfer systems to be optimal - that is, how do governments apparently weigh individuals based on income when setting their systems? The optimal tax and transfer system developed in Section 2 may be used to back out welfare weights g_i implied by actual tax-and-transfer systems, shedding light on underlying social welfare preferences. To back out the weights, I must assume that individuals are responding optimally according to their behavior in equations 2.8-2.10 and governments optimize according to the optimal tax and transfer system in equation 2.17. From equation 2.17, solving for a single welfare weight g_i yields:

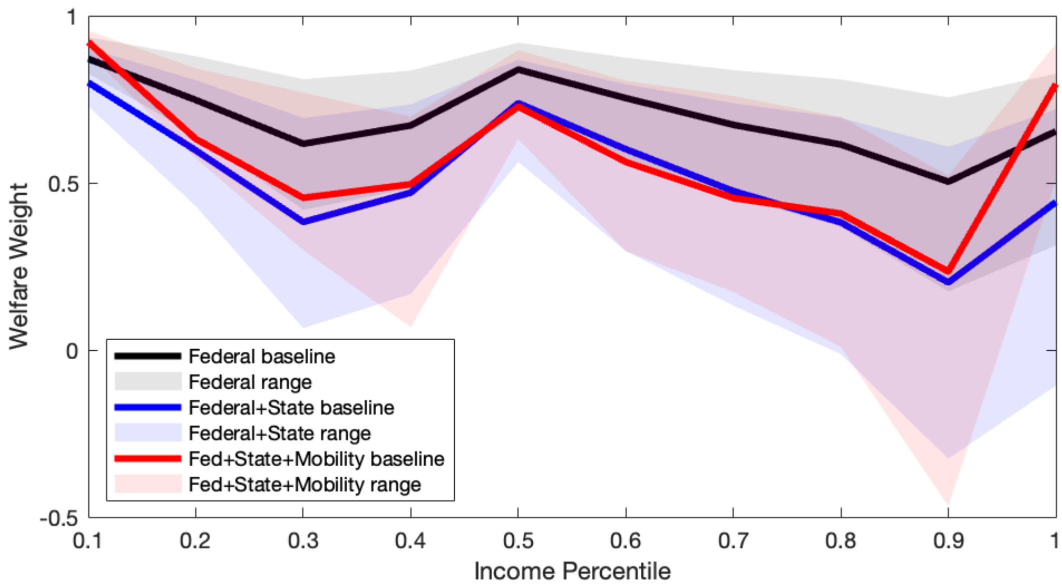
$$\begin{aligned} g_{iK} = 1 - \eta_{iK}\frac{T_{iK} - T_{0K}}{c_{iK} - c_{0K}} - \psi_{iK}\frac{T_{iK}}{c_{iK} - \bar{c}_{iK}} - \zeta_{iK}\frac{T_{iK} - T_{(i-1)K}}{c_{iK} - c_{(i-1)K}} + \\ \frac{1}{h_{iK}}\sum_{i=j+1}^{I} h_{jK}\left(1 - g_{jK} - \eta_{jK}\frac{T_{jK} - T_{0K}}{c_{jK} - c_{0K}} - \psi_{jK}\frac{T_{jK}}{c_{jK} - \bar{c}_{jK}}\right) \end{aligned} \tag{2.20}$$

This combined with the normalization that the weighted sum of the welfare weights g_{iK} sum to 1 as in equation 2.7 yields a system of $I+1$ equations: equation 2.20 for each $i = 1, \ldots, I$ g_{iK} and equation 2.7. In this system, the densities h_{iK} correspond to the empircal density, and transfers T_{iK} are the default linear approximations of government tax and transfer systems outlined in 2.3.2. The resulting welfare weights, as outlined in Section 2.2.1, fit a broad variety of utility functions.

Figure 2.4 plots the resulting welfare weights under alternative assumptions of the optimization problem for governments.[18] In black I plot the welfare weights resulting from fitting the federal model to the data, in blue the federal + state model, and in red the federal + state model with mobility. Given that the resulting welfare weights depend on assumptions of behavioral elasticities, I plot results for the weights when these are varied 50% above and below their baseline values in the shaded areas. Under higher elasticities, optimal tax systems tend to be less progressive than in reality. Thus, assuming higher elasticities vis a vis lower elasticities generates a greater spread between weights placed on low vs high incomes in order to rationalize actual, more progressive systems.

[18] I use less granular categories of income than the preceding figures due to considerable noise in the estimates. That is, as in Embree (2023), the welfare weights generate an oscillating pattern across income. The plotted values are the average welfare weights for the given income bins.

Figure 2.4: Implied Welfare Weights



Note: Implied welfare weights under the assumption that individuals and state governments are behaving optimally given assumptions about labor elasticities, and responding to a linear approximation of tax-and-transfer schedules. The black line is the welfare weights estimated from the baseline federal-only model. The blue line is the weights estimated from the federal+state model. The red line is the weights estimated from federal+state model with mobility. Shaded areas are ranges of welfare weight estimates under different elasticities.

The resulting welfare weights are non-monotonic, suggesting that governments do not value the welfare of individuals as a purely declining function of income. This violates a common assumption of social welfare functions, including those used above, in which welfare weights continuously decline with income.[19] As shown in Figure 2.4, this non-monotonicity is robust to a range of assumptions on behavioral elasticitities. Welfare weights increase as incomes rise from the 30th to 50th percentiles of income, before generally declining with income. Nevertheless, welfare weights for individuals earning less than 30th percentile of income are higher than for those earning at the 30th percentile. That is, in order to generate the GMIs as seen in the data, governments must be placing a greater weight on those with little to no income than slightly higher-income households. Besides at the lowest percentile, the highest welfare weights often take place at the 50th percentile of income. Given that

[19]I am not the first to find this result - Embree (2023) finds the same. Moreover, Hendren (2020) in a paper backing out weights from the US federal system also finds non-monotonicity in weights for top incomes, reflecting a change in the income distribution from a log-normal shape to a Pareto distribution. His federal estimates are more negatively-sloped than mine, primarily due to his assumption of higher compensated labor elasticities.

this is where the bulk of incomes reside, this suggests a potential political economy angle - perhaps governments (or those who elect their governments) place the most value on where the most potential votes reside.

There does not appear to be a large differences in the shape of welfare weights between levels of government, with and without taking into account mobility. Similarities between weights result from estimating off similar tax systems: the federal tax system with the federal government, and federal + state tax systems with federal + state governments where states take the federal tax system as given, as in Section 2.3.2. In reality, state taxes and transfers tend to be dwarfed by the federal government's. Hence, in order to rationalize existing federal + state tax systems, the shape of welfare weights for the average state is not too dissimilar to the shape of welfare weights for the federal government.

Nevertheless, important differences remain across levels of government and their implied welfare weights. Overall, the average state with mobility needs to prefer more redistribution relative to the federal government - or state governments that do not account for mobility - in order to rationalize linear approximations of their tax-and-transfer systems. States with mobility place a higher weight relative to the federal government on incomes in the 10th percentile, before placing lower relative weights on higher incomes. As income increases, the gap between the federal and state weights widens. This result is suddenly reversed for the top decile of incomes, while implied welfare weights by states with mobility are higher than their federal counterpart.

What could be driving these differences? As seen in Section 2.3.2, states desire to set less progressive tax-and-transfer systems than the federal government, for fear of attracting low-or-no-income earners and driving away high-earners. In reality, states tend to supplement federal transfers. This implies that to rationalize the average tax system of state governments that take the federal tax system as given, states must be more progressive in their weighting. The reversal in weighting for top incomes stems from the particularly high mobility elasticities of these individuals, which other calibrations do not account for by construction. States, in contrast, fear losing high-income taxpayers. In order to rationalize low top marginal state income tax rates, a greater weight must be placed on these taxpayers.

In Appendix B.3, I chart the ratio of welfare weights for incomes below to above median income under mobility for every US state. This reveals that using an "average state" masks some variation, with ratios varying from about .8 to about 1.25 (for the baseline case above, the value is 0.98). Some notable patterns, perhaps contrary to expectations, emerge: some of the highest ratios are in the Deep South and rust belt states, despite other traditionally Democratic states having much more progressive tax systems. This is in part explained by empirical income densities: these states in the have some of the highest fractions of no-and-low-income earners. In order to generate the approximated linear income tax systems while meeting revenue requirements, welfare weights for low-income individuals in these states must be considerably higher than for top incomes. In contrast, other states, such as California, have some of the highest densities of top-income earners and lower amounts of no-or-low-income individuals. Given that high-income individuals are assumed to be more mobile, this increases the welfare weights on top incomes relative to low earnings.

2.5 Policy Implications and Conclusion

The above results lead to several policy implications. On average, sub-national jurisdictions find it optimal to provide less generous transfers to no-or-low-income earners than a purely national system. A back-of-the-envelope calculation may be informative: consider no-or-low-income potential after-tax-and-transfer consumption across incomes ranging from zero to the break-even point of the work credit. Then on average, consumption for these lower incomes decreases by 20.57% from the federal transfer system when states set their own system, and by 46.98% from the federal system when states set their own system and take into account mobility.

Thus, if the federal government is concerned about welfare along these lines, it should make welfare payments untaxable so that states are unable to use them as sources of revenue. It in fact already does this for TANF, WIC, and SNAP payments, but unemployment benefits are taxable. Although, along the lines of Austin, Glaeser, and Summers (2018), it may be optimal – and an improvement in the sense of increasing employment – for states to be able to tax benefits in order to increase the difference in consumption between those with earned income and not, as they may lack the fiscal capacity to expand the EITC to accomplish the same effect. An alternative may be for the federal government to expand the EITC in particular states to get the desired difference in consumption between those with no and earned income if states are unable to do so, but again, such benefit expansions would have to be untaxable.

Compared to the empirical calibration, matching the prescription of optimal income transfers on the federal level would be a significant fiscal undertaking. Starting from the calibration in Table 2.1, if the federal EITC were expanded to match the results of Table 2.2, federal expenditures would increase by about \$175-\$245 billion annually, or nearly \$2.1 trillion over the 2018-2027 budget window.[20] For comparison, this is roughly the same amount that the

[20] I calculated this using Tax-Calculator, an open-source federal individual income and payroll tax microsimulation model maintained by the Open Source Policy Center (OSPC) via the American Enterprise Institute (AEI). Specifically, I increased the maximum EITC for each household by \$4,778 (regardless of the number of children) to match the prescribed peak EITC. I then made about 30% of the maximum credit

2017 Tax Cuts and Jobs Act costs over the 2017-2026 budget window if all of its provisions are made permanent.

This work leaves open multiple avenues for expansion. For one, this model does not take into account the negative externalities of non-employment and the positive externalities of the EITC.[21] Hence, the optimal GMI may be an upper-bound and optimal EITC a lower-bound. Another is calibrating the model to work within the filing-unit-based framework of the current US tax system, rather than be individually-based.[22] Finally, directly calibrating mobility elasticities leaves much to be desired - as suggested by Kleven et al. (2020)), future work should endogenize these elasticities.

Now recall our motivating question – what would an optimal income transfer system look like, taking into account the potential for both federal and state-level programs? To answer this, I developed a model that accounts for three labor elasticities: extensive, intensive, and mobility. I then calibrate the model to a database of state GMIs and income densities. I find that in every case, optimal income transfers involve a combination of a guaranteed minimum income with a phased-in-and-phased out work credit – an "Earned and Basic Income Tax Credit" or EBITC, as it were.[23] On average, the optimal income transfers from the perspective of states is lower than that offered by the federal government, given their fiscal constraints. When states take into account the mobility of the adult population, the after-tax transfers are even less generous on average. However, states with a greater amount of high-income earners and a lower share of no-income workers find it optimal to supplement the federal transfers. Nevertheless, states with greater non-employment set transfers so that they increase the relative difference between no-income and earned income consumption, compared to states with greater employment.

payable at zero income, so as to increase the GMI by about $1,500 to match the actual to the prescribed transfer. Going on, I increased the phase-in rate to 97% for every level of the EITC, and the phase-out rate to 52% starting at an earned income of $2,000, per Table 2. I also lowered the eligibility age from 25 to 18 and made the credit individual-based rather than filing-unit-based. Overall, these changes increased the after-tax income of the bottom 20% of the income distribution by about 26.2%. In all, 52.4% of filers would receive tax cuts, and 8.4% would face tax increases, mostly in the 20%-40% income quintile, due to the steeper EITC phase-out.

[21] The negative externalities include fiscal (less taxes and more spending) as documented in Benjamin and Summers (2018), social as in Killewald (2016), and spillovers as in Conley and Topa (2002). Positive externalities of the EITC flow through increased education as in Bastian (2018) and Bastian and Michelmore (2018) and increased labor supply as in Bastian and Jones (2021).

[22] Bart and Horton (2024) do this for households with children, although not in a federalist framework.

[23] I am indebted to economist Arindrajit Dube at the University of Massachusetts, Amherst for this phrase.

CHAPTER 3

A Transparent Look at How Taxes Affect Growth: Evidence from Cross-Country Panel Data

Coauthored with Meng Hsuan Hsieh, Laura Kawano, and Joel Slemrod

3.1 Introduction and Motivation

No question about taxation is more important than its impact on the level of economic prosperity and its growth. For this reason, it has drawn an enormous amount of research attention from both macroeconomists and public finance economists, with the research designs used differing on a number of dimensions such as the time horizon, level of aggregation, and estimation method. One of the most common research strategies analyzes the medium-term effect of the tax system on national output using cross-country panel data. Even within the literature that adopts this approach, no clear answer has emerged.

In what follows, we revisit this issue. We begin with a critical review of the literature that attempts to shed light on why, with an inherently small set of common data that in principle all the analyses share, no consensus set of findings has emerged. We address what are the crucial research design choices that lead to divergent results.

Building on the insights from this exercise, we then pursue our own analysis of the cross-country panel data. Our approach follows several principles. First, we focus on data analysis that is transparent, often graphical, and therefore is inevitably relatively simple. Second, we concentrate on the impact of tax systems on medium-term prosperity, defined here as GDP growth five years later, and thereby de-emphasize the short-term effects of tax policy on business cycles; this reduces, but does not eliminate, the need to carefully distinguish so-called endogenous and exogenous policy changes. Third, we take seriously that tax policy is multi-dimensional, and therefore make use of a newly-assembled data set containing tax rates of the three major categories of taxes supplemented by information about the bases to which each of these tax rates apply. Finally, in contrast to much (but not all) of the macroeconomics literature on this topic, we eschew measuring changes in the tax system with changes in the associated level of tax revenues; we do so because revenues are the product of arguably exogenous parameters, such as rates and aspects of bases, with a certainly (or at least hypothesized to be) endogenous base such as aggregate income or consumption.

We conclude that the analysis of modern cross-country aggregate panel data does not credibly support any claim that prominent aspects of tax policy have a statistically robust medium-term impact on national output. This is true regardless of whether the set of tax changes studied is pared by eliminating consideration of changes induced by current macroeconomic conditions and therefore endogenous. Earlier findings that purport to establish the role of reduced taxation to stimulate national production are based on arguably endogenous tax policy shocks. Consistent with the narrative approach literature, we find that removing policy changes that are deemed endogenous with respect to the trajectory of the economy attenuates the estimated effect of taxes on economic growth. However, much of this literature measures tax policy shocks as changes in tax revenues as a percent of GDP, which is also likely endogenous even when limiting attention to the narratively-determined exogenous policy changes. When we instead use exogenous statutory tax rate changes, we find no statistically significant relationship between tax rate changes and economic growth.

We further assess the literature in light of the recent econometric advances that have made clear that in a setting like this—where countries generally undergo tax policy shocks at multiple times—the methodologies that have been employed may yield biased causal treatment effect estimates. The central problem is the difficulty of finding a valid comparison group that serves to approximate the evolution of outcomes had tax rates not changed. We show, in a simple example, why the commonly used linear projection approach yields biased results in this setting, and show what can be estimated causally. In particular, we estimate the causal medium-run effect of the first tax policy change observed using a method developed by de Chaisemartin and D'Haultfoeuille (2022). In this limited setting, we again find no statistically significant effect of personal income, corporate income, and consumption tax rate changes five years later.

3.2 Literature Review

We argue that there are three crucial empirical challenges to providing credible estimates of how tax systems affect the level and growth of national output. The first is that taxes change for myriad reasons, some of which are correlated with ongoing developments in the economy and possibly directly related to recent or expected future output changes. For example, legislators may cut tax rates in anticipation of a recession. Second, it is difficult to construct a comprehensive measure of tax policy, which comprises many levies, sometimes featuring a graduated rate structure, imposed on several different tax bases, each of which generally changes over time. Lastly, it is difficult to obtain credible estimates of causal treatment effects using panel data when policies change in different periods and can change multiple times. In this section, we review the approaches that have been employed to address these challenges.

Barro (1991) pioneered the modern empirical analysis of the impact of taxes on economic growth. To address the endogeneity of tax policy changes, early papers use a panel of cross-country data that is typically collapsed to a single cross-section or shorter panel of differenced data of countries using either a long difference in economic growth (e.g., Barro, 1991) or the average growth rate over a long period (e.g., Lee and Gordon, 2005) as the dependent variable. This method implicitly assumes that any short-run endogeneity of tax changes with respect to economic conditions cancels out over the longer horizon, and the key identifying assumption is that conditional on the set of included controls, there must be no omitted factors that affect both tax rates and potential growth rates. Although these linear regression models offer a straightforward interpretation, the identifying assumption is difficult to satisfy. As a result, the estimated effects of taxes using this approach can vary widely depending on the set of included controls, as shown by, e.g., Easterly and Rebelo (1993).[1]

To better address the endogeneity of tax policy changes, Blanchard and Perotti (2002) championed the use of structural vector auto-regressions (SVARs), which hinge on assumptions about the exogeneity and timing of tax policy changes.[2] Tax effects are identified by imposing short-run restrictions on the system: output can contemporaneously respond to changes in tax policy, but tax policy can only respond to economic conditions with a lag. The justifications for this assumption are that legislative processes are lengthy, and tax policy changes are rarely motivated by output stabilization. This assumption is violated, however, if tax policies are anticipated or forward-looking, or also if governments quickly respond to changes to economic conditions. Because panel VARs often suffer from the curse of dimensionality—when the number of estimated parameters is large relative to the number of observations in the data—SVAR analyses typically focus on a single country.[3]

[1] Note that fixed effects with lagged regressors are biased in samples with short time periods (Arellano and Bond, 1991).

[2] Blanchard and Perotti (2002) estimate the impact of government spending and tax shocks on growth in the United States between 1947 and 1997.

[3] In her review of the literature, Ramey (2019) finds that SVAR analyses of tax increases tend to find

To obtain estimates using cross-country data, an influential paper by Arnold et al. (2011) instead uses a pooled mean group (PMG) estimator. The PMG approach estimates country-specific autoregressive distributed lag models for individual variables, and for tractability the authors impose homogeneity in the long-run relationship between taxes and growth across countries. Their headline result—that shifts from direct to indirect taxes are associated with higher levels of GDP growth—has influenced the "conventional wisdom" that has shaped policy recommendations by the IMF and OECD.[4] However, later work has shown that these results are fragile, and that the rank ordering of taxes varies using different assumptions about the long-run and short-run parameters (Xing, 2012), clustering standard errors at the country level (Baiardi et al., 2019), or expanding the set of countries and years included in the analysis (Widmalm, 2001; Xing, 2012; Angelopoulos et al., 2007; Baiardi et al., 2019).[5] Generally, all of the methodologies described thus far rely on modeling assumptions to deal with the endogeneity of tax policy changes.

Recently, the literature has converged towards the narrative approach introduced by Romer and Romer (2010) as the preferred method for dealing with the fact that some tax policy changes are directly (or indirectly) related to changes, actual or expected, in economic output. The narrative approach uses analysis of government documents and official speeches to classify which tax policies are arguably exogenous with respect to economic growth in the short to medium run. The exogenous policy shocks are then used either directly to estimate the causal effect of taxes on growth, or as instruments for the full series of policy shocks (e.g., Barro and Redlick, 2011; Mertens and Montiel Olea, 2018).[6] The econometric models used

large, contractionary effects—tax multipliers range from -1.1 to -5, averaging about a 2.75% contraction in GDP after 5 years following a 1% increase in revenue as a share of GDP.

[4]The IMF and OECD, in numerous publications, have suggested a "growth ranking" of tax instruments similar to the Arnold et al. (2011) paper, with taxes on consumption being the least harmful and corporate income taxation being the most harmful to economic growth. For example, in their October 2013 Fiscal Monitor publication, IMF authors note that "[t]he literature suggests that corporate income taxes have the most negative effect, followed by labor income taxes, then consumption taxes, and finally property taxes" while citing Arnold et al. (2011). The OECD report "All on Board: Making Inclusive Growth Happen" cites Arnold et al. (2011) in warning that raising progressive labor income taxes could be detrimental to growth in the long run.

[5]Xing (2012) studies 17 OECD countries for the period 1970-2004 and finds "no clear evidence that corporate income taxes are 'worse' than personal income taxes" (p. 381). Baiardi et al. (2019) extend the data to 34 OECD countries (and due to data limitations restrict the time period to 1995-2014), and find that a revenue-neutral shift to corporate taxation is significantly positively related to GDP per capita in the long run. An exception is Acosta-Ormaechea et al. (2019), which finds results consistent with Arnold et al. (2011) using a large sample of countries (69, including non-OECD countries) and years (through 2009). Note also that the user-written pooled mean group estimator Stata command employed by Arnold et al. (2011) has undergone major updates across versions. Our personal correspondence with the program author reveals that the standard error calculation was corrected over time, and our replication exercise showed that standard errors grew with the program updates. More detail is available upon request.

[6]While Mertens and Montiel Olea (2018) find a negative effect of increases in the average marginal tax rate (AMTR) on output, the relationship between changes in federal revenues and output conditional on

to capture the dynamic responses to policy changes are either an SVAR or the linear local projection (LP) framework proposed by Jordà (2005) and Stock and Watson (2007).[7] These studies initially addressed the U.S. in the post-WWII era, and have since been extended to several other countries.[8] Subsequent work has implemented the narrative approach in a cross-country setting (e.g., Alesina et al., 2015, 2017; Dabla-Norris and Lima, 2023).[9]

Estimates using the narrative approach for the U.S. find large, negative short-run effects of tax increases on economic output, with short-run tax multipliers for the U.S. ranging from 1.1 (Barro and Redlick, 2011) to 2.5-3 (Romer and Romer, 2010; Mertens and Ravn, 2014). Cross-country estimates for a sample of OECD countries also suggest that fiscal consolidations that rely primarily on tax increases are associated with deep and long-lasting recessions (Alesina et al., 2015, 2017). Importantly, these estimates are much larger in absolute value than those obtained using all, and not just arguably exogenous, tax policy innovations, suggesting there is considerable bias in non-narrative approaches.

Regardless of the econometric methodology used, this literature typically measures tax policy shocks as changes to average tax rates, frequently computed as projected revenue implications relative to baseline GDP.[10] This is problematic because, measured in this way, tax policy changes are by construction endogenous to the predicted economic growth rate itself. Even under so-called static scoring methods, official revenue estimating techniques commonly take into account predicted behavioral responses to tax policy changes, leading to a simultaneity bias in regressing growth rates on projected revenues. Using revenue estimates of narratively-defined exogenous policy shocks as an instrument for all tax policy changes will not correct for this source of endogeneity. An important exception is Mertens and

changes in the AMTR (which should capture wealth effects), are statistically insignificant. They conclude this provides suggestive evidence that the primary channel for tax effects on GDP is substitution, rather than wealth, effects.

[7]Impulse responses from local projections are approximately equivalent to impulse responses from VARs when each estimator uses the same number of lags, and the horizon of the impulse response is less than or equal to the number of lags used (Plagborg-Møller and Wolf, 2021).

[8]Mertens and Ravn (2012, 2013) further differentiate the Romer and Romer (2010) series by those policies that could be anticipated (i.e., when there are significant gaps between the enactment and implementation date of a policy) versus unanticipated, and by impacts of changes to average personal income and corporate taxes (Mertens and Ravn, 2013).

[9]These papers typically rely on a single-equation based on a linear local projection. Alesina et al. (2015, 2017) instead use a truncated moving average model, similar to Romer and Romer (2010), with country-specific fixed effects. Although the local projections method is able to estimate the impacts of fiscal shocks, it does not provide an estimate of the average impacts of fiscal plans due to the autocorrelation between fiscal shocks.

[10]There are exceptions. For example, Mendoza et al. (1997) use the ratio of source-based tax revenue to the tax base (e.g., labor income tax revenue to pre-tax household income), representing average tax rates on factor incomes and consumption, and Lee and Gordon (2005) use measures of top statutory marginal tax rates on different sources from the World Tax Database constructed by the Office of Tax Policy Research at the University of Michigan.

Montiel Olea (2018), who instrument for the observed change in average marginal tax rates (AMTRs) with the predicted change in AMTRs based on income of a base year, akin to instruments used in the public finance literature to estimate the elasticity of taxable income (e.g., Gruber and Saez, 2002).

By collapsing the multitude of tax policy levers into a single, summary amount, this tax policy shock measure also does not account for the multidimensionality of tax systems, instead combining the effects of changes to tax rates and the definition of the tax base, the effects of taxes applied to labor, capital and consumption, and so on. Yet, there is little reason to presume that effects are uniform across a diverse set of tax instruments. The few papers that attempt to disentangle these responses support that the effects of different aspects of the tax system are non-uniform, finding that substitution effects, rather than income effects (e.g., Barro and Redlick, 2011; Mertens and Montiel Olea, 2018), and changes to tax rates, rather than the definition of the tax base (Dabla-Norris and Lima, 2023), are the important drivers of tax effects. Following the insights from Kawano and Slemrod (2016) that there is significant correlation between changes to both rates and tax base definitions, Dabla-Norris and Lima (2023) control for changes to the tax base definitions when estimating the impact of changes to tax rates, and vice versa.

This review of the literature makes clear that the empirical analyses of the last few decades start from approximately the same underlying data but differ on a long list of features: (1) country coverage; (2) period coverage; (3) outcome variables; (4) tax system measures; (5) empirical methodology, including the strategy for dealing with endogeneity; and (6) horizon of effect considered. Our goal in this paper is to systematically and transparently document the implications of alternative strategies for addressing the empirical issues that arise, and evaluate what causal relationships between aspects of tax systems and medium-term economic performance can be credibly claimed. We do this utilizing comprehensive new cross-country data sets now available that measure key aspects of tax systems and apply the narrative approach to classify exogenous tax changes with respect to economic growth.

In addition to examining the strategies that have been extensively used, we address a burgeoning econometrics literature that shows the potential for significant bias in these methods when applied to panel settings with treatments staggered over time (e.g., Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeuille, 2022). We employ a new econometric estimator that can identify a particular causal average treatment effect when countries experience multiple changes in tax policy developed by de Chaisemartin and D'Haultfoeuille (2022).

3.3 Data Sources and Descriptive Statistics

We draw on several sources to construct a data set that describes changes in the top rate and aspects of the tax base for the personal income tax (PIT), corporate income tax (CIT), and value added tax (VAT) for 23 countries over 34 years.[11] In this section, we describe these data sources in detail.

3.3.1 Statutory Tax Rate Data

We construct cross-country panel data on statutory tax rate changes using various data sources. Our primary source is the OECD Tax Database, which provides data on tax rates for 34 OECD countries from 1981 through 2015. We supplement these data with information on statutory tax schedules from the American Enterprise Institute (AEI) International Tax Database. This database covers 159 countries from 1981 through 2011. We identify the top statutory marginal CIT and PIT rates, and the standard VAT and sales tax rates for each country in each year. The CIT rate refers to the statutory rate that applies generally in cases where there are multiple tax rates for different sectors, and to the rate that applies to publicly-traded companies if a different rate applies to privately-held businesses.

Occasionally, the OECD Tax Database and AEI International Tax Database disagree on top statutory rates. In such cases, we turn to other data series, such as government finance web pages or other academic articles, the University of Michigan's World Tax Database, and information available from the Tax Policy Center. Appendix C.1 provides detailed documentation on how we construct these series.

Because our tax rate series capture changes in tax rates at the time of implementation, we use information from the IMF's Tax Policy Reform Database (TPRD) to construct a series based on the time of announcement to account for anticipation of tax rate changes. The TPRD covers 23 advanced and emerging market economies[12] since the early 1970s, and reports the direction of tax rate changes, along with information on whether a policy affected the top statutory rate or other portions of the tax schedule and both the announcement and implementation dates of the policy.

[11] We define the top rate for the VAT as the standard VAT rate, ignoring potentially higher rates on luxury goods. Our focus is on top tax rates, as they are arguably the most salient features of tax systems. Very few businesses (as a share of GDP) face marginal CIT rates other than the top rate, with many CIT systems having a flat rate. The top PIT tax rate is often the "headline" rate for personal income taxes, and applies to a larger share of the income distribution particularly in Europe.

[12] The 23 countries are Australia, Austria, Brazil, Canada, China, the Czech Republic, Denmark, France, Greece, Germany, India, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Poland, Portugal, Spain, Turkey, the United Kingdom, and the United States.

3.3.2 Statutory Tax Base Data

Data on changes to the definition of tax bases come from the TPRD. In the spirit of Kawano and Slemrod (2016), it reports the direction of change in the base (and rate) of six tax types: personal income tax, corporate income tax, value-added tax, excise taxes, social security contributions, and property taxes. In this paper, we focus on the first three of these taxes. Constructing this database entailed processing information from more than 950 OECD country reports and 53,000 International Bureau of Fiscal Documentation (IBFD) news clips using text-mining techniques to extract information that potentially describes tax policies, and identified which proposed policies were actually implemented. For further details on the database, see Amaglobeli (2018).[13]

For the CIT, the base measures recorded refer to R&D promotion (e.g., tax credits), investment promotion (e.g., depreciation rules), loss carry-forward and loss carry-back rules, thin capitalization rules, capital gains taxation, and all other changes to the CIT base.[14] For the PIT, the base measures recorded refer to standard relief (e.g., single or family deductions or tax credits), child relief, capital gains relief, interest relief, relief for social security contributions, insurance premiums, and private pensions, and all other changes to the PIT base. Changes to the VAT base are recorded as either exemptions on food items, exemptions on medical supplies, exemptions on education, and all other VAT base changes. For most specific measures, the database documents the announcement and implementation dates (e.g., day or month and year). For each measure, a variable indicates whether the change entails an increase or a decrease in the breadth of the tax base. A measure is coded to increase (decrease) the tax base if it will contemporaneously increase (decrease) tax revenues holding constant other aspects of the tax system, including tax rates, as well as behavioral responses. There are also indicator variables that denote whether the measure represents a "major" base change[15], whether it was part of a policy package, and whether it is multi-year in nature.

Each observation in the TPRD provides information on a specific type of tax rate or base change. If there are several tax policy changes that occur in the same calendar year or tax reform packages that affect several aspects of the tax system, each measure is documented separately. We collapse these observations at the country-year level to construct a panel

[13] Our analyses are based on the TPRD versions 3.0-4.0. The database is a work in progress, with additional information over countries and time periods covered expected over time. The earliest announcement and implementation year is 1930; the latest announcement year is 2018, and the latest implementation year is 2020. The TPRD reports that its coverage is most comprehensive between 1988 and 2018.

[14] The measures that capture CIT base changes in the TPRD differ somewhat from those that are contained in Kawano and Slemrod (2016).

[15] Changes to the tax base are considered major changes when the language describing the policy change is deemed to point to a large change (e.g., changes "reported to affect large groups of taxpayers or potentially mobilize significant resources").

data set. For each tax base measure, we tabulate the number of times that there was an "increase" or "decrease" in that tax base measure. The TPRD describes whether a specific type of policy narrowed or broadened a particular tax base, and does not permit comparisons of the relative magnitudes of its effect. To avoid overweighting country-years that contain several small changes to a particular aspect to a base, we also create indicator variables for having any policy that "increases" or "decreases" each aspect of a tax base. In a small number of cases, there are counteracting policies, such as when a country enacted a policy that made investment incentives more generous while in the same year enacting another policy that reduced investment incentives. Changes that both broaden and narrow a particular aspect of a tax base occur in 10.5% of CIT base changes and 11.9% of PIT changes, almost exclusively in changes to standard relief and other PIT base changes. These conflicting changes to the VAT base occur only once, for exemptions on food items. In these cases, the indicator variables for having any "increases" or "decreases" to those tax base measures are both set equal to one. We use the implementation year to collapse the data.

For the CIT and PIT systems, we consider any country-year in which there is no information on a particular aspect of the tax system in the TPRD as one in which there is "no change" to that aspect of the system. We assume that if tax rate data are available for a country-year in the OECD Tax Database or AEI International Tax Database, then information about the tax system would have been captured in the OECD Economic Surveys and IBFD news archives that form the basis of the TPRD. We use this information to fill in country-years when there was "no change" to a country's tax bases. We make this choice so that we do not conflate years with no tax policy changes with those in which there is simply no information available in the TPRD's underlying data sources. We similarly fill in country-year observations with "no change" to the VAT system beginning with the first observed TPRD entry regarding the VAT system. We also limit our sample to 1982-2015, the years with information on the change in tax rates. Appendix Table C.1 provide information on the countries included in the TPRD and the first observed CIT, PIT and VAT system change.

While it is important to control for changes to the definition of the tax base in our analyses, a limitation of our work is that these measures only capture whether the tax base was broadened or narrowed. These measures do not capture the magnitude of these policy changes, and so we are unable to fully account for the impact of tax base changes when estimating the effects of tax rate changes.

3.3.3 Exogenous and Endogenous Tax Policy Shocks

We utilize two sources that classify tax changes as exogenous or endogenous using a narrative approach, in addition to our own classification exercise. The first is a publicly-available database of fiscal consolidations from Alesina et al. (2017). This database provides annual data on exogenous tax revenue changes for 17 countries between 1981 and 2014. The policy impacts are disaggregated into changes in corporate income, personal income, and consumption tax revenues.

The second data source we use is from Dabla-Norris and Lima (2023)[16], which further disaggregates the fiscal consolidations identified by Alesina et al. (2015) into changes to tax rates and the definition of tax bases, and narratively identifies the exogeneity of each individual policy. These more granular data are available for 10 OECD countries. Tax rate changes include any change to the main statutory tax rate(s), and tax base changes are identified though any change in the legal definition of the tax base to which statutory rates are applied, or through changes in credits or exemptions that change tax liabilities. Tax policy changes are separated by their source (e.g., corporate, personal or consumption tax base).

We also build our own data set of exogenous tax rate changes, starting with the exogenous rate changes identified during episodes of fiscal consolidation in Dabla-Norris and Lima (2023). Our data collection effort uncovered 104 additional tax rate changes occurring outside consolidation episodes for their same sample of 10 OECD countries. We cross-reference these changes with the TPRD and national sources to ensure their validity, finding the announcement dates for each. Along with budget announcements recorded in the TPRD, we use supporting information from the IBFD, which includes news articles surrounding the announcement and implementation of these tax changes. This is combined with data on the state of the economy at the time of announcement and the ensuing years to judge the extent to which these tax changes may be motivated by the business cycle. We use these sources and narratively determine that 32 of these rate changes are exogenous (described below).

Appendix Table C.2 summarizes the key differences among these three data sets regarding the countries and year coverage as well as the definition of an "exogenous" tax change. For comparison, we also include those for the seminal paper by Romer and Romer (2010). All series treat tax measures aimed at reducing the fiscal deficit as exogenous. The other exogenous policy shocks in these sources are those that do "not appear related to other factors affecting output in the near future" (Romer and Romer, 2010, p.770), are motivated by reasons independent of the state of the business cycle (Alesina et al., 2015), or were

[16]We are grateful to Era Dabla-Norris and Frederico Lima for sharing their database with us.

primarily aimed at increasing long-run growth (Dabla-Norris and Lima, 2023). Endogenous tax changes are those that fail to meet any one of these conditions.

For each tax rate change, the associated change in revenue as a percent of GDP is often used to estimate the impact of tax changes on economic growth. For comparison, we also measure policy shocks as revenue-to-GDP for our series of exogenous tax policy shocks. One data source is from Dabla-Norris and Lima (2023). Here, the expected revenue impact of each tax change is compiled from country authorities at the time of policy change announcements. Alesina et al. (2017) do the same. In each paper, changes in revenues are expressed as a percent of annual GDP from the quarter preceding the announcement. For our own set of exogenous tax policy shocks, we use estimates from Dabla-Norris and Lima (2023) and Alesina et al. (2017) where applicable. If unavailable, we use the change in the tax-revenue-to-GDP ratio from the quarter prior to the policy announcement to a year after as the resulting revenue change. We impute values in just under 30% of our sample. This imputation could be a source of bias in our analyses that use tax revenue measures, as the actual change in revenues is the sum of projected revenue changes plus the error term. However, as we describe below, our preferred measure of tax policy changes are statutory tax rate changes that are not subject to this concern.

We formally test the exogeneity of our narratively-defined policy shock series with Granger causality tests. We estimate the following specification:

$$\text{Tax Rate Shock}_{c,t} = \alpha + \sum_{i=1}^{T} \beta_c x_{t-i} + \sum_{i=1}^{T} \gamma_c \text{Tax Rate Shock}_{c,t-i} + \delta_c + \delta_t + \epsilon_{c,t}$$

where our tax rate shocks are regressed on T lags of potentially predictive variables x, controlling for T lags of exogenous tax changes, country fixed effects δ_c for idiosyncratic country characteristics, and time fixed effects δ_t for global business cycles that may impact all countries in the sample. We use two years of lags ($T = 2$), and test each predictive variable one at a time. Appendix Table C.3 presents F-statistics for joint tests of the significance of the lagged predictor variables and their corresponding p-values for all tax rate changes together as well as by tax type. Panel A presents results for our set of exogenous shocks, which importantly shows that exogenous tax changes are not driven by the business cycle, and lagged changes in output are not predictive of any rate change measures. The only variable that has predictive power is government purchases for changes in the corporate income tax rate. Nevertheless, to prevent these factors from confounding our estimates, we include these variables as controls in our main specifications. Panels B and C repeat this exercise to show that the Dabla-Norris and Lima (2023) tax rate series are not predictable, but the Alesina et al. (2017) corporate tax rates are statistically significantly associated with lagged changes in output.

For a discussion of the issues around categorizing tax policy changes as exogenous or endogenous with respect to the state of the economy, a comparison of alternative data sets for doing so, and comparing the composition of all tax rate changes to exogenous rate changes, see Appendix C.

3.3.4 Summary Statistics

Table 3.1 provides summary statistics of our compiled tax data. Overall, our sample covers 350 country-year observations: 35 each for 10 countries. Tax rate changes deemed exogenous skew slightly more towards cuts than tax rate changes overall. There is a change to at least one of the corporate tax, individual tax or VAT rates in nearly one-fifth of country-year observations. Nearly half of the observations contains some change to an aspect of a tax base. Policy changes to the VAT system—either in the rate or in some aspect of the base—are the least commonly observed, occurring in fewer than 10 percent of observations. This relative infrequency of policy changes to the VAT system will matter for power in estimating the effects of VAT policy shocks.

In addition to the tax rates and base data, we consider several other macroeconomic variables. Measures of economic activity come from the World Bank's World Development Indicators (WDI) database. We collect the annual real GDP growth rate, real GDP (measured in local currency units), and GDP per capita (measured in constant 2010 US dollars). From the WDI, we also gather data on the national unemployment rate. Our measure of private sector investment comes from gross fixed capital formation as a share of GDP. Total tax revenue, measured as a share of GDP, comes from the OECD Global Revenue Statistics Database. Finally, data on short-and-long-term nominal interest rates, employment levels, inflation, and government expenditure come from the OECD Economic Outlook No. 102 (November 2017), with similar series for pre-unification Germany coming from the St. Louis Federal Reserve Bank's Federal Reserve Economic Data (FRED) database.

3.3.5 Raw correlation between changes in tax rates and GDP

Figure 3.1 presents a visual representation of the correlations between a one-year change in different tax rates and GDP per capita growth rates over the five-year period beginning with the year of announcement of the tax change. Panel A shows this relationship using only the set of exogenous tax rate changes, while Panel B includes all tax rate changes. The corresponding point estimates are presented below each figure. Here, and in all remaining

Table 3.1: Summary Statistics of Tax Data

Variable	Mean	Std. Dev.	Min	Max	Share of Country-Years
PIT Rate Change	-2.94	8.21	-28	17.87	20.00%
CIT Rate Change	-2.67	5.41	-25	6.2	23.71%
VAT Rate Change	0.84	3.59	-18.5	6.5	15.14%
Exogenous PIT Rate Change	-3.16	8.83	-28	10	14.00%
Exogenous CIT Rate Change	-3.33	6.11	-25	5	16.00%
Exogenous VAT Rate Change	0.6	4.44	-18.5	5	9.43%
PIT Base Change	-0.74	1.51	-4	4	46.57%
CIT Base Change	-0.48	1.32	-3	3	40.57%
VAT Base Change	0.07	1.07	-2	1	7.71%
Count	350				

Notes: Summary statistics of tax rate and base changes. Observations are at the country-year level and statistics are calculated from the share country-years that feature a tax change. Rate changes are in points, and base changes are indicator variables for broadening (+1) and narrowing (-1). The share of country-years refers to the percentage of observations which feature the indicated variable.

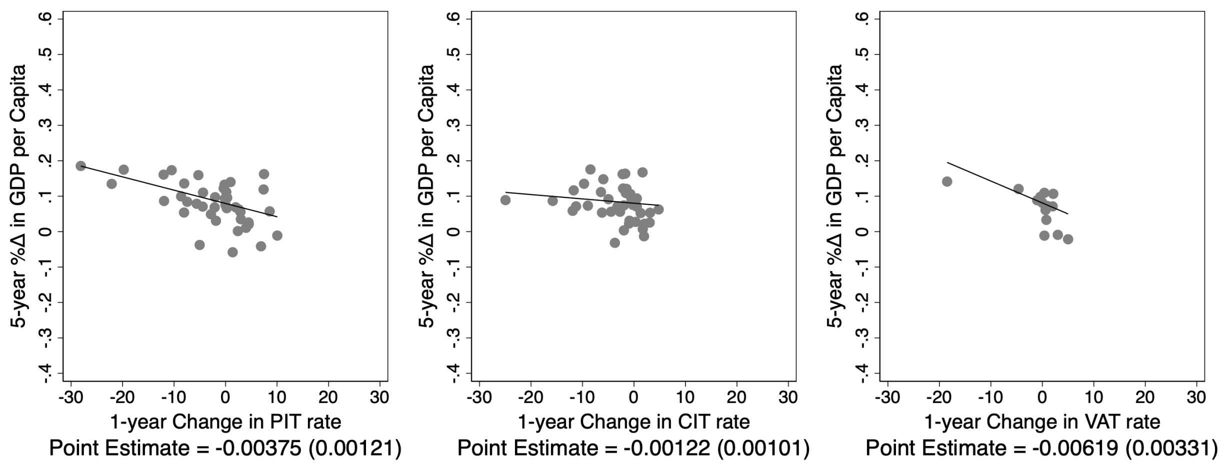
estimates, we present Driscoll-Kraay standard errors.[17] We choose a five-year horizon because we are interested in the medium-term effects of tax changes, and not the near-term cyclical effects. We consider these relationships for PIT, CIT and VAT rates separately. To isolate the effect of changes in tax rates, we control for changes in the tax base for each tax system by residualizing each variable on changes in the tax base before plotting.[18] The change in the tax base is computed as the sum of base broadening (+1) or base narrowing measures (-1) that occur across all aspects of the base that are captured in the Tax Policy Reform Database.

Several interesting patterns arise from these simple correlations. First, Panel A suggests that there is a negative relationship between an exogenous tax rate increase and the subse-

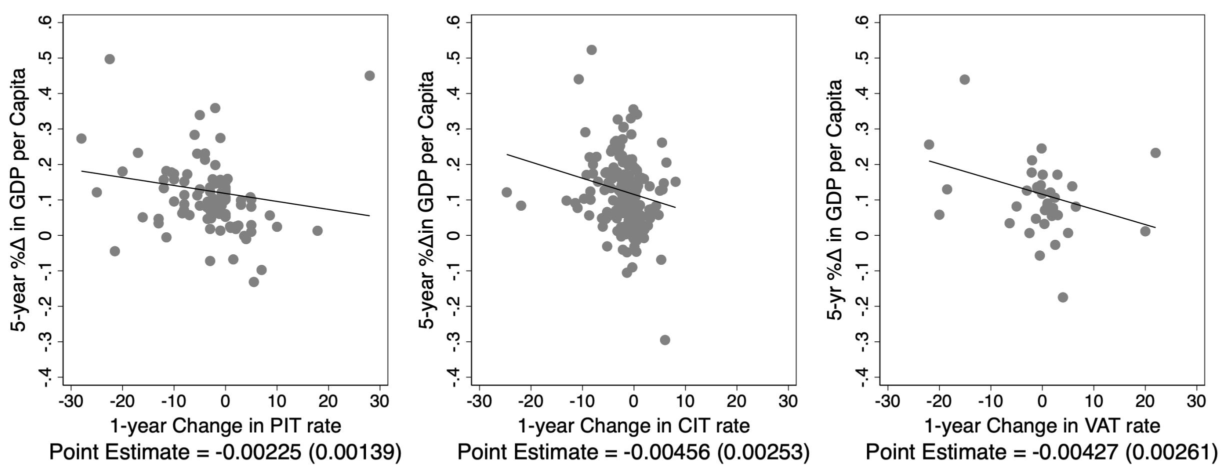
[17] Valid inference for local projection impulse responses can be obtained using HAC (heteroskedasticity and autocorrelation-consistent) standard errors (Jordà, 2005). Driscoll-Kraay standard errors are, in addition, robust to cross-sectional dependence—an important feature given that the countries in our sample are all in the OECD.

[18] To residualize, we regress changes in GDP growth on changes in tax bases. The residuals of this regression are changes in GDP growth unexplained by changes in tax bases. Plotting these against changes in tax rates gives the correlation between changes in tax rates and GDP growth unexplained by changes in tax bases.

Figure 3.1: Correlation of 1-Year Change in Tax Rates and Subsequent 5-Year Change in GDP per Capita Growth Rates, Controlling for Tax Base Changes



Panel A: Exogenous Rate Series



Panel B: All Rate Series

Notes: The correlations between one-year changes in tax rates and growth in GDP per capita over the subsequent five years. Panel A shows the correlations only for narratively-identified exogenous tax rate changes. Panel B shows the correlations for all tax rate changes. Each panel shows correlations separately for each tax type—the top marginal personal income tax rate (PIT), the top marginal corporate income tax rate (CIT), and the standard consumption tax rate (VAT). Changes in tax bases are controlled for by residualizing GDP per capita growth on indicators for base changes for each tax type. Point estimates are below each plot, with Driscoll-Kraay standard errors in parenthesis.

quent five-year economic growth rate for the PIT and VAT, but this is statistically significant at the 5% level for only the PIT. Second, the relationship between exogenous CIT rate increases and economic growth is flat and statistically indistinguishable from zero. Lastly, a comparison between Panels A and B suggests that restricting attention to exogenous tax rates is important for interpreting the causal relationship between tax policy shocks and economic growth, particularly for the CIT. Using the full set of CIT rate changes suggests that the negative relationship between rate increases and economic growth is almost four times as large as that when using exogenous changes alone, and this relationship for the CIT is statistically significant at the 10% level. For the PIT and VAT, the (negative) relationships between rate increases and economic growth become weaker when using the full series of tax rate changes, as does their statistical significance. These simple comparisons of unconditional means ignore other factors that may affect both tax policy and economic growth. In the analyses that follows, we examine how these relationships change as we move from such correlations to estimates that attempt to reveal the causal effects of tax policy.

3.4 What are the Implications of Various Modeling Choices?

In this section, we attempt to uncover the implications of various modeling choices for estimating the causal effects of tax policy shocks on economic growth. Guided by the existing body of research and what is available in the comprehensive data set that we compiled, we systematically show how the estimated relationships change along several dimensions highlighted in our literature review.

We pursue two estimation strategies that have been frequently used in the literature. We begin with a transparent approach that directly relates changes in tax policy to medium-term growth rates in a simple linear regression framework. We then turn to a location projection (LP) model more common in the macroeconomics literature that traces the path of responses over the same time horizon. For each strategy, we begin by presenting estimates using our preferred measure of tax policy shocks: narratively-determined exogenous changes to statutory tax rates. We compare these estimates to those obtained when making alternative choices. Because the set of country-year changes when moving across these different tax policy measures, we also assess whether the conclusions change if we hold the set of country-years constant across these choices. In all of our analyses, we focus on economic growth over a five-year time horizon.

3.4.1 The Linear Regression Method

We first estimate a simple linear regression model that relates the annual economic growth rate to contemporaneous and five lags of annual changes in a particular aspect of the tax system,[19] given by:

$$\ln Y_{c,t} - \ln Y_{c,t-1} = \alpha + \sum_{\underline{\ell}=0}^{\bar{\ell}=5} \beta_\ell \left(\tau_{t-\ell} - \tau_{t-\ell-1}\right) + \sum_{\underline{\ell}=0}^{\bar{\ell}=5} \gamma_\ell X_{t-\ell} + \delta_t + \delta_c + u_{c,t} \tag{3.1}$$

In this specification, $Y_{c,t}$ is real output per capita of country c in time period t, α is a constant, and $\tau_{c,t}$ is the tax rate of interest. The vector δ_t contains year fixed effects that control for global economic conditions that may affect both tax policy and economic growth rates; country-specific fixed effects are controlled with δ_c. Because there can be significant correlation between changes in tax rates and tax base definitions, failing to account for the simultaneity in these policy shocks can lead to estimation bias (e.g., Kawano and Slemrod, 2016). Thus, we control for changes in the tax base as well as changes to the other tax rates with X_t. Given the potential predictive power of other economic variables as discussed in Section 3.3.5, we control for lags of GDP growth, debt levels, government spending growth, inflation, short-run interest rates, and tax revenue growth, as in Dabla-Norris and Lima (2023).

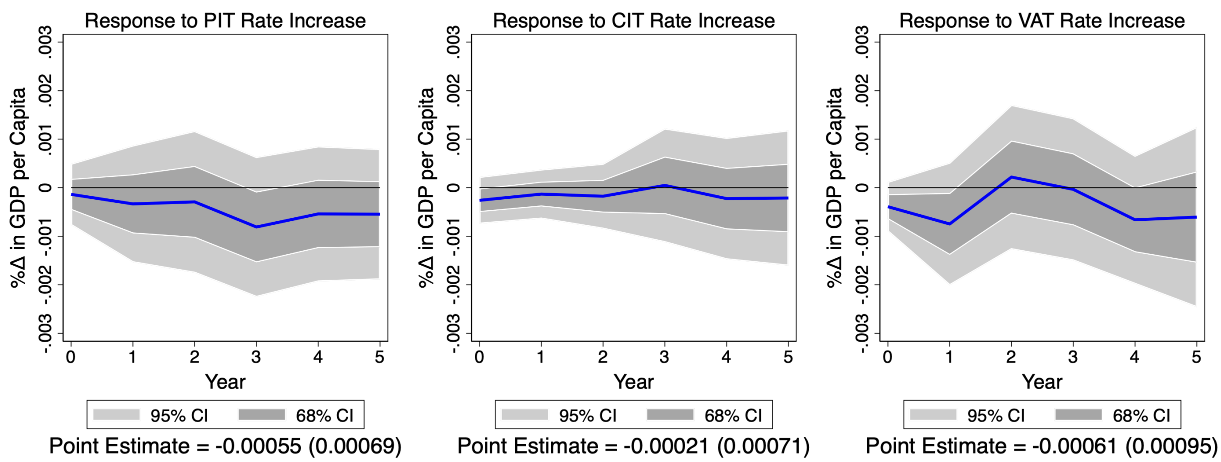
This specification allows us to trace out the relationship between a tax rate change and economic growth rate over the subsequent five years. The effect of a tax rate change in its contemporaneous year is captured by β_0, the effect in its first year is captured by β_1, and so on through its effect in the fifth year as estimated by β_5. To obtain an estimate of the five-year effect of a tax policy change, we calculate the cumulative sum of the six estimated βs.

In Figure **??**, we show how these tax policy effects on GDP growth accumulate over the five-year window along with 68% and 95% confidence intervals for a 1-point increase in each tax rate. We focus in Panel A on the set of exogenous tax changes identified using the narrative approach and, for comparison, use the full set of tax rate changes in Panel B. The horizontal axis corresponds to the years elapsed since a tax policy announcement occurred and the plotted estimates are the corresponding cumulative sum of β_0 through β_5 over the time horizon. The final 5-year point estimate and Driscoll-Kraay standard error is given below each plot.

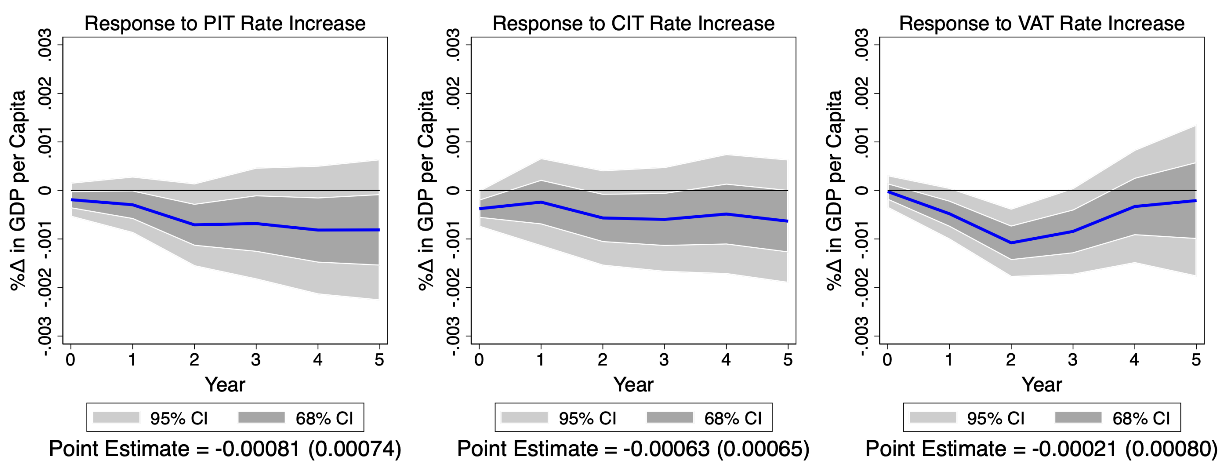
When considering only exogenous rate changes in Panel A, responses are attenuated compared to regressing on all rate changes—exogenous and not—in Panel B. This is particularly true for the VAT rate, which is correlated with statistically significant declines in GDP per capita 2 years after a tax change in Panel B, but not at all in Panel A. Relative to the simple correlations, the general patterns remain unchanged: there is little evidence to support a

[19] This specification is based on equation (35) in Suárez Serrato and Zidar (2016).

Figure 3.2: Cumulative Five-Year Effect of Tax Rate Changes



Panel A: Exogenous Rate Series



Panel B: All Rate Series

Notes: The cumulative effects of a one-point increase in tax rates on GDP per capita over the subsequent five years from linear regressions. Panel A shows the effect only for narratively-identified exogenous tax rate changes. Panel B shows the effect for all tax rate changes. Each panel shows effects separately for each tax type—the top marginal personal income tax rate (PIT), the top marginal corporate income tax rate (CIT), and the standard consumption tax rate (VAT). Darker and lighter shading indicate 68% and 95% confidence intervals, respectively. Point estimates for the fifth year are below each plot, with Driscoll-Kraay standard errors in parenthesis.

conclusion that changes to different tax rates have statistically significant medium-run effects on economic growth rates. In Appendix Figure C.4 and Panel A of Appendix Table C.5, we show this is also true for the series of exogenous tax rate changes from Dabla-Norris and Lima (2023) and Alesina et al. (2017).

3.4.2 The Local Projection Method

We now turn to a local projection (LP) model, which is a more common estimation strategy in the macroeconomics literature for unpacking the dynamic response of economic growth to a policy shock. We trace the path of responses over the same time horizon as in our linear regression framework. Specifically, at each time horizon, $h = 0$ through $h = 5$, we estimate the regression:

$$\begin{aligned} \ln Y_{c,t+h} - \ln Y_{c,t-1} = \alpha + \beta_0 \left(\tau_{c,t} - \tau_{c,t-1}\right) + \beta_1 \left(\tau_{c,t-1} - \tau_{c,t-2}\right) + \cdots \\ + \beta_5 \left(\tau_{c,t-5} - \tau_{c,t-6}\right) + \delta_t + \delta_c + u_{c,t+h} \end{aligned} \tag{3.2}$$

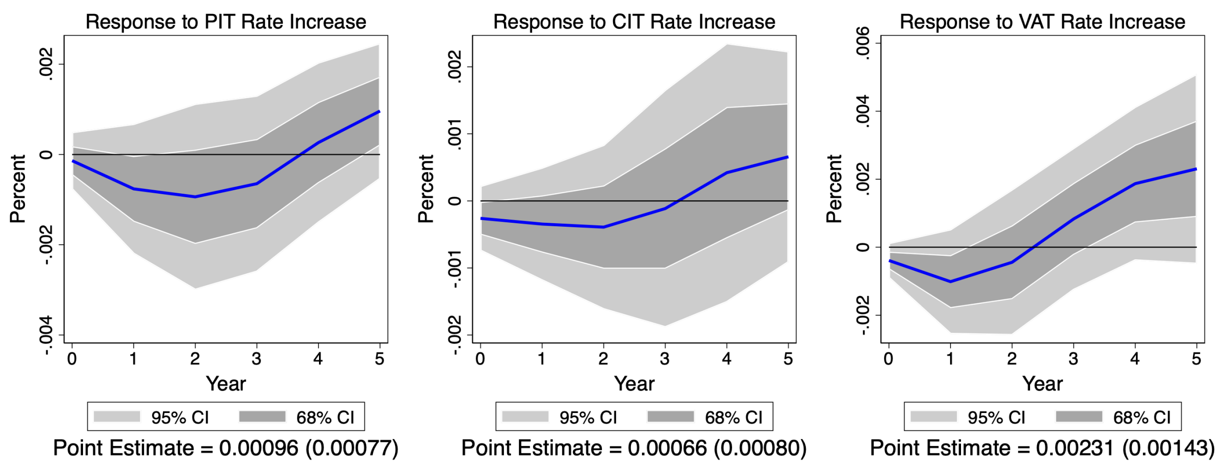
The impulse response of the log change in output to a tax policy shock over the 5-year window is traced out by the six estimates of β_0 resulting from these regressions. For the regression corresponding to $h = 0$, the estimate of β_0 captures the contemporaneous log change in GDP to a tax policy change that occurs between $t = -1$ and $t = 0$. For higher values of h, the estimates of β_0 capture the impacts of that same tax policy shock over longer time horizons, controlling for past tax changes. When $h = 5$, the parameter estimate reflects the cumulative change in log GDP to a policy shock that occurred five years prior. As before, we control for lagged changes to tax bases, tax rates other than the one of interest, GDP growth, debt levels, government spending growth, inflation, short-run interest rates, and tax revenue growth.

The impulse response function from the LP model is similar in spirit to the cumulative sum approach of (3.1), but it aggregates the effects of tax changes in a slightly different way. At longer time horizons, the LP model controls for additional lags of tax policy shocks and other control variables. In contrast, the cumulative sum of estimated coefficients from the linear regression model does not control for these longer lags.[20] Because of this difference in lag structure, the LP model may be preferred over the cumulative sum approach, as including more lags allows for the possibility that the effects of policy shocks take considerable time to materialize (Kilian and Lütkepohl, 2017).

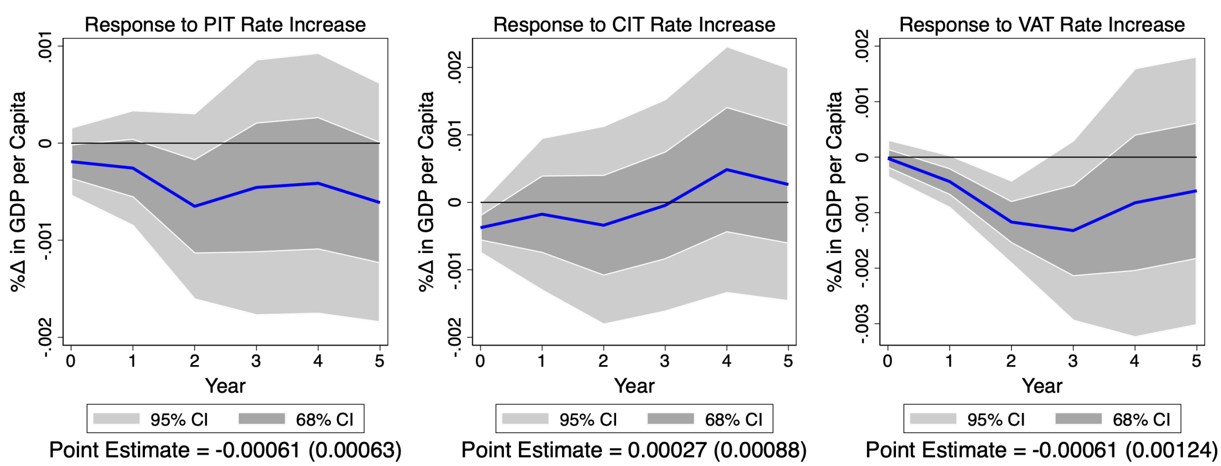
Figure 3.3 depicts the results from estimating (3.2). The horizontal axis denotes the years elapsed since a tax rate change was announced, and each point estimate corresponds to an estimated β_0 from a regression for a time horizon $h = 0$ through $h = 5$. The figures show the response to an exogenous 1-point increase in each tax rate, and the 5-year point estimate are presented below each plot. Overall, the results are quite similar to those in Figure **??**, but now the downward bias of using the full set of tax policy shocks (Panel B)

[20] For example, 5 years out, the point estimate for the local projections is influenced by five years of previous tax changes as controls. The cumulative sum at the same horizon does not have these controls.

Figure 3.3: Five-Year Effects of Tax Rate Changes



Panel A: Exogenous Rate Series



Panel B: All Rate Series

Notes: The effects of a one-point increase in tax rates on GDP per capita over the subsequent five years from local projections. Panel A shows the results only for narratively-identified exogenous tax rate changes. Panel B shows the results for all tax rate changes. Each panel shows results separately for each tax type—the top marginal personal income tax rate (PIT), the top marginal corporate income tax rate (CIT), and the standard consumption tax rate (VAT). Darker and lighter shading indicate 68% and 95% confidence intervals, respectively. Point estimates for the fifth year are below each plot, with Driscoll-Kraay standard errors in parenthesis.

rather than exogenous shocks alone (Panel A) is even more pronounced. Again, none of the exogenous tax rate changes have a statistically significant impact on GDP per capita growth at the 95% level, but regressions using the full tax change series would imply some significant short-term negative effects on growth. Appendix Figure C.5 and Panel B of Appendix Table

C.5 presents similar estimates that use alternative exogenous tax change series.

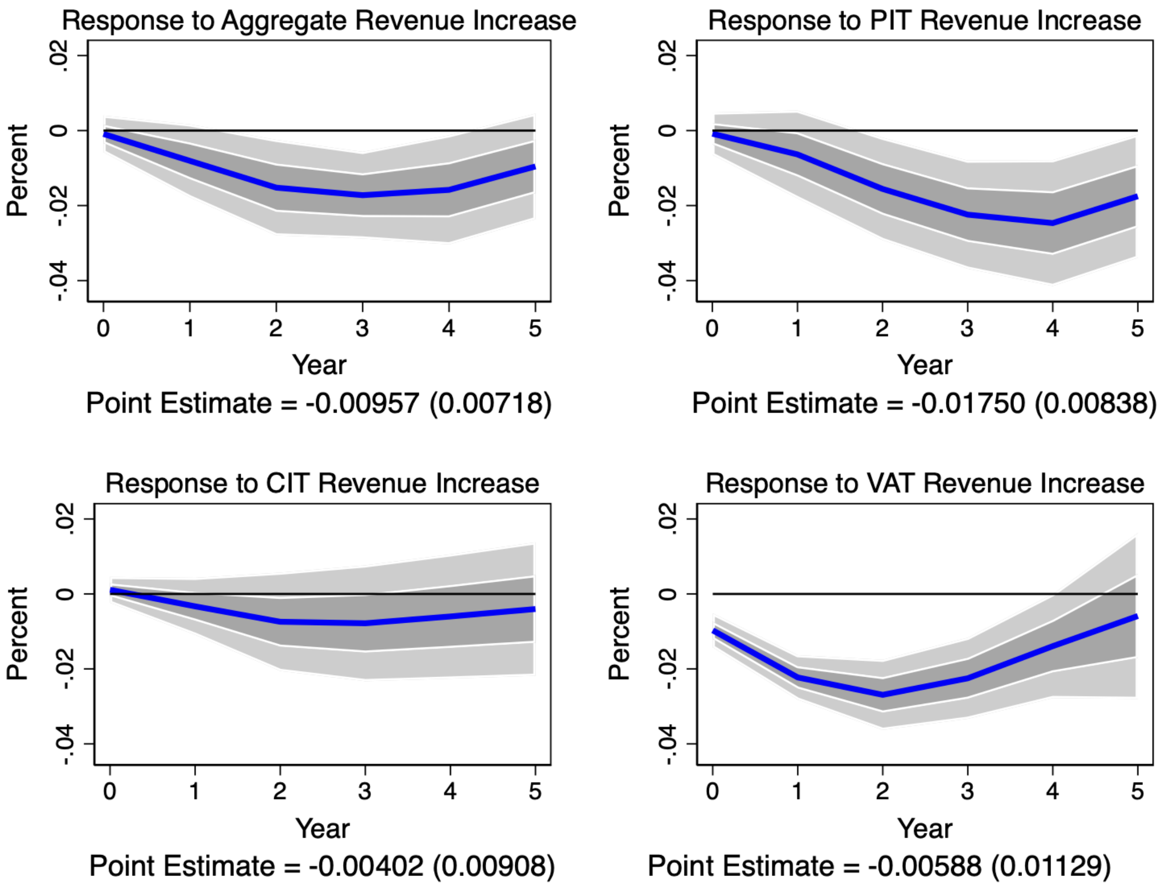
Because some of the previous macroeconomics literature measures tax policy shocks as changes in tax revenues, for comparison Figure 3.4 presents estimates using exogenous changes in tax revenues as a percent of GDP as our tax policy shock measure. Now the shape of the relationships between tax rate changes and economic growth are dramatically different. An increase in the PIT rate has a statistically significant negative impact on GDP growth rates beginning 2 years after announcement and continues to persists until 5 years later. VAT rate increases have an even larger negative immediate effect on economic growth, but the five-year effect is not statistically significant. The CIT rate is also associated with more negative economic growth rates, but these point estimates are not statistically significant. Thus, the choice between measuring tax policy as a change in statutory rates or the expected change in tax revenues matters quite a bit for the implied effects of taxes on growth. In our setting, using exogenous revenues as a percent of GDP as the measure biases the estimated effect of tax changes downwards, particularly in the two years following the tax change. This may be due to simultaneity bias, as changes in tax revenues as a percent of GDP are by construction endogenous to the economic growth rate itself, and official revenue estimating techniques often already account for some behavioral responses to tax changes.[21]

We examine the implications of collapsing PIT, CIT and VAT changes into a single summary measure on estimates in a linear projection model in Figure 3.4. Aggregating tax changes that occur across the three tax bases considered obscures the heterogeneity in effects that might exist. For example, it appears that increases in VAT revenues are more contractionary soon after tax increases, compared to other revenue sources, although the effect is not statistically significant after 5 years.

What have we learned so far? First, consistent with previous work (e.g., Romer and Romer, 2010; Alesina et al., 2015), including endogenous tax policy changes biases the estimated effect of tax rate increases on economic growth in the medium run—in our setting, including endogenous tax rate changes overstates the negative impact on GDP per capita. Second, the estimated impact of tax on growth depends on how the tax shocks are measured. Using revenue changes as a percent of GDP leads to more contractionary estimated tax effects than our preferred statutory tax rate measure. Lastly, we confirm the findings in Dabla-Norris and Lima (2023) that collapsing tax rate changes to a single measure conceals heterogeneity in the impacts of different tax types. Under what might be considered a preferred set of choices to this point—the linear projection model using statutory tax rate

[21] As noted in Section **??**, some address this issue by instrumenting for the exogenous revenue change as a percent of GDP using the statutory tax rate change. While this can successfully correct for the endogeneity of tax revenues, the potential for significant bias when applied to panel settings with staggered treatments remains.

Figure 3.4: Five-Year Effects of Exogenous Tax Revenue Changes



Notes: The effects of a one-point increase in tax revenue as a percent of GDP on GDP per capita over the subsequent five years from local projections. The results are only for narratively-identified exogenous tax revenue changes. The first panel shows the results for all tax types combined (aggregate). The following panels show results separately for each tax type—the personal income tax (PIT), the corporate income tax (CIT), and consumption taxes (VAT). Darker and lighter shading indicate 68% and 95% confidence intervals, respectively. Point estimates for the fifth year are below each plot, with Driscoll-Kraay standard errors in parenthesis.

changes—there is suggestive evidence of increased economic growth five years after tax rate increases, regardless of the tax base we consider, although none of the estimates here is statistically significant at conventional levels.

There is, though, a reason to be concerned that all of the results presented thus far may be biased. In our setting, countries experience tax policies that vary both in their timing and in the direction and intensity of their associated changes. While the approaches in this section are commonly used to exploit variation from these types of staggered treatments,[22] a burgeoning microeconometrics literature shows that they can produce biased estimates if treatment effects are heterogenous across countries and over time (e.g., Sun and Abraham, 2021; Goodman-Bacon, 2021; de Chaisemartin and D'Haultfoeuille, 2020). In fact, estimators can fail to satisfy the "no-sign reversal" property. When this property fails to hold, we can observe, for example, a positive average treatment effect on the treated (ATT) when in fact some jurisdictions experience negative treatment effects (de Chaisemartin and D'Haultfoeuille, 2022). We address this issue in the next section.

3.5 Estimation Accounting for Staggered Treatments

Any research design that uses panel data to estimate tax effects will face challenges in obtaining credible causal estimates because it relies on policy variation that occurs across jurisdictions at different times and with different treatment intensities. With variations in treatment timing, an issue arises when using a comparison group that comprises already-treated units. Countries change their tax policies frequently and, as a result, the set of country-years that can serve as part of a valid counterfactual for a given tax shock diminishes quickly. Yet the problem of making "forbidden comparisons" is prevalent in the literature that uses staggered policy variation to estimate treatment effects with two-way fixed effects models (Goodman-Bacon, 2021), and it is now well understood that the estimation methods in this literature, and in Section 4, can thereby produce biased treatment effect estimates (see Propositions 2 and 3 of de Chaisemartin and D'Haultfoeuille (2022) for discussion of the bias in the distributed lag regression and linear projections model, respectively).

To understand what induces the bias, we first examine a simple example to illustrate why the local projections estimates are problematic. We analyze a setting with two countries and three time periods, and analyze the effect of being exposed to one period of treatment on an outcome, captured by the parameter β_0. We assume that each country is treated exactly once, with country 1 treated in period 1 and country 2 treated in period 2. For simplicity, we also assume that these countries have the same population over time. In Appendix C.5.1, we derive the local projections estimand for β_0 in this case is given by:

$$\beta_0 = \frac{\mathbb{E}\left[Y_{1,1} - Y_{2,1}\right] + \mathbb{E}\left[Y_{2,2} - Y_{1,2}\right]}{2} \tag{3.3}$$

[22] By "staggered treatments," we mean that countries experience tax changes at different times.

where $Y_{g,t}$ is outcome for country g at time t. This simple example makes clear that this estimand does not represent a causal treatment effect for any subgroup. First, this estimand is a simple average of the differences in outcomes in different treatment timing periods. In general, this estimand is a weighted average of differences in outcomes, where the weights are some function of treatment timings (note that (3) is a special form of what is derived in Appendix C.5.1, where algebraic cancellations lead to the simple, relatively interpretable form). Second, this estimand makes comparisons in outcomes between treated and already-treated groups. The first difference in the numerator compares the period 1 outcomes for country 1 (treated) and country 2 (untreated), which is a valid comparison for estimating treatment effects. The second difference in the numerator, however, is invalid: it compares the period 2 outcomes for country 2 (newly treated) and country 1 (previously treated). This parallels the problem pointed out by Goodman-Bacon (2021) in the two-way fixed effects setting.

These issues with the linear projection estimator carry through to our more complex setting. The same conclusions apply to analyses of the local projections estimand for $h \geq 1$ periods out from treatment. More substantively, with more than two countries with staggered treatment timings, the local projections estimand introduces more pairwise comparisons in outcomes between countries, thereby potentially increasing the number of comparisons between treated- and already-treated groups (Goodman-Bacon, 2021). To further complicate analyses, in settings with repeated and/or non-monotone treatments (i.e., multiple tax increases that change in sign over time), Appendix C.5.1 shows that we will likely end up with a more complicated estimand that continues to make comparisons between treated- and already-treated units, and weights these comparisons by some function of treatment timings. Presenting formal results in these complicated settings is beyond the scope of our paper, and remains an open problem.

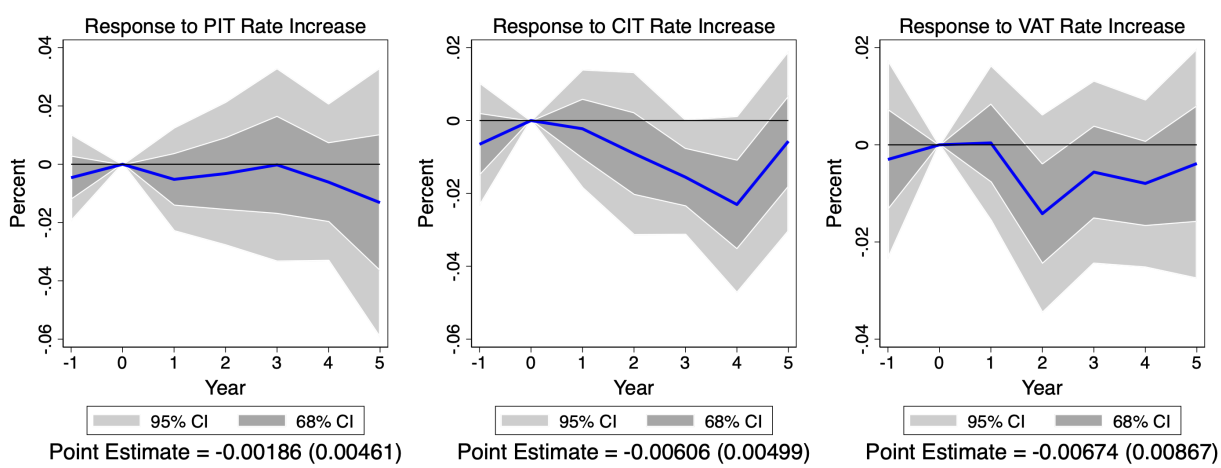
For distinct situations with variation in treatment timing, there now exist estimators that solve the problem of forbidden comparisons. Goodman-Bacon (2021), Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Borusyak et al. (2023) derive estimation strategies for staggered binary treatments, allowing for the possibility of heterogeneous treatment effects across treated cohorts. Some work also tackles the issue of accounting for staggered continuous treatments (Callaway and Sant'Anna, 2021). However, in our setup—where a country can experience both tax increases and decreases multiple times—none of these new approaches applies directly. In fact, it remains an open problem to design heterogeneity-robust estimators in settings with multiple treatments that allows for treatments to vary in sign and intensity (Schmidheiny and Siegloch, 2023).

To our knowledge, the only estimator that can currently provide a causal treatment effect estimate in our context is that developed in de Chaisemartin and D'Haultfoeuille (2022) (henceforth referred to as the dC-D'H estimator), although the treatment effect that is identified is limited. This estimator is a generalized event-study approach that estimates the causal effect of the first observed tax change for a country that occurs within our panel. Effectively, this proposed approach avoids making comparisons between treated- and already-treated groups by aligning event times. The treatment effect for country g is estimated by tracing the evolution of its economic growth rate relative to that of countries that have not yet changed their tax rates. Specifically, for the set of countries g, that experience the first treatment in the same year, we compute the difference in outcome between periods $F_g - 1$ and $F_g + h$, where F_g is the time period that group g is exposed to treatment. We then compute and aggregate the differences in outcome between periods $F_g - 1$ and $F_g + h$, for all countries that were untreated from period 0 to $F_g + h$. The estimate of timing group g's treatment effect, after exposed to h periods of treatment, is the difference between these two differences. These differences are aggregated into an average effect of being first exposed to a weakly higher tax rate, representing average treatment effects on the treated (ATTs) estimates.

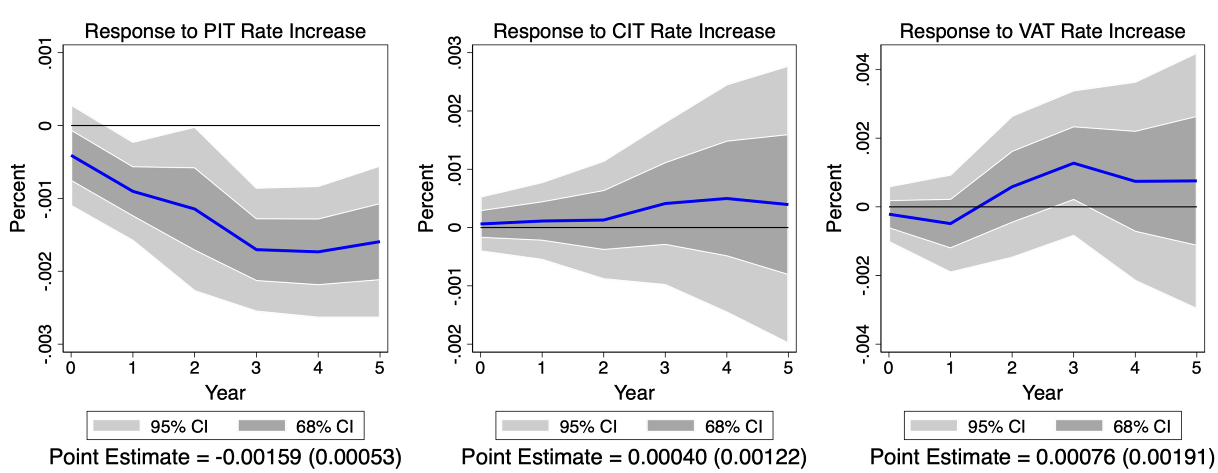
The resulting event study graphs, presented in Panel A of Figure 3.5, trace out effects relative to the distance to the first tax rate change on the horizontal axis. For each year $h \geq 0$ since first treatment, the estimated parameters, β_h, are ATTs among groups that are exposed to $h + 1$ periods of treatment. In these specifications, we include the same controls as before. To aid in interpretation, we also present the average total effect of a 1-percentage point increase in the tax rate below each figure. The total effect is the sum of instantaneous and dynamic effects of a treatment, and is the weighted average across h of the reduced-form estimates divided by the weighted average across h of the first-stage estimators (shown in Appendix Figure C.7). The first stage replaces the outcome with the treatment and shows the average exogenous tax rate increase.

An important caveat for this approach is that the estimator assumes that treatments that occur prior to the start of the panel *do not affect potential outcomes*, i.e., that these pre-panel tax changes do not affect the anticipated behavioral response. With the presence of never-treated units at the start of the panel, the main theoretical results of de Chaisemartin and D'Haultfoeuille (2022) hold even if potential outcomes depend on pre-panel treatments. In our case, we know that tax changes have occurred prior to the start of the panel and these may have associated behavioral responses that are not accounted for by the estimator. To account for this, we employ a correction for scenarios where there is uncertainty about when a country g first receives treatment. This correction drops from the estimation of parameters of interest country-years where earlier treatments were not observed, but includes them as

Figure 3.5: Five-Year Effects of Exogenous Tax Rate Shocks



Panel A: de Chaisemartin and D'Haultfoeuille (2022) estimation



Panel B: Local Projections

Notes: The effects of a one-point increase in tax rates on GDP per capita pver the subsequent five years. Panel A shows the results using the methodology of de Chaisemartin and D'Haultfoeuille (2022). Panel B shows the results using local projections with the sample restricted to be the same as used in Panel A. Each panel shows effects separately for each tax type—the top marginal personal income tax rate (PIT), the top marginal corporate income tax rate (CIT), and the standard consumption tax rate (VAT). Darker and lighter shading indicate 68% and 95% confidence intervals, respectively. Point estimates for the fifth year are below each plot. In parenthesis, Panel A uses standard errors constructed using the statistical package in de Chaisemartin and D'Haultfoeuille (2022) and Panel B uses Driscoll-Kraay standard errors.

controls once we observe changes in treatment.

While the estimates in Figure 3.5 use only a small subset of the policy variation contained in our data,[23] it is clear that the dynamic effects obtained from the dC-D'H estimator point to quite different medium-run responses to tax policy shocks than those found in Section 3.4. While neither are statistically significant at the 95% level, a PIT rate increase now has a contractionary effect on the economy five years later, whereas the linear projection model suggested a positive effect. Rather than a relatively flat response to a CIT rate increase, there are shorter-run decreases in economic activity suggested by the dC-D'H estimator, although there are still no statistically significant effects on GDP five years out. And the effect of a VAT rate increase now becomes flatter and closer to zero when allowing for heterogenous treatment effects. Regardless, it is important to note that the treatment effects estimated here correspond to the average treatment effect of the *first* observed tax treatment that some countries experience.

Panel B of Figure 3.5 shows that even when we restrict the linear projections model to use the same estimation sample and policy variation exploited in the dC-D'H estimator, the conclusions drawn from these estimators diverge. The contractionary dynamic effects of a PIT rate are statistically significant in the local projections, in contrast to the wide standard errors of the new estimator. The dynamic effects of a CIT rate hike are less contractionary in the local projections, as are VAT hikes.

3.6 Conclusion

We set out to offer a transparent look into the consequences of the crucial design choices that underlie existing estimates of how tax policy affects economic growth using cross-country panel data. By evaluating how the estimated effects change when making different choices—to use the full set of tax policy changes or only those that are plausibly exogenous, to measure policy shocks as tax rate changes or a function of projected tax revenues changes, and to employ a linear regression or linear progression model—we shed light on the extent to which we can rely on the existing estimates to inform tax policy. We find that these alternatives matter and can meaningfully alter the lessons drawn from the resulting estimated effects. Accounting for the fact that tax policies are sometimes endogenously undertaken with respect to the state of the economy and that some measures of tax shocks partially reflect anticipated responses to these changes both attenuate the estimated relationship between tax rates and economic growth. It is also important to acknowledge that not all tax parameters are expected to have the same effect on economic growth, and to disentangle effects across different tax bases.

But recent advances in the econometrics literature also demonstrate that because the

[23] Specifically, 1981-1996 for PIT rates, 1981-2011 for CIT rates, and 1981-2006 for VAT rates.

standard approaches previously used fail to account for the fact that countries frequently change their tax policies, they generally yield biased estimates. The sole estimator that can produce an unbiased estimate in our setting—where countries can experience multiple tax rate changes over the panel—is limited to estimating the effect of only the first observed tax change, and points to quite different medium-run responses than our previous analyses using these earlier approaches. It remains an open problem to design heterogeneity-robust estimators in settings where groups can be treated with different signs (non-monotonically) and potentially in multiple periods. Given that settings with multiple tax policy changes within a cross-country panel are ubiquitous, empirical researchers may be better off in the meantime focusing on specific natural policy experiments, where credible and meaningful treatment and comparison groups can be constructed.

Causal estimates of the impact of taxation on economic growth are paramount for the optimal design of tax systems. Our exercise reveals that previous analyses that attempt to estimate these effects using cross-country data do not credibly support claims that tax rate changes have a statistically robust medium-term impact on national output. Moreover, the existing econometric methods available at present are unable to remedy this situation. As such, there remains much uncertainty around any claims made about the effects of tax policy shocks on economic growth based on current estimates.

APPENDIX A

Appendix for "Benefits Cliffs and Aggregate Fluctuations"

A.1 Empirical Appendix

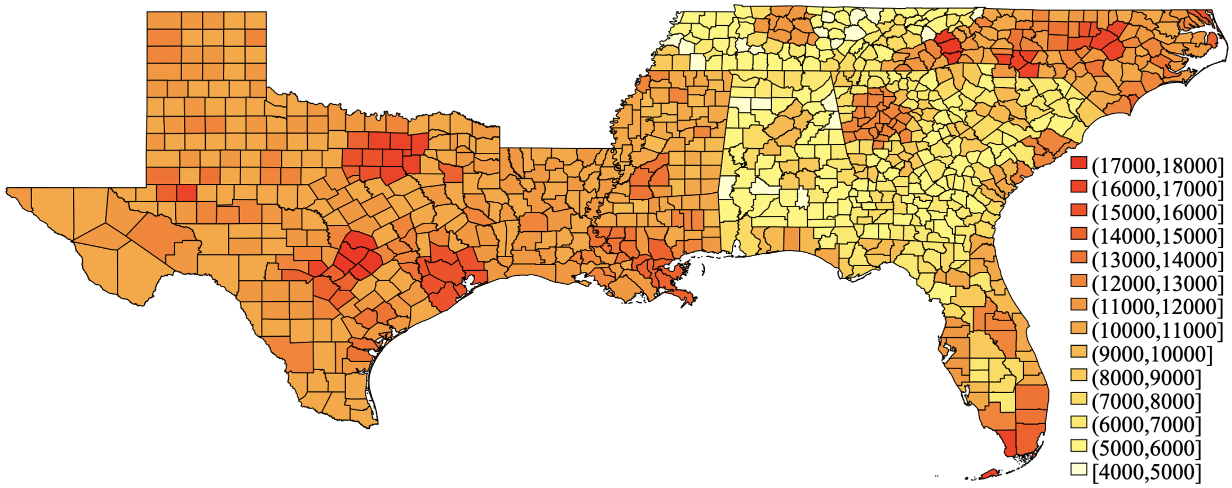
A.1.1 Data

I assign every household a location relative to the cliff by mapping all available demographic and geographic variables to those reported in the BCW. I record household incomes and reported hours worked by all working members of the household in the ACS. Incomes are inflated to 2020 dollars, and I crosswalk the smallest unit of geography in the ACS - Public Use Microdata Areas (PUMAs) - with counties using GEOCORR from the Missouri Census Data Center. I exclude Section 8 housing benefits temporary programs enacted during the COVID-19 pandemic from consideration. This is due to the rationing and thus very limited take-up of Section 8 housing benefits, and excessive churn in labor markets and threat to external validity from the pandemic.

The BCW assumes fungibility of benefits and does not take into account non-pecuniary benefits or costs of these programs, such as desperation to maintain Medicaid status for children or the hassle costs of signing up for programs. Some benefits, like SNAP, are given in dollar amounts. Other benefits, like Medicaid, are in-kind benefits, for which the BCW uses outside data to make assumptions about the value of these benefits. For example, for Medicaid benefits, the BCW uses data from actuarial tables provided by the Centers for Medicare and Medicaid services, which have estimates for benefits paid per beneficiary per month based on state, age, and disability status. To the extent that these valuations are wrong, they should be wrong across all observations. Finally, these valuations would only affect the sizes of cliffs, but not their pre-tax/transfer earnings locations.

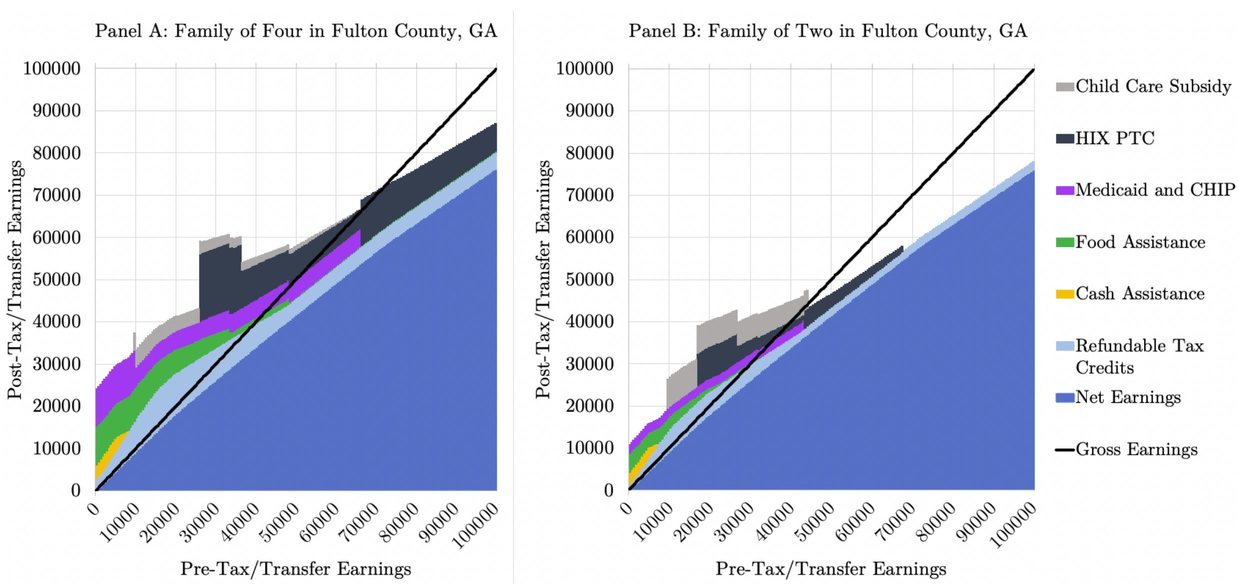
The constructed dataset contains two important sources of variation that I will use to discern the impacts of benefits cliffs: across households within the same geographic subunit,

Figure A.1: Maximum Eligible Benefits for a Single 30-year-old Woman



Note: Values are in 2020 USD. Darker colors indicate a higher maximum eligible benefits.

Figure A.2: Pre/Post-Tax-And-Transfer Earnings for Two Families in Georgia



Note: "HIX PTC" stands for health insurance premium tax credits as part of the Affordable Care Act. CHIP is for the Children's Health Insurance Program. The figure abstracts away from Section 8 housing vouchers and subsidies. Values are in 2020 dollars.

and geographic subunits with the same type of household. Figure A.1 shows the maximum benefits a single 30-year-old woman would be eligible for each county in the BCW. Although the household is exactly the same in each county, differences in state and county policies cause maximum benefits range from over $4,000 to nearly $18,0000. While federal policy

Table A.1: Descriptive Statistics

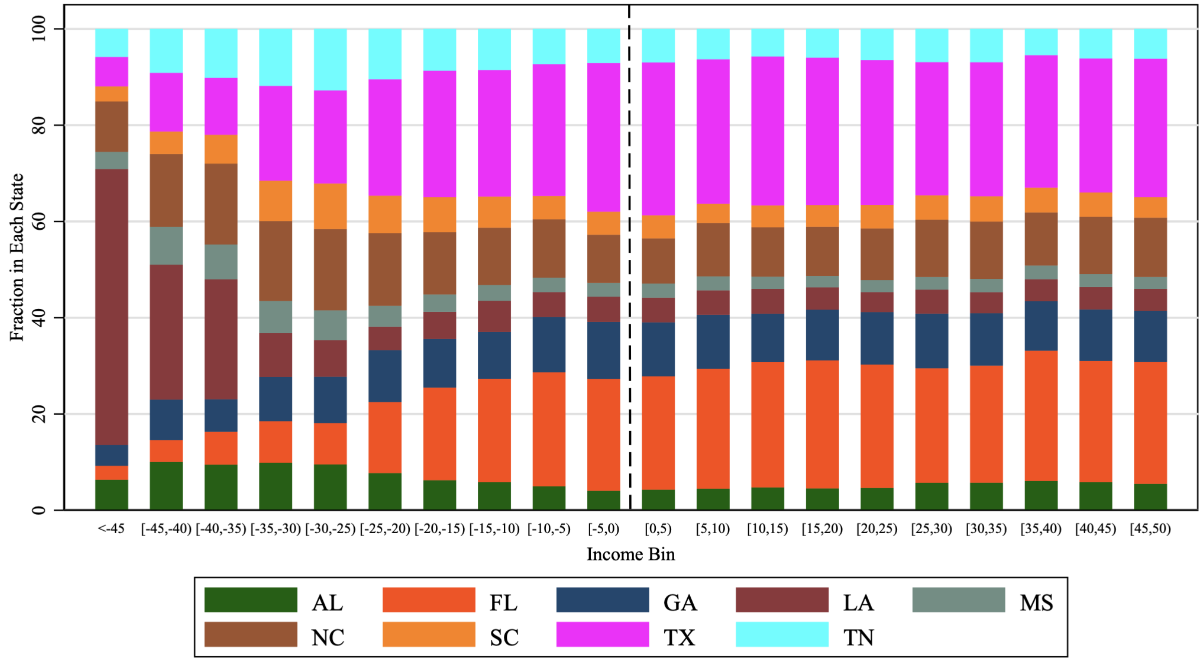
Variable	Restricted Sample		Full Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Household income (USD)	51158.160	25453.260	46568.42	27263.3
Personal income (USD)	26254.580	21013.970	23716.17	21395.44
Annual hours worked	1844.757	736.687	1702.374	850.7356
Number of Children	0.739	1.113	0.7241966	1.111889
Age (years)	40.071	13.431	40.09461	13.61606
Cliff Size (USD)	-4105.964	4886.898	-4464.681	4897.274
Earnings less threshold (USD)	5345.213	19005.850	6174.57	19408.6
% Female (binary)	0.513		0.5214444	
% Black (binary)	0.238		0.2455421	
% Married	0.303		0.284599	
N	162,856,318		194,101,124	

Note: The restricted sample drops those with no reported income or working hours, and restricts ages to 18-64. The full sample removes these two restrictions.

should drive no geographic variation here, state and county policy may. For example, Alabama did not expand its Medicaid program following the passage of the Affordable Care Act. Items such as free and reduced lunch and childcare subsidies drive the variation between counties. The second source of variation is illustrated in A.2. Here, I replicate the pre/post-tax/transfer diagram of 1.1, with pre-tax/transfer earnings along the x-axis and post-tax/transfer earnings along the y-axis. This figure shows the budget schedules for two different families living in the same county in Georgia. Panel A shows the schedule for a married couple with two kids, whereas Panel B shows the same for a single mother of one. The size and location of benefits cliffs may change with the change in household composition. For example, the members of the family of four experience a $5,000 benefits cliff at around $35,000 of pre-tax/transfer income, where no such cliff exists there for the family of two.

Table A.1 lays out some descriptive statistics of the data. Given my focus on the intensive margin of labor supply, I restrict the sample to persons who report strictly positive wage income and annual working hours. I drop persons with reported ages of above 64 or below 18. Given that the BCW does not contain data for household incomes exceeding $100,000, those observations are also dropped. This leaves me with over 1.4 million observations that, when weighted, cover over 162 million persons. When it comes to benefits cliffs, the average

Figure A.3: Composition of Income Bins by State



Note: The x-axis is the person's household's distance relative to their nearest cliff. The figure illustrates that the composition of income bins around the cliff is relatively stable.

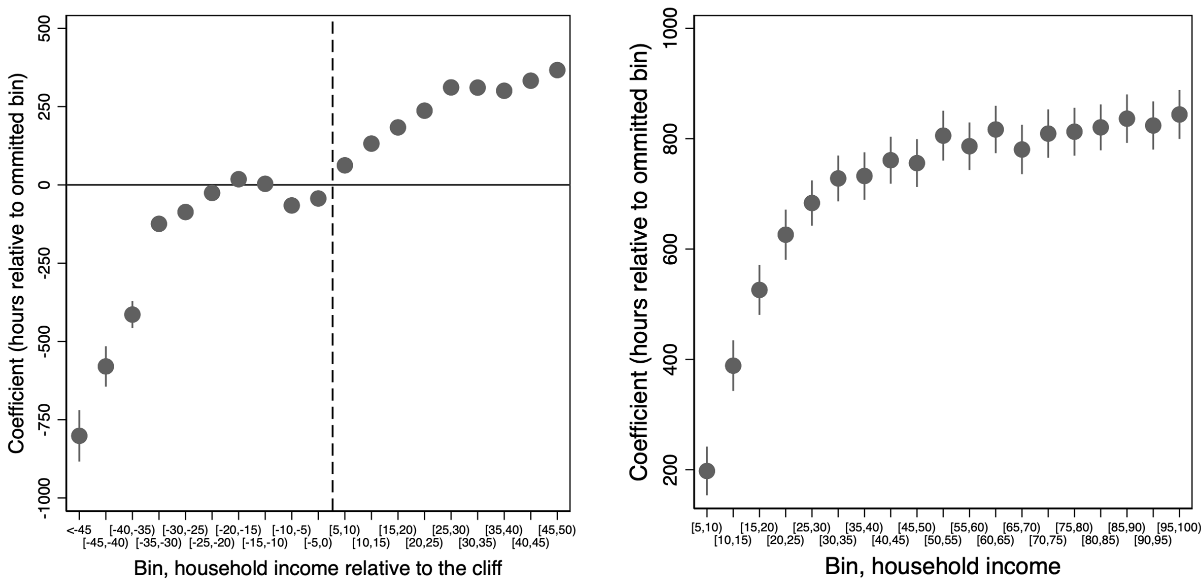
size is $4,105.96, though most households are ahead of their closest benefits cliff. A larger fraction of the sample population identifies as Black than in the US as a whole since the sample comes from the Southeastern US.

A.1.2 Additional Notes on Binned Regressions

In the binned regressions, there may be concern that the composition of geographies in bins changes substantially around the cliff. While I include county fixed effects, Figure A.3 also shows that the state composition of bins around the cliff is smooth. There may also be a temptation to interpret the sum of reduced-form coefficients as the net effect of benefits cliffs, though this ignores the natural positive relationship between hours and income. Figure A.4 shows this relationship, as seen just after the cliff in the binned regressions, though there is a dip in hours worked when plotted against the cliff.

Another concern may be the curious result that the dip in reported annual hours is not strongest when immediately prior to the cliff. This can be happening for several reasons: one is measurement error - the ACS is survey data, where households are prone to rounding incomes and hours worked. There may also be measurement error in the form of surveys

Figure A.4: Zoomed-Out Binned Regressions



Note: The left panel contains the same outcomes as in Figure 3, but zoomed out to include more bins of household income relative to the cliff (the omitted bin). The right panel shows the relationship between annual hours worked and income, with the omitted bin being below $5,000 of income.

failing to capture work done under the table - which may be very likely for those near cliffs who want additional income without losing their benefits. Finally, there is a whole range of properties from "unsophisticated agents" that could lead to the observed results. People may interpret their average tax rate as their effective marginal tax rate, or be unaware of the existence of cliffs at all. These individuals would have fewer qualms with being just behind benefits cliffs, and from their perspective would have muted or no reasons to hedge their hours. By underestimating their marginal cost of labor, these individuals very close to cliffs may push up the average reported annual hours.

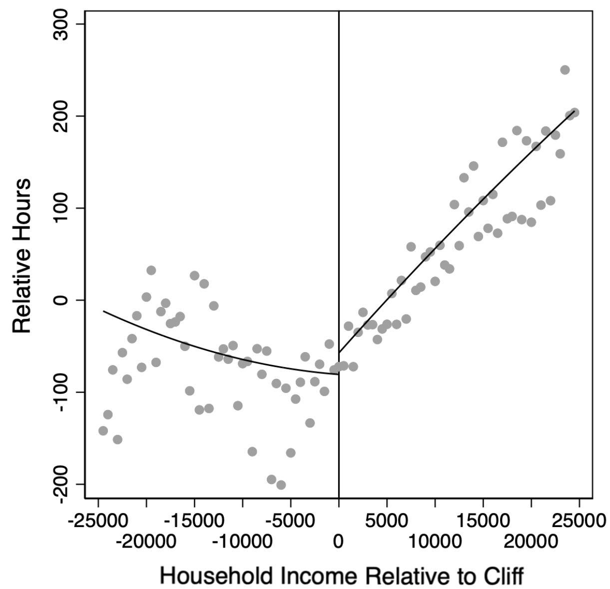
A.1.3 Regression Discontinuity Design

To gauge the robustness of results to this binned methodology, I also employ a regression-discontinuity design (RDD). Here, I use a parametric design with a second-order polynomial:

$$H = \alpha + \beta D + \gamma_0 c + \gamma_1 c^2 + \gamma_2 D \times c + \gamma_3 D \times c^2 + \gamma_4 X + \delta_c + \delta_y + \delta_c \times \delta_y + \varepsilon \qquad \text{(A.1)}$$

where many variables have the same definition as before. D is an indicator for households whose pre-tax/transfer income is greater than the pre-tax/transfer level of the cliff and c

Figure A.5: Regression Discontinuity Design



Note: The x-axis indicates a person's household's income relative to their nearest cliff. The y-axis indicates values of residuals from equation A.1.

is household income relative to the cliff. I include the wage and demographic controls in the vector X. I present visual evidence from this by plotting the residuals of regressing the outcome variable on controls in Figure A.5.

Here, the fitted line is a result of regressing the residuals on a quadratic function of the person's household's distance from their nearest cliff, interacted with a post-cliff dummy. The x-axis is the household's income relative to the cliff and changes along the y-axis correspond to changes in hours. The coefficient on the after-cliff indicator implies a 12.81 [SE: 4.25] reduction in annual hours worked just before the cliff. Regressing the residuals on a linear function (as opposed to quadratic) captures more of the dip in hours for persons in households with their nearest cliff $10,000 or less ahead of them. Here, the coefficient on the after-cliff indicator implies a 29.73 [SE: 3.15] loss in hours - closer to the losses found in the binned regressions.

A.1.4 System of Equations

As mentioned in Section 1.2.3, the main specification of the paper is reduced form. Equation 1.7 from the main paper, reproduced below, demonstrates that there is endogeneity between the hours choice $h_{i,t}$ and distance to the cliff $\hat{\theta}_{i,t}$:

$$\hat{\theta}_{i,t} = z_i - a_i w_t h_{i,t}$$

which prevents the results from the reduced form from having a causal interpretation. What may be the effect of this simultaneity on estimated coefficients? Consider a simplified version of this problem:

$$\begin{aligned} h_{i,t} &= \alpha + \beta_b \hat{\theta}_{i,t} \\ \hat{\theta}_{i,t} &= z_i + \beta_2 a_i w_t h_{i,t} \end{aligned}$$

where β_b is the coefficient on distance to the cliff and β_2 is the strength of the relationship between income earned through wages and distance to the cliff.[1] Solving each expression for $h_{i,t}$ and equating the two yields:

$$h_{i,t} = \left(\alpha + \frac{z_i}{\beta_2 a_i w_t}\right) + \left(\beta_b - \frac{1}{\beta_2 a_i w_t}\right) \hat{\theta}_{i,t}$$

which reveals that the coefficient on distance to the cliff in reduced form is $\left(\beta_b - \frac{1}{\beta_2 a_i w_t}\right)$. Given that as distance to the cliff declines the hours choice falls and as incomes rise the distance to the nearest cliff ahead should shrink, both coefficients β_b and β_2 should be negative. Hence, the weaker the relationship between wage earnings and distance to the cliff β_b, the more positive bias in estimates for β_b. Moreover, one of the control variables - tax rates - is endogenous to the hours decision, as the choice of hours affects income and therefore the tax bracket agents are in.

To account for this endogeneity, I estimate a system of equations that includes these interactions between distance to the cliff, the hours decisions and tax rates, which allows for accounting of correlations across equations. Specifically, I estimate

$$\begin{aligned} H &= \alpha + \sum_{b=1}^{n} \beta_b B_b + \gamma X + \delta_c + \delta_y + \delta_c \times \delta_y + \varepsilon \qquad \text{(A.2)} \\ \Pr\left(B_b = 1 | z, H, X\right) &= \Phi\left(\rho + z + \eta H + \kappa X + \mu\right) \ \forall \ b \\ 1 - \tau &= \nu + \phi H + \chi X + \zeta \end{aligned}$$

where ν is a constant, and other variables are the same as before - with the exception of an

[1] It is not necessarily the case that $\beta_2 = -1$, since households may have sources of income that are not from wages

Table A.2: Results for System of Equations

Coefficients on Income Bins (thousands of dollars)		
DV: Total annual hours	System of Equations	Reduced Form
For [-10,-5) bin	-79.74989	-65.5903
	[16.96578]	[5.390936]
For [-5,0) bin	-52.59004	-43.39509
	[13.54418]	[3.202076]
Coefficients on Hours in Probit Regressions. DV: Indicator for bin		
For [-10,-5) bin	5.547998	
	[0.82643]	
For [-5,0) bin	6.169395	
	[0.97749]	
Coefficients in Tax Regression. DV: Net tax rate		
Constant	0.83008	
	[0.07251]	
Hours	-0.0001438	
	[0.00003]	
Demographic, wage, and tax controls	X	X
State and Year FE	X	X
N	3,228,921	129,869,922

Note: Standard errors are in brackets and clustered at the county level. The first column of coefficients show results for system of equations, whereas the second column of coefficients shows the results of Figure 3 in the main paper for bins of interest. The results from the system of equations are estimated using 5% of the sample for speed, increasing their standard errors.

additional control for the size of the cliff, in line with the simple structural model of equation 1.8. z is the location of the nearest cliff ahead, and both μ and ζ are an additional error terms. η and ϕ are newly-estimated coefficients on hours in the equations for distance to the nearest cliff ahead and tax rates. The first equation is the binned regression, augmented so that the bins B_b are distance to the nearest cliff ahead. The second equation captures the endogeneity between bin location and the hours decision using a probit model, and the

Table A.3: Heterogeneous Responses to Cliff Sizes

Cliff Size (percent of household income)	Binned Regression		Parametric RDD	
	Baseline	System of Equations	Quadratic	Linear
First Quartile	-31.69	-34.67	-1.98	-20.68
(0-1.75%)	[2.07]	[8.67]	[8.66]	[6.55]
Second Quartile	-28.20	-38.98	-4.97	-36.5
(1.75-5.65%)	[2.36]	[9.36]	[8.42]	[5.89]
Third Quartile	-46.24	-56.88	-21.08	-45.76
(5.65-15.41%)	[2.67]	[13.08]	[8.13]	[5.95]
Fourth Quartile	-67.06	-82.48	-44.02	-66.88
(>15.41%)	[7.31]	[18.15]	[9.88]	[14.14]
N (for each quartile)	39,863,840	1,623,374	26,010,562	26,010,562

Note: Standard errors are in brackets and clustered at the county level.

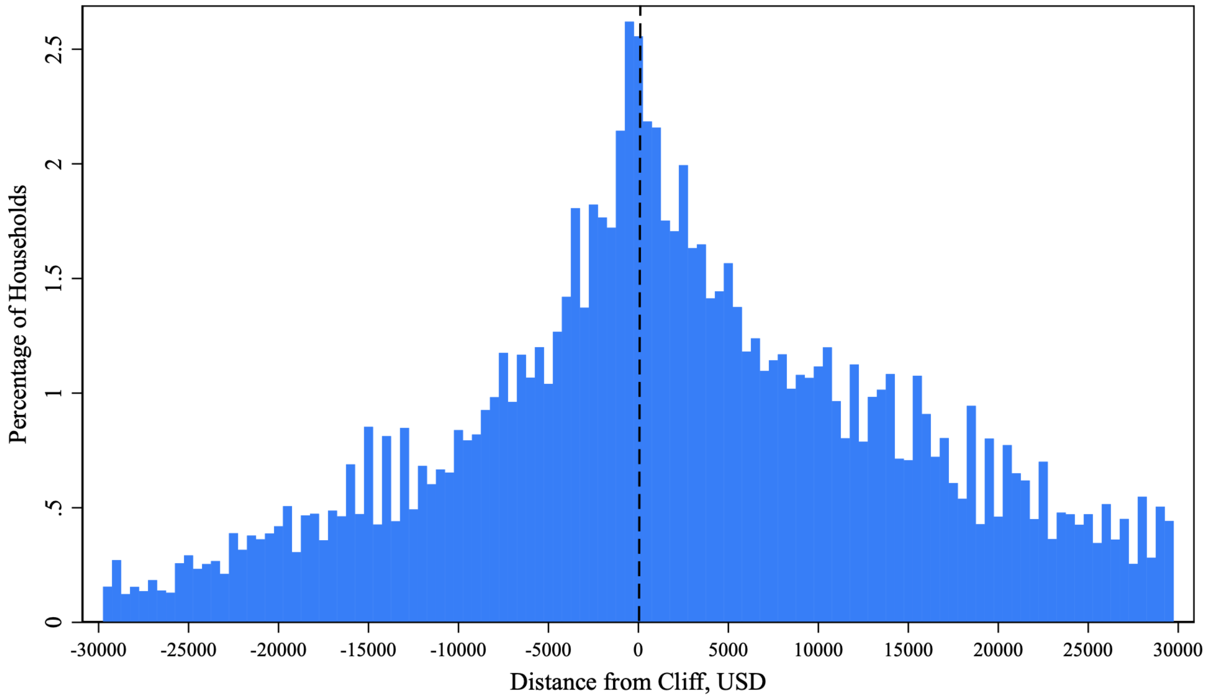
final equation captures the endogeneity between tax rates and working hours. Again, if the effect on hours increases as individuals get closer to the cliff, we should expect more negative coefficients β_b as the cliff approaches. I jointly estimate this system of equations using generalized method of moments (GMM), with the above equations acting as moment conditions.

I show the results of this exercise in Table A.2. These generally conform to the theory outlined above, finding point estimates slightly more negative than in the binned regression. For persons whose household is within $10,000 of their nearest cliff ahead, annual hours worked fall by about 52.6-79.8 hours [SE:13.54-16,97].

A.1.5 Heterogeneous Outcomes

In order to discern if these are real responses to benefits cliffs, I explore heterogeneous reactions to different cliff sizes, both in the binned regression and RDD frameworks (results for the system of equations are in progress). I start by dividing the sample into quartiles based on the size of their next cliff as a share of household income. Table A.3 shows the results for the reduced form, structural estimation, and the RDD with quadratic or linear trends. Standard errors are in brackets below the point estimates and clustered at the county level. Across two of the three specifications, annual hours worked decrease as the cliff size

Figure A.6: Earnings Relative to Benefits Cliffs



Note: The x-axis indicates a person's household's income relative to their nearest cliff. The y-axis is the density of incomes at that point.

increases. In all specifications, households subject to the highest quartile of cliff size reduce their hours at least about twice more than households facing the lowest quartile of benefits cliffs. I interpret these heterogeneous responses as evidence in favor of a behavioral response to benefits cliffs.

A.1.6 On Bunching

Using administrative data, past work - such as Kleven and Waseem (2013) - finds that taxpayers bunch their incomes before notch points. I started out this project hoping to find the same. Again borrowing from the methodology of Hamersma (2013) and Haider and Loughran (2008), I normalize persons by their household's distance to their nearest cliff. Figure A.6. plots the resulting histogram.

While it does appear that taxpayers may be piling around the notch point, the characteristic missing mass to the right of the notch is, well, missing. Slicing and dicing the data - for different family structures, cliff size, etc. - at most marginally improves the results. How is it then possible that I find behavioral responses in hours worked, if earnings are unaffected?

For one, reducing working hours to target a specific level of earnings is entirely consistent with my microfoundations of Section 1.2.1. The peak at the notch may be the result of this targeting, with many agents "missing" the exact notch point due to optimization errors. A more likely explanation is data quality. Papers that find bunching often use administrative records of income which allow for precise determination of household income relative to the cliffs. The survey data I use, in contrast, is known to have substantial measurement error that makes this approach less useful. Moreover, bunching in earnings may be the result of strategic misreporting to tax authorities, whereas surveys do not carry the same incentives to misreport (Abraham et al., 2021). Finally, the Georgia Center for Opportunity has programs that work directly with individuals encountering benefits cliffs. In my correspondence with the GCO, their director of research, Erik Randolph, notes:

> "Your findings match what we've observed. It was not our expectation to find lots of households clustering near the cliffs' edge for several reasons. One big reason is that for many cliffs, [after-tax/transfer earnings] can be fairly flat ramping up to the edge of the cliff, meaning the incentives to earn more have already been skewed."

Indeed, several cliffs - particularly for households with multiple children - face effective marginal tax rates tax rates exceeding 70% before the cliff. These limit the incentives to bunch right at benefits cliffs.

A.1.7 Anecdotal Evidence

From an interview with Frankie Johnson, recipient of government assistance in Gwinnet County, Georgia. Johnson detailed an experience in which she felt forced to turn down a job placement that would have earned roughly $70,000 a year because of the loss of housing and childcare benefits.

> "They want to see your pay stubs, your bank statements. They want to make sure you're poor. ... If you have a car, they want to know what kind of car you're driving and if you have insurance. They want to make sure there's no possible way you can work a job. ... We need to get them their GEDs and diplomas. Start them off as home health aides, CPAs, LPNs, RNs, physician's assistants, or doctors. ... But no one's willing to help. They just want to enable their programs to get money for housing us. After that, you're out on the street like a dog."

From a case study of New York state's policies to increase wages for home care workers:

> "One worker expressed frustration over not wanting to lose hours at work, yet needing to work one less hour per week to avoid losing SNAP benefits. A $17 gain for that one additional hour of work would trigger a $300-per-month loss in SNAP benefits, a benefit that the client reported having received for the past three years."

From an interview with a recipient (name not given) of childcare subsidies in Colorado:

> "I don't want to make any less, but if I accept an increase in pay, I'm subjected to more child care costs. A couple dollars increase in pay doesn't give me the $600 extra I need to pay for child care (if I lost the subsidy). It sucks. It's literally holding me back from pursuing better opportunities... I'm worried because if I get a raise, or make more money, I can't afford child care."

A.2 Theoretical Appendix

A.2.1 Motivating Model

Consider a person with household income below the threshold of a single cliff. Suppose that individuals are considering labor supply $h_{i,t}$ before the cliff as given and pre-tax/transfer income $y_{i,t}$ is uncertain, where

$$y_{i,t} = a_i w_t h_{i,t} + \theta_{i,t} \tag{A.3}$$

where $\theta_{i,t} \sim F$ with the probability density function f is an income shock. If income is more than the income location of the cliff $y_{i,t} \geq z_i$, then after-tax/transfer income is

$$c_{i,t} = (1 - \tau)\left[a_i w_t h_{i,t} + \theta_{i,t}\right] - T \tag{A.4}$$

where T is the size of the benefits cliff. If $y_{i,t} < z_i$ then after-tax income is

$$c_{i,t} = y_{i,t} = (1 - \tau)\left[a_i w_t h_{i,t} + \theta_{i,t}\right] \tag{A.5}$$

Next, let $\hat{\theta}_{it}$ be the income necessary to reach the income at the cliff z_i, or

$$\hat{\theta}_{i,t} = z_i - a_i w_t h_{i,t} \tag{A.6}$$

which I will also refer to as "distance to the cliff." With labor supply already chosen, utility as a function of hours is

$$E\left[u_{i,t}|h_{i,t}\right]=u(h_{i,t})=E\left[\frac{c_{i,t}^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}\right]-\psi\frac{h_{i,t}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \tag{A.7}$$

where the expectations operator is over the income shocks $\theta_{i,t}$. I now investigate how this income risk, which affects consumption through the changing after-tax income, affects the hours decision before the cliff. Using the integral definition of the expectations operator,

$$\begin{aligned}u(h_{i,t})=&\int_{-\infty}^{\hat{\theta}_{i,t}}\frac{\left[(1-\tau)\left(a_iw_th_{i,t}+\theta_{i,t}\right)\right]^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}f\left(\theta_{i,t}\right)d\theta_{i,t}\\+\int_{\hat{\theta}_{i,t}}^{\infty}&\frac{\left[(1-\tau)\left(a_iw_th_{i,t}+\theta_{i,t}\right)-T\right]^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}f\left(\theta_{i,t}\right)d\theta_{i,t}-\psi\frac{h_{i,t}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}\end{aligned} \tag{A.8}$$

Using the Liebniz integral rule, deriving with respect to labor supply $h_{i,t}$ yields

$$\begin{aligned}&(1-\tau)\,a_iw_t\int_{-\infty}^{\hat{\theta}_{i,t}}\left[(1-\tau)\left(a_iw_th_{i,t}+\theta_{i,t}\right)\right]^{-\frac{1}{\sigma}}f\left(\theta_{i,t}\right)d\theta_{i,t}\\&\quad+\frac{\partial\hat{\theta}_{i,t}}{\partial h_{i,t}}\frac{\left[(1-\tau)\left(a_iw_th_{i,t}+\hat{\theta}_{i,t}\right)\right]^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}f\left(\hat{\theta}_{i,t}\right)\\&+(1-\tau)\,a_iw_t\int_{\hat{\theta}_{i,t}}^{\infty}\left[(1-\tau)\left(a_iw_th_{i,t}+\theta_{i,t}\right)-T\right]^{-\frac{1}{\sigma}}f\left(\theta_{i,t}\right)d\theta_{i,t}\\&-\frac{\partial\hat{\theta}_{i,t}}{\partial h_{i,t}}\frac{\left[(1-\tau)\left(a_iw_th_{i,t}+\hat{\theta}_{i,t}\right)-T\right]^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}f\left(\hat{\theta}_{i,t}\right)-\psi h_{i,t}^{\frac{1}{\epsilon}}=0\end{aligned} \tag{A.9}$$

Combining like terms and using equation A.6 for $a_iw_th_{i,t}+\hat{\theta}_{i,t}$ yields

$$\begin{aligned}&(1-\tau)\,a_iw_t\left\{\int_{-\infty}^{\hat{\theta}_{i,t}}\left[(1-\tau)\left(a_iw_th_{i,t}+\theta_{i,t}\right)\right]^{-\frac{1}{\sigma}}f\left(\theta_{i,t}\right)d\theta_{i,t}\right.\\&\qquad\left.+\int_{\hat{\theta}_{i,t}}^{\infty}\left[(1-\tau)\left(a_iw_th_{i,t}+\theta_{i,t}\right)+T\right]^{-\frac{1}{\sigma}}f\left(\theta_{i,t}\right)d\theta_{i,t}\right\}\\&+\frac{\partial\hat{\theta}_{i,t}}{\partial h_{i,t}}\frac{\left[(1-\tau)\,z_i\right]^{1-\frac{1}{\sigma}}-\left[(1-\tau)\,z_i+T\right]^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}f\left(\hat{\theta}_{i,t}\right)=\psi h_{i,t}^{\frac{1}{\eta}}\end{aligned} \tag{A.10}$$

From equation A.6, note that $\partial\hat{\theta}_{i,t}/\partial h_{i,t} = -a_i w_t$. Also note that the integrals are the marginal utilities times the probability of being above or below the cliff. Thus,

$$h_{i,t} = \left\{ \frac{a_i w_t}{\psi} \times \left\{ (1-\tau)\Big(\Pr[\text{BC}] \times E[MU_{i,t}|\text{BC}] + \Pr[1-\text{BC}] \times E[MU_{i,t}|\text{AC}]\Big) \right.\right. \\ \left.\left. - \frac{[(1-\tau)\, z_i]^{1-\frac{1}{\sigma}} - [(1-\tau)\, z_i - T]^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} f\left(\hat{\theta}_{i,t}\right) \right\}^{\epsilon} \right. \tag{A.11}$$

where $\Pr[BC]$ and $\Pr[AC]$ are probabilities of being below and above the cliff, respectively, and $MU_{i,t}$ are the marginal utilities. For a simple case, assume that $\sigma = \infty$ so that $MU_{i,t} = 1$ and

$$u_{i,t} = c_{i,t} - \psi \frac{h_{i,t}^{1+1/\epsilon}}{1+1/\epsilon} \tag{A.12}$$

Under this assumption, the term with probabilities in equation A.6 collapses and this expression now becomes

$$h_{i,t} = \left\{ \frac{a_i w_t}{\psi} \left(1 - \tau - T f(\hat{\theta}_{i,t})\right) \right\}^{\epsilon} \tag{A.13}$$

which is the normal labor supply condition under this utility, except with an added term consisting of the size of the cliff T and the density of income shocks at size $\hat{\theta}_{i,t}$. If the distance to the cliff is outside the range of income shocks, then $f\left(\hat{\theta}_{i,t}\right) = 0$, the last term of equation A.13 drops out, and we recover the normal labor supply condition

$$h_{i,t} = \left\{ \frac{a_i w_t}{\psi} (1-\tau) \right\}^{\epsilon} \tag{A.14}$$

Otherwise, there is a negative effect on the hours decision before the cliff, with a testable prediction that as the cliff size increases, the hours decision declines. Finally, note that this equation only applies for individuals choosing hours before the cliff.

Special Case 1: What if distance the cliff is zero, or $\hat{\theta}_{i,t} = 0$? In this case, agents would be inframarginal and behave according to equation 1.5, as the tax-and-transfer system is not differentiable at this point. However, if we consider a distance $\hat{\theta_{i,t}} = \varepsilon$ where $\varepsilon > 0$, then we are at a differentiable part of the budget constraint. Here, the majority of the possible income shocks will be past the cliff, and the detrimental effect on working hours increases as ε approaches zero. Finally, consider the assumption that agents at the cliff consider their effective marginal tax rate for another dollar to be the same as their current effective marginal tax rate - or, alternatively, they consider their average effective tax rate to their marginal effective tax rate. If this is the case, agents at the cliff behave marginally, but again the majority of possible income shocks will be past the cliff, increasing the detrimental effect on working hours.

Special Case 2: What if distance to the cliff is negative, or $\hat{\theta}_{i,t} < 0$ and agents find themselves ahead of the cliff? Without constant marginal utility, agents may react in one of two ways: one, the loss of benefits induces an income effect that causes individuals to work more.[2] Two, there is a dominated region of working hours in which agents would not wish to locate. If working more hours to achieve the same or higher level utility beyond the cliff as at it is not possible, they would instead wish to reduce their hours to locate at the cliff. Specifically, in the steady-state for a given ability and preference parameters, agents will reduce their hours from their hours decision after the notch h_{AN} to the hours necessary to earn z_i and locate at the notch, or h_N if

$$\frac{c_N^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \psi \frac{h_N^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \geq \frac{c_{AN}^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - \psi \frac{h_{AN}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \tag{A.15}$$

with consumption at the notch given by c_N and consumption after the notch by c_{AN}. This states that agents who would otherwise choose hours h_{AN} after the notch would instead choose to consume and work at the earnings location of the notch, z_i if doing so grants higher utility. While closed-form solutions are difficult to obtain, we can see the likelihood of hours reductions increasing as the cliff size T becomes more negative (the cliff size grows). This can be seen most easily by solving the equation for h_N:

$$h_N \leq \left(h_{AN}^{1+1/\epsilon} - \frac{(1+1/\epsilon)\left(C_{AN}^{1-1/\sigma} - C_N^{1-1/\sigma}\right)}{\psi(1-1/\sigma)} \right)^{\frac{\epsilon}{\epsilon+1}} \tag{A.16}$$

where the difference in benefits at and after the cliff induces differences in consumption C_{AN} and C_N. Assuming again that $\sigma = \infty$ for simplicity, the hours decision here is

$$h_N \leq \left(h_{AN}^{1+1/\epsilon} + \frac{1+1/\epsilon}{\psi(1-1/\sigma)} T \right)^{\frac{\epsilon}{\epsilon+1}} \tag{A.17}$$

which makes it clear that as the cliff size grows, an agent with a given ability is more likely to reduce their hours to locate at the notch and earn higher utility. Moreover, we see the importance of the parameters for the elasticity of intertemporal substitution σ and Frisch

[2]In the setting with constant marginal utility, consumption does not enter the hours decision and there is no income effect on labor supply from the loss of benefits

elasticity ϵ, as they govern the shape of utility and thus the range of dominated hours. As either σ or ϵ grow, ceteris paribus, the likelihood of hours reductions fall. This is because as these parameters grow, the relationship between hours and ability becomes more positively-steeped, decreasing the range of dominated hours. Hence, the calibrations of the benefits cliffs T, as well as preference parameters σ and ϵ, will be crucial.

A.2.2 Log-Linearization

For concreteness, I will assume a distribution of the income shocks. Specifically, due to downward wage rigidity, shocks should have a mean μ above zero. For tractability I will assume a normal distribution:

$$f(\hat{\theta}_{i,t}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{\theta}_{i,t}-\mu}{\sigma}\right)^2} \tag{A.18}$$

where σ in the above expression is the standard deviation. Under the same utility assumptions, this changes equation A.13 to

$$h_{i,t} = \left\{ \frac{a_i w_t}{\psi} \left(1 - \tau - T\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{\theta}_{i,t}-\mu}{\sigma}\right)^2} \right) \right\}^{\epsilon} \tag{A.19}$$

I then start log-linearizing by taking the log of both sides

$$\ln\left(h_{i,t}\right) = \epsilon\ln(a_i) + \epsilon\ln(w_t) - \epsilon\ln(\psi) + \epsilon\ln\left(1 - \tau - T\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{\theta}_{i,t}-\mu}{\sigma}\right)^2} \right) \tag{A.20}$$

I then do a first-order Taylor expansion around steady-state values (denoted by S):

$$\begin{aligned} \ln\left(h^S_{i,t}\right) + \frac{h_{i,t}-h^S_{i,t}}{h^S_{i,t}} &= \epsilon\ln(a^S_i) + \epsilon\frac{a_i - a^S_i}{a^S_i} + \epsilon\ln(w^S_t) + \epsilon\frac{w_t - w^S_t}{w^S_t} - \epsilon\ln(\psi^S) - \epsilon\frac{\psi-\psi^S}{\psi^S} \\ &+\epsilon\ln\left(1 - \tau - \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{\theta}^S_t-\mu}{\sigma}\right)^2} \right) + \epsilon\frac{T\left(\hat{\theta}^S_t - \mu\right) e^{\frac{1}{2}\left(\frac{\hat{\theta}^S_t-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma^3\left(1 - \tau - \frac{Te^{\frac{1}{2}\left(\frac{\hat{\theta}^S_t-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma}\right)}\left(\hat{\theta}_{i,t} - \hat{\theta}^S_{i,t}\right) \end{aligned} \tag{A.21}$$

Note that many terms on both the left and right-hand side cancel in the steady-state and that parameters - including abilities - are constant, and thus their differences from their steady-state values are zero. Finally, I multiply the last term by $1 = \frac{\hat{\theta}^S_t}{\hat{\theta}^S_t}$ this yields the log-linearized expression

$$\tilde{h}_{i,t} = \epsilon \tilde{w}_t + \epsilon \frac{T\hat{\theta}_t^S \left(\hat{\theta}_{i,t}^S - \mu\right) e^{\frac{1}{2}\left(\frac{\hat{\theta}_t^S - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma^3 \left(1 - \tau - \frac{Te^{\frac{1}{2}\left(\frac{\hat{\theta}_t^S - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma}\right)} \tilde{\hat{\theta}}_{i,t} \tag{A.22}$$

where $\tilde{x}$ denotes the percent deviation of x from its steady-state value. As income grows relative to the cliff, $\hat{\theta}_{i,t}$ falls via equation A.6. The more this occurs, $\tilde{\hat{\theta}}_{i,t}$ grows increasingly negative. Thus, this again shows the desirable property that, under a notched tax system with income uncertainty, the hours choice before the cliff falls as agents get closer to the cliff.

Next, note that $\tilde{x} \approx \ln(x) - \ln(x^S)$. Thus, equation A.19 is approximately equivalent to

$$\ln(h_{i,t}) - \ln(h_{i,t}^S) = \epsilon \left(\ln(w_t) - \ln(w_t^S)\right) + \epsilon \frac{T\hat{\theta}_t^S \left(\hat{\theta}_{i,t}^S - \mu\right) e^{\frac{1}{2}\left(\frac{\hat{\theta}_t^S - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma^3 \left(1 - \tau - \frac{Te^{\frac{1}{2}\left(\frac{\hat{\theta}_t^S - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma}\right)} \left(\ln(\hat{\theta}_{i,t}) - \ln(\hat{\theta}_{i,t}^S)\right) \tag{A.23}$$

I now add the steady-state value of labor supply to both sides:

$$\begin{aligned} \ln(h_{i,t}) = \epsilon \ln(w_t) + \epsilon \frac{T\hat{\theta}_t^S \left(\hat{\theta}_{i,t}^S - \mu\right) e^{\frac{1}{2}\left(\frac{\hat{\theta}_t^S - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma^3 \left(1 - \tau - \frac{Te^{\frac{1}{2}\left(\frac{\hat{\theta}_t^S - \mu}{\sigma}\right)^2}}{\sqrt{2\pi}\sigma}\right)} \left(\ln(\hat{\theta}_{i,t}) - \ln(\hat{\theta}_{i,t}^S)\right) + \\ \epsilon \ln(a_i) - \epsilon \ln(\psi) + \epsilon \ln \left(1 - \tau - T\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\hat{\theta}_{i,t} - \mu}{\sigma}\right)^2}\right) \end{aligned} \tag{A.24}$$

Collecting all constants into a single term $\alpha_{i,t}$ yields

$$\ln(h_{i,t}) = \alpha_{i,t} + \epsilon \ln(w_t) + \beta \ln(\hat{\theta}_{i,t}) \tag{A.25}$$

where β, the effect of distance to the cliff on labor supply, is a nonlinear combination of parameters including the Frisch elasticity ϵ, the tax rate τ, and those that govern the distribution of income shocks.

A.2.3 Relaxing Discrete Households

Assume for a moment that rather than have discrete households, some of which are inframarginal, I have a continuum of households. I would need to use value-function iteration to construct policy functions for each household type. Aggregates with this method would be generated by simulating thousands of agents and collecting the resulting moments, so that results stem from more of a black box. Solving for the steady-state and implementing aggregate shocks via perturbation in my model takes 2 minutes on a 2020 MacBook Pro. Solving a version of the model via value function iteration with aggregate shocks via simulation would take substantially longer. Already, lots of ink has been spilled showing that much of what a continuum of heterogeneous agents gives you in terms of aggregate implications can be mostly be accomplished with two types of agents (see work done by Florin Bilbiie in this area). The value-added from full-fledged models comes from more precise analysis of distributional implications. This paper tries to strike at the happy middle.

The most important thing relaxing discrete households yields for dynamics is that, in response to positive wage shocks, this allows for the possibility for people to "jump" past dominated regions of benefits cliffs for sufficiently high wage shocks. This should be of second-order importance. As stated in 1.2.1, all households with abilities $[a_1, a^*)$ will reduce their working hours, where a^* is a critical value of ability $a^* \in (a_1, a_2]$. Agents above this critical value of ability will increase their hours in response to a positive wage shock. To find this critical value, we can solve for it assuming that agents with this critical value would be exactly indifferent between locating at the notch point or beyond, and equating the utilities. Assuming log utility in consumption allows for a closed-form solution:

$$\log(c_{NP}) - \psi \frac{h_{NP}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} = \log(c_{BN}) - \psi \frac{h_{BN}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} \tag{A.26}$$

where subscript-NP variables correspond to choice variables at the notch point, and subscript-BN variables correspond to those beyond the notch point. These agents will be subject to a budget constraints that contains the critical value a^*:

$$c_i = a^* h_i w - T\left(a^* h_i w\right) \text{ where } i = \{NP, BN\} \tag{A.27}$$

For simplicity, I assume one tax rate τ and and benefit of size d that the agent beyond the notch point has no access to. Plugging these and the budget constraints into their respective utility functions and solving for the critical value a^* yields:

$$a^* = \frac{d}{h_{BN} w \left(e^{\psi \frac{h_{NP}^{1+\frac{1}{\epsilon}} - h_{BN}^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}} - \frac{h_{NP}}{h_{BN}} \right)} + \tau \tag{A.28}$$

Note that the critical value a^* is increasing in the size of the benefits cliff d and tax rate τ. These both increase the region of dominated abilities: a greater benefit via a greater loss in consumption after the cliff, and higher marginal tax rates via tilting the budget constraint of Figure 1.2 clockwise. The denominator of equation A.28 further reveals that as the ratio of the hours choice at and beyond the notch point, h_{NP}/h_{BN} approaches one, the critical value explodes. Hence, the critical value a^* increases in relevance only for large potential jumps in hours. As mentioned in Appendix A.1.6, effective marginal tax rates approaching benefits cliffs already tend to be high. The wage increases in my simulation are also relatively small, with shock sizes not exceeding 3% annually. These assumptions work to increase the critical value a^*, thereby decreasing the number of agents who will jump in response to wage shocks.[3] My model abstracts away from these jumpers entirely, thereby increasing the detrimental effects of benefits cliffs towards my goal of delivering an upper bound of their effects.

Nevertheless, what would the aggregate implications of abstracting away from these jumpers be? Observe, for example, the change in aggregate consumption in response to a wage increase. For a model with a single cliff at pre-tax/transfer income z, a shock given by x_t, household income as $w_{i,t} = a_i w_t h_{i,t}$, the income density f, and a dominated region whose income ends at z^+, then aggregate consumption is

$$C_t = \int_0^z c\left(w_i, x_t\right) f\left(w_i, x_t\right) dw_i + c\left(z, x\right) f\left(w_z, x_t\right) + \int_{z^+}^{\infty} c\left(w_i, x_t\right) f(w_i, x_t) dw_i \tag{A.29}$$

Where the first term is consumption before the cliff, the second term is consumption at the cliff, and the third term is consumption after the dominated region of the cliff. $c(\cdot)$ is the consumption function for individual agents. The assumption of "no jumping" constrains the densities $f(\cdot)$ to be constant, whereas consumption is free to move in response to a wage shock. For small densities at the notch $f\left(w_z, x_t\right)$ the aggregate implications are small. In reality, a fraction of the density at the cliff $f\left(w_z, x_t\right)$ will increase their consumption from $c\left(z, x\right)$ to at least $c\left(w_i, x_t\right)$.

[3] As a back-of-the-envelope calculation, assume my model's baseline calibrated parameter values, an EMTR of 50%, and a benefits cliff of $1,000. About 98% of the population has estimated idiosyncratic abilities below the critical value $a^* = 30.537$ for an agent indifferent between working 1900 and 2100 hours annually - even more if the ratio of hours is closer to 1.

Finally, while constructing this model to be solvable with perturbation techniques - as opposed to value function iteration and simulation - significantly improves computational speed, the resulting model is linear. This would imply that for a negative productivity shock, findings would simply be negated. In this case, the model likely assumes an unreasonable increase in working hours for those at cliffs, as they would target the cliff to maintain their maximum benefits. While this may be theoretically accurate, in reality many may simply maintain their working hours. Thus, the model of this paper is not as suitable for analyzing negative aggregate productivity shocks.

A.3 Modeling Appendix

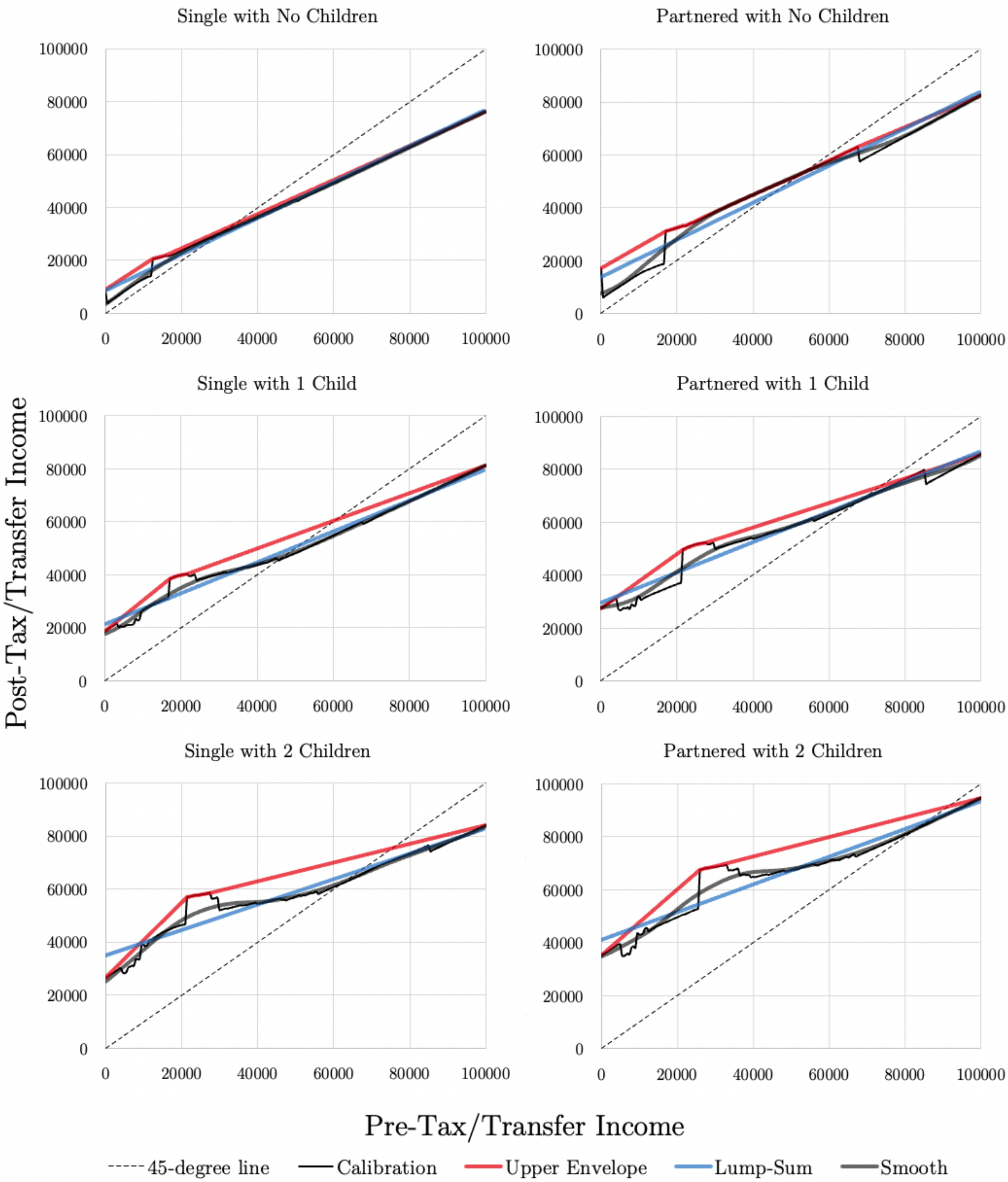
A.3.1 Additional Counterfactuals

A Linear Income Tax and Demogrant Description: This counterfactual - referred to as the "lump-sum counterfactual," replaces the entire tax-and-transfer schedule with a linear income tax and demogrant. Specifically, I fit a linear regression to each separate tax-and-benefit schedule: $y_{ij} = \alpha + \beta z_{ij} + \epsilon_t$. Here, y_{ij} are the statutory post-tax/transfer incomes, and z_{ij} are pre-tax/transfer incomes. The resulting constant α is the demogrant, with the resulting coefficient β acting as the net-of-tax rate. The demogrants range from $8,311.01 for singles with no children to $40,835.23 for those married filing jointly with two children. Effective marginal tax rates range from 30.00% for those married with no kids to 47.05% for those married with two children.

Smoothed Cliffs with No Benefit Loss Description: The previous counterfactuals above result in benefit losses for some households. This counterfactual - referred to as the "upper envelope counterfactual", replaces all benefits so that the new tax-benefit schedule fits at least the upper envelope of the baseline schedule. However, a pure upper envelope of the tax-benefit schedules would still leave cliffs in place, and/or regions with 100% effective marginal tax rates. Thus, I set benefits so that they phase out linearly from their statutory peak to their statutory level at $100,000 of income, separately for each household type j.

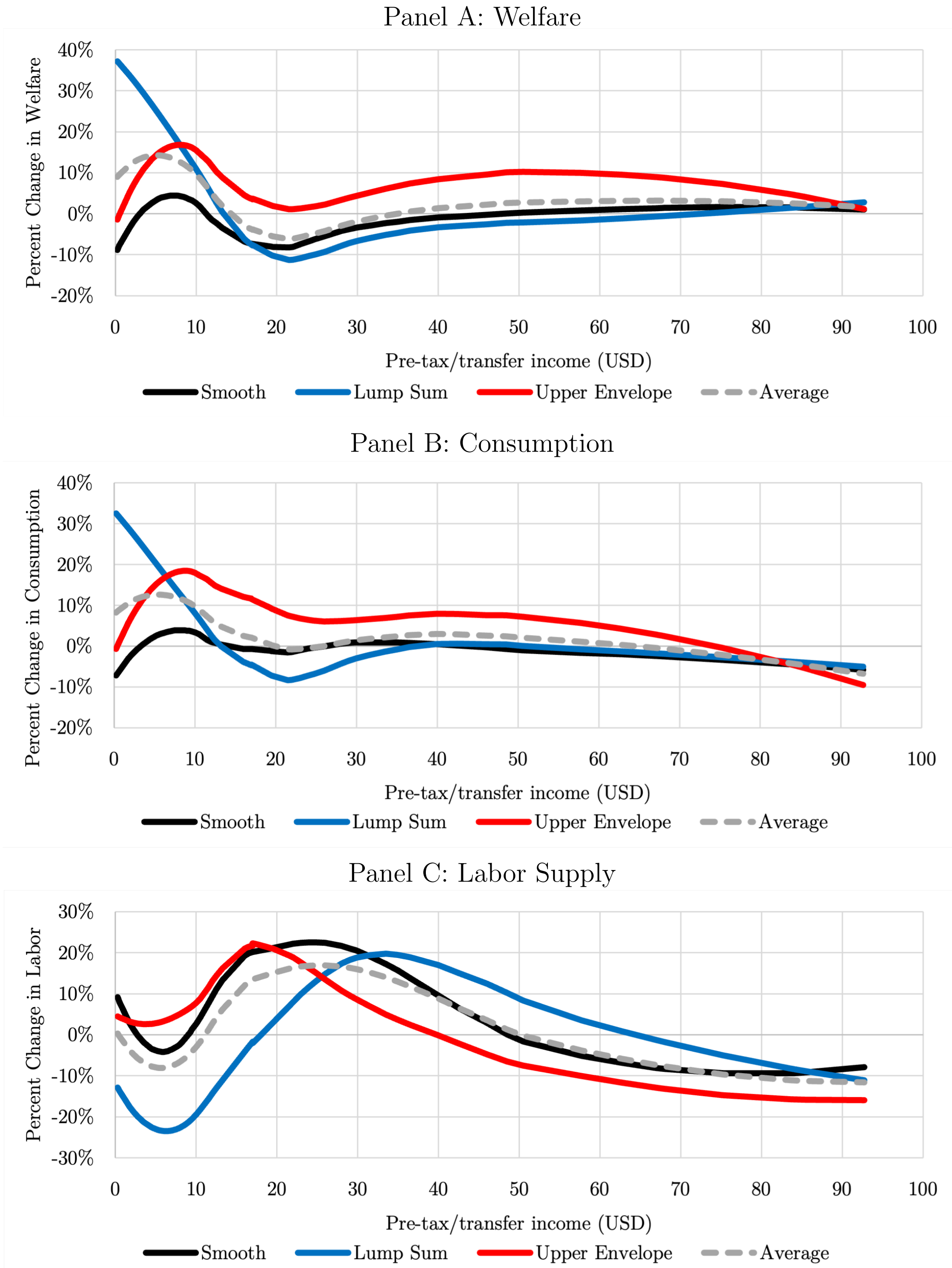
Figure A.7 presents a comparison of the counterfactual tax/benefit schedules and the calibrated baseline schedules. Each panel is a tax-and-benefit schedule for each household type. Each x-axis is pre-tax/transfer earnings, and each y-axis is post-tax/transfer earnings. Black is for the calibrated baseline schedule, grey is for the smoothed counterfactual, blue is for the lump-sum counterfactual, red is for the upper-envelope counterfactual, and the dashed line is where pre-and-post-tax/transfer incomes equal one another. As mentioned above, both the smooth and lump-sum counterfactuals feature regions where post-tax/transfer income

Figure A.7: Counterfactual Illustrations by Family Structure



Note: For each of the subfigures, the axes remain the same, as indicated. Values are in 2020 USD.

Figure A.8: Percent Changes in Steady-State Values from Baseline, Distribution



Note: Lines are the result of fitting a LOWESS function to the outcomes for individual households. The simple average of outcomes across the three counterfactuals is plotted to showcase general trends.

drops relative to the calibrated baseline - in particular around the region where health insurance premium tax credits kick in for families. The upper-envelope counterfactual makes it so no household should see a drop in their post-tax/transfer earnings, at least according to the schedule. Results for these alternative counterfactuals are presented in Figure A.8 and Table A.4.

Lump-Sum Counterfactual Results: Here, due to the large disincentive effects from more generous benefits in some regions, output decreases by about 3.8% in the steady state. Nevertheless, output still increases for those in the \$20,00 to \$60,000 range. Aggregate welfare in this counterfactual steady-state increases by over 6%, with slightly larger improvements for those formerly subjugated to cliffs. Changes in steady-state welfare broadly fall in line with that of the smooth counterfactual across the income distribution, except for those with earned income less than \$10,000, as seen in Figure A.8, Panel A. Here, the generosity of the lump-sum payment increases consumption and lowers labor supply to a much greater extent than the smooth counterfactual, so that the poorest agents have steady-state welfare gains exceeding 30%. This counterfactual still has the effect of reducing benefits for families in the region of premium tax credits, increasing labor supply and reducing welfare the most for those in this region.

In response to an aggregate 1% TFP shock, output increases by over 2% more in this counterfactual compared to the baseline. However, the positive response in welfare is smaller under this counterfactual, by about 1.7%. This is primarily driven by diminishing marginal utility: with an increase in welfare in the steady-state, additional changes in consumption and labor supply from a shock matter less. Still, despite the decline in aggregate welfare's response to the shock, those formerly subjugated to cliffs have their welfare response increase over 150% compared to baseline.

Upper Envelope Counterfactual Results: This counterfactual avoids reducing benefits at any point along the tax/benefit schedule while eliminating all benefits cliffs. However, doing so comes are a considerably high cost: an increase of nearly 50% in the lump-sum payment that must be paid to close the government's budget constraint. Due to the generosity of benefits, output losses and welfare gains are largest for this counterfactual steady-state: -4.6% and 14.3% respectively. Output still increases for those in the \$20,000-\$40,000 range. In contrast to the other counterfactuals, welfare increases across nearly the entire income distribution, as shown in Figure A.8, Panel A. Steady-state welfare gains are still the largest for those earning less than \$12,000 - up to just over 15% more. The generosity of benefits reduces the labor supply for higher-income households the most here among the counterfactuals.

Under the upper envelope counterfactual, output increases over 3.5% more on impact in response to a 1% TFP shock relative to baseline. Welfare increases 9.43% less on impact in

Table A.4: Additional Counterfactual Results

Panel A: Steady-State Output

Counterfactual	Aggregate	By income				
		\$0-\$20k	\$20k-\$40k	\$40k-\$60k	\$60k-\$80k	\$80k-\$100k
Lump-Sum	-3.801%	-15.324%	13.783%	5.353%	-0.406%	-5.177%
Upper Envelope	-4.615%	-5.020%	4.542%	-2.450%	-2.877%	-10.543%

Panel B: Steady-State Welfare

Counterfactual	Aggregate	formerly on cliffs	Never on cliffs	Point difference	%Δ in Lump-Sum Payment
Lump-Sum	6.235%	7.043%	6.219%	0.823	2.0%
Upper Envelope	14.286%	16.877%	14.237%	2.640	49.1%

Panel C: Response to TFP shock

Counterfactual	Output	Welfare			
		Aggregate	Formerly on cliffs	Never on cliffs	Point Difference
Lump-Sum	2.041%	-1.716%	156.875%	-4.746%	161.620
Upper Envelope	3.560%	-9.434%	149.282%	-12.466%	161.748

Note: Values are in percent changes from the baseline to counterfactual models. This includes Panel C, which is the percent improvement of the response of output under the counterfactual model compared to the same response in the baseline. Panel A shows changes in output in aggregate and across incomes, and panel B shows changes in consumption-equivalent welfare.

comparison to the same shock in the baseline model, due to declining marginal utility and the substantial welfare improvement already inherent in the new steady-state. Still, welfare improves nearly 150% for those formerly on cliffs.

A.3.2 Alternative Assumptions on Parameters

No debt-holding costs: To test the sensitivity of my model to the assumption of debt-holding costs, I rerun the smooth counterfactual and set the parameter that governs these costs, μ, to zero. The results for the smooth counterfactual without debt-holding costs is given in Table A.5. Removing these costs does nothing to affect steady-state values, because these costs only apply when bond-holdings deviate from its steady-state values. However, it does slightly attenuate the percent improvement in output on impact, which falls 2.4% from its original value. Percent changes in welfare in response to a TFP shock are more sensitive: the aggregate welfare response falls 2 percentage points, representing a 9.8% and 37.9% drop in the additional welfare gained from a shock under the counterfactual compared to their original values, for those previously at cliffs and not, respectively.

Richer Frisch Elasticities: In the baseline model, all agents are assumed to have the same Frisch elasticity of labor supply. However, empirical estimates of the Frisch elasticity often differ depending on the number of children and cohabitation status. In their literature review, Reichling and Whalen (2012) notes that women with children have a higher Frisch elasticity than those without - up to 75% greater (Blundell et al., 1993). Dual-income households also have a smaller elasticity than single earners - around 40% higher (Kimball and Shapiro, 2008). With this evidence in mind, I recalibrate the Frisch elasticities. Singles with 0-2 children now have Frisch elasticities of 0.256, 0.352, and 0.448, respectively, and couples with 0-2 children have Frisch elasticities of 0.359, 0.455, and 0.551. The results for the smooth counterfactual with new Frisch elasticities are given in Table A.5, in comparison to a baseline model with cliffs and updated elasticities. The results are largely similar, save for one exception: the improvement in the welfare response for those formerly at cliffs in reaction to a 1% TFP shock increases by over 70% to nearly 313%. This is because many households - particularly those with multiple earners and children - have higher Frisch elasticities than in the original smooth counterfactual model, considerably increasing their responsiveness to wage gains when moving from the constrained to unconstrained scenario.

Allowing for government debt: In the baseline model, the government's budget constraint does not allow for the accumulation of any public debt. Here, I relax this assumption by augmenting this constraint to:

$$\sum_{i=1}^{I} \omega_i T_j \left(a_i h_{i,t} w_t\right) + D_{t+1} = \sum_{i=1}^{I} \omega_i B_t + (1 + r_t) D_t \tag{A.30}$$

where D_t is the stock of government debt. That is, issuing new debt is now a form of financing for the government, though it must be paid back with interest r_t. I subject the government budget constraint to a fiscal feedback rule:

$$B_t = B_{t-1} - \chi(D_t - \bar{D}) \tag{A.31}$$

where $\bar{D}$ is the target debt. Hence, the lump-sum payment B_t is adjusted when we deviate from the target stock of debt, with the strength of the adjustment governed by χ. Including a stock of government debt essentially amounts to numerical error in changes in the responsiveness of aggregate output on impact in response to a TFP shock. This is due to both the baseline and counterfactual economies are subject to the same fiscal feedback rule. Improvements in welfare from a TFP shock are lessened, but this is simply the result of dis-

Table A.5: Percent Changes Steady-State Output from Baseline, Extensions

Panel A: Steady-State Output

Aggregate	By Income				
	\$0-\$20k	\$20k-\$40k	\$40k-\$60k	\$60k-\$80k	\$80k-\$100k
	Richer Frisch Elasticities				
1.207%	2.848%	19.128%	-5.242%	-1.575%	-3.186%
	No Debt-holding Costs				
1.448%	3.787%	15.368%	-4.002%	-1.312%	-2.832%

Panel B: Steady-State Welfare

Aggregate	Formerly on cliffs	Never on cliffs	Point difference	%Δ in Lump-Sum Payment
	Richer Frisch Elasticities			
-4.572%	-1.785%	-4.626%	284.098%	-8.532%
	No Debt-holding Costs			
-4.582%	-2.179%	-4.628%	244.883%	-8.829%

Panel C: Response to TFP shock

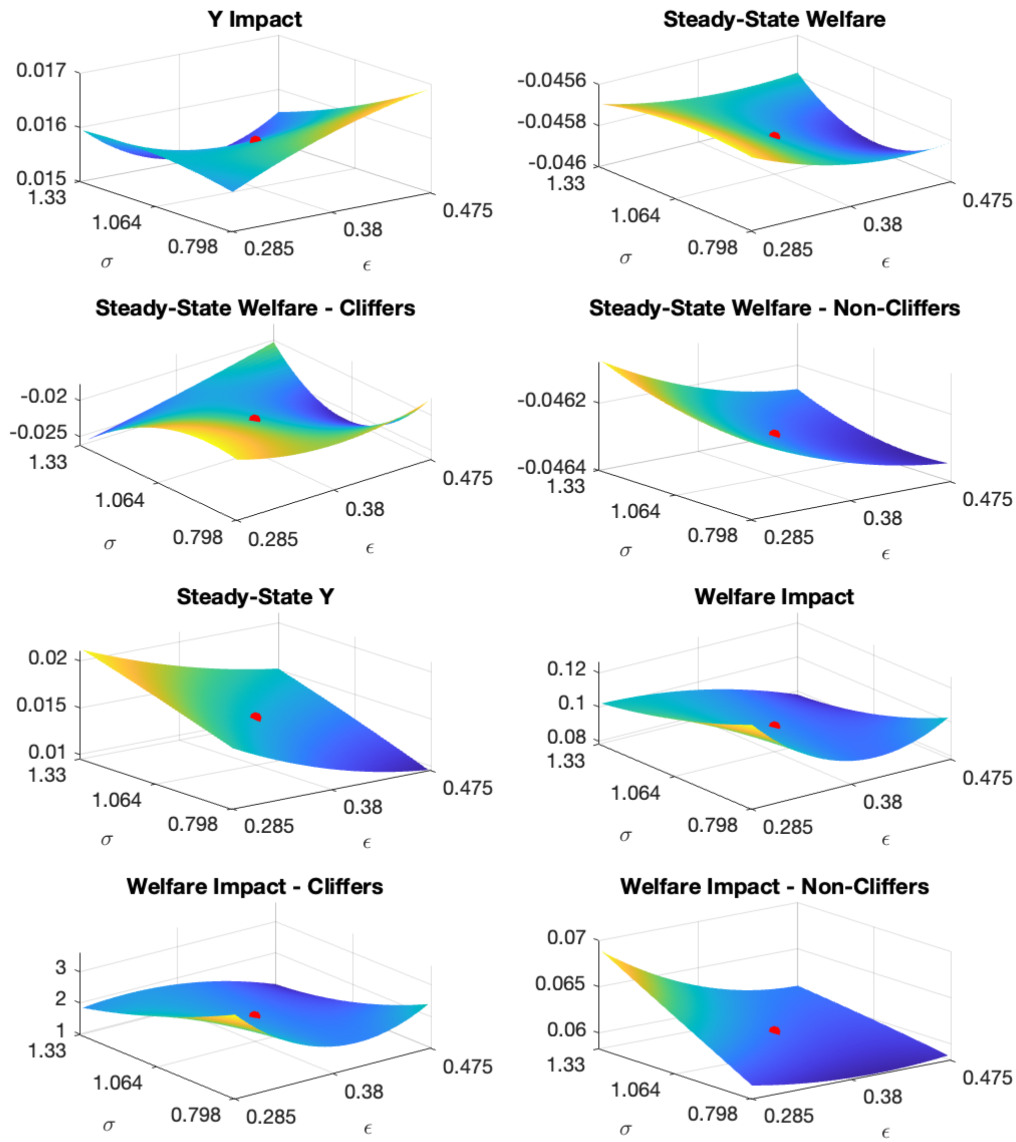
Output	Welfare			
	Aggregate	Formerly on cliffs	Never on cliffs	Point Difference
	Richer Frisch Elasticities			
1.490%	11.617%	312.982%	5.859%	307.123
	No Debt-holding Costs			
1.545%	7.414%	165.561%	4.392%	161.169

Note: Values are in percent changes from the baseline to counterfactual models. This includes Panel C, which is the percent improvement of the response of output under the counterfactual model compared to the same response in the baseline. Panel A shows changes in output in aggregate and across incomes, and panel B shows changes in consumption-equivalent welfare.

counting. Rather than immediately redistributing additional revenue from a positive TFP shock, the government with a fiscal feedback rule does so with a delay. Essentially, a fraction of the gains from a productivity shock are shifted forward in time. Even if these gains are nearly the same in absolute value between models with a fiscal feedback rule and without, agents discount delayed payments more. Moreover, the presence of HTM households means many lack the ability to save. Hence, for essentially the same aggregate outcomes, welfare will improve less with a fiscal feedback rule. Perhaps unsurprisingly, the more quickly the government distributes gains from positive productivity shocks, the more welfare improves in response to a shock.

Parameter Sensitivity: In order to see if my results are particularly sensitive to my calibration of preferences, I calibrate and run the model over a range of values for the Frisch elasticity ϵ and elasticity of intertemporal substitution σ, separately and in combination. I do this ranging from values 25% larger or smaller for each. Visual evidence of the effects of changing these parameters are in the various panels of Figure A.9 for the smooth counterfactual. These plot the percent differences in the steady-state values and impact response of output and welfare, disaggregated between those formerly on cliffs and not. The impact response on output is very stable - its standard deviation is 0.052% from its baseline value. Changes in overall steady-state welfare and the steady-state welfare of those never on cliffs in the first place are even less affected. The standard deviation of changes in steady-state output is 0.362%, and even less so for the steady-state change in welfare for those formerly on cliffs. Parameter values matter most for the change in the impact response in welfare. The standard deviation of changes in the aggregate welfare response on impact of a TFP shock is 1.424%. More strikingly, the standard deviation in the welfare response for those formerly at cliffs in reaction to a shock is just over 75%. Thus, while aggregates are relatively insensitive to my parameter assumptions, the welfare gain for those formerly at cliffs in the counterfactual is not. Values for it range from 97% to 361.76%, so the effect of removing cliffs is always at least nearly a doubling of the welfare response to a shock.

Figure A.9: Parameter Sensitivity



Note: Red dots indicate the outcome for the baseline model for each indicated variables. For brevity, I label outcomes for those formerly at cliffs as "cliffers" and those never at cliffs as "non-cliffers."

APPENDIX B

Appendix for “Place-Based Policy and Optimal Income Transfers in a Federalist Framework”

B.1 A Database of State GMIs

This dataset of guaranteed minimum incomes was constructed to calibrate the federal and state default tax schedules. The data for transfers and income tax rates comes from the sources below.

On the federal level, I set the actual guaranteed minimum income equal to the sum of average annual benefits from Temporary Assistance for Needy Families (TANF), Women, Infant and Children (WIC), Supplemental Nutritional Assistance Program (SNAP), and unemployment benefits. I select these programs only, rather than include additional programs like Medicaid and other in-kind transfers, because they come closest to a pure cash-assistance program. As for the unemployment benefits, since in most cases they last 26 weeks and are funded in-part by the states, I compute an annualized measure (as if half of the benefits are paid for each 6-month segment) for the fraction that comes from federal funding. In addition, because only the “unemployed by definition” are eligible for unemployment benefits (those who do not have a job but are looking for work), and not all those who have no income, I weight the measure by the fraction of the density with zero income who meet the official definition of unemployed in March 2018. This generates an “average GMI” for those who have no income. The GMI for the US federal government is \$6,662, as seen in Table 2.1.

For the individual states, I calculate the average GMI using the same process above, but with state values for each program. I also weight the GMI by the fraction of current transfer programs financed by the states, for each state. This reveals that some states, such as Tennessee, are very reliant on the federal transfer system to supplement their own transfers. I present the table of GMIs that I calculated below in Table B.1.

Table B.1: GMIs for Calibrating Default Linear Schedules

Jurisdiction	GMI	Jurisdiction	GMI
Alabama	$ 942.16	Montana	$ 2,433.83
Alaska	$ 3,616.44	Nebraska	$ 1,880.74
Arizona	$ 552.52	Nevada	$ 1,829.30
Arkansas	$ 857.70	New Hampshire	$ 145.19
California	$ 5,368.37	New Jersey	$ 3,402.12
Colorado	$ 2,152.24	New Mexico	$ 202.90
Connecticut	$ 8,486.43	New York	$ 8,763.50
Delaware	$ 2,578.70	North Carolina	$ 1,592.26
D.C.	$ 5,027.48	North Dakota	$ 863.71
Florida	$ 1,446.76	Ohio	$ 1,176.27
Georgia	$ 1,219.93	Oklahoma	$ 915.17
Hawaii	$ 3,219.64	Oregon	$ 3,645.22
Idaho	$ 1,205.09	Pennsylvania	$ 1,872.65
Illinois	$ 2,529.88	Rhode Island	$ 2,069.66
Indiana	$ 1,525.60	South Carolina	$ 579.97
Iowa	$ 2,733.11	South Dakota	$ 2,490.77
Kansas	$ 1,796.45	Tennessee	$ 25.09
Kentucky	$ 1,094.56	Texas	$ 1,510.93
Louisiana	$ 746.65	Utah	$ 1,723.10
Maine	$ 1,782.26	Vermont	$ 2,281.85
Maryland	$ 3,775.96	Virginia	$ 2,351.48
Massachusetts	$ 5,577.24	Washington	$ 2,705.54
Michigan	$ 1,535.77	West Virginia	$ 1,329.42
Minnesota	$ 3,985.94	Wisconsin	$ 2,488.65
Mississippi	$ 433.98	Wyoming	$ 2,716.64
Missouri	$ 1,300.38	State Averages	$ 2,276.61

Dollar values are expressed in March 2018 dollars.

B.2 States Reacting to Optimal Federal Transfers

In the main body of the paper, I have states optimize taking a linear approximation of the current federal tax and transfer system as given. Here, I have states optimize taking the optimal federal transfer system, as seen in Table 2.2 and Figure 2.1, as given. The results are shown below in Table B.2 and Figure B.1, assuming that states take into account the mobility elasticity.

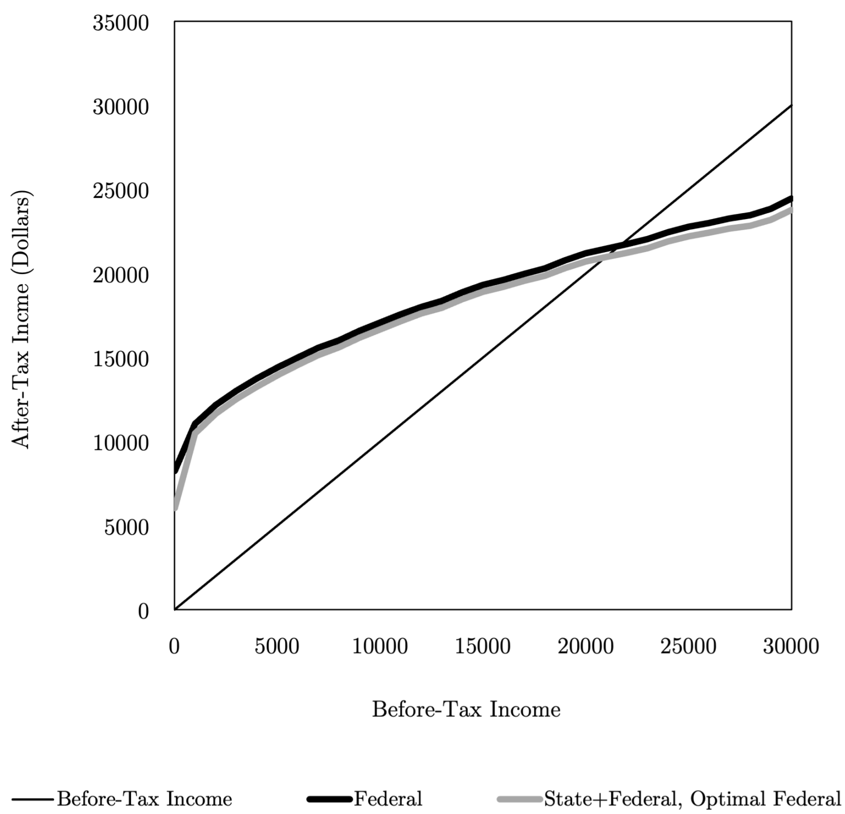
The results indicate that states would change very little from the optimal federal system, in contrast to the results in the main body of the paper. That is, states do not tax away the GMI and peak EITC to the same extent as above. Hence if the federal government were to implement the optimal income transfer system, they need not be too concerned about states taxing away substantial amounts of transfers, as states already find them mostly optimal. However, this result only holds on average, and individual states may deviate substantially, still necessitating a policy that federal transfers be untaxable should the federal government wish to promote consumption of those with no or low incomes.

Table B.2: Average State Reacting to Optimal Federal Transfers

GMI:	$6,053
MTR for phase-in:	180%
Peak EITC:	$9,644
Pre-tax Income at Peak*:	$2,000
MTR for phase-out:	52%
Break-even point*:	$21,000
Top State MTR above break-even point:	5%

Dollar values are expressed in March 2018 dollars. * indicates within $1000 of income.

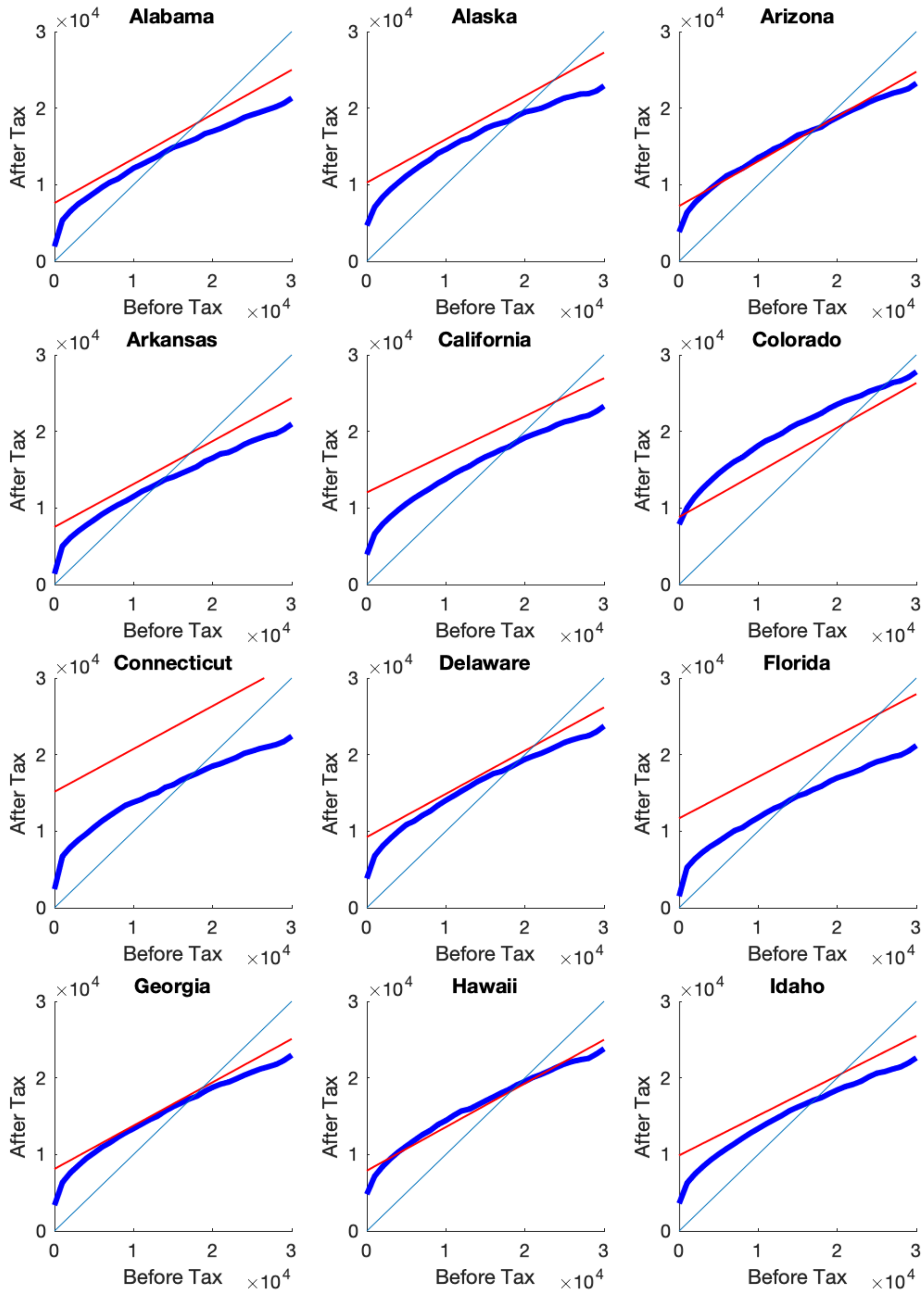
Figure B.1: Optimal State Transfers Under Optimal Federal Transfers

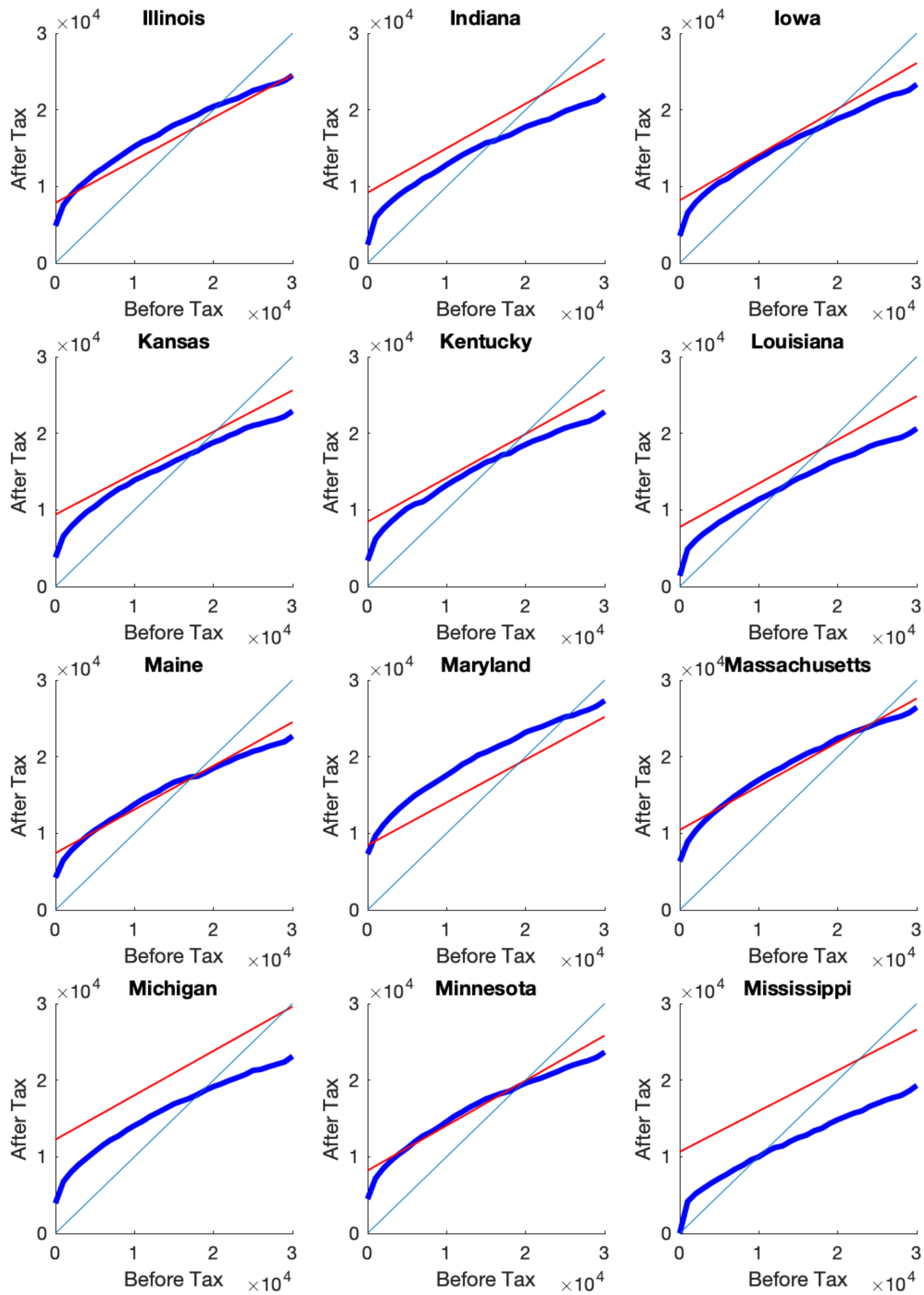


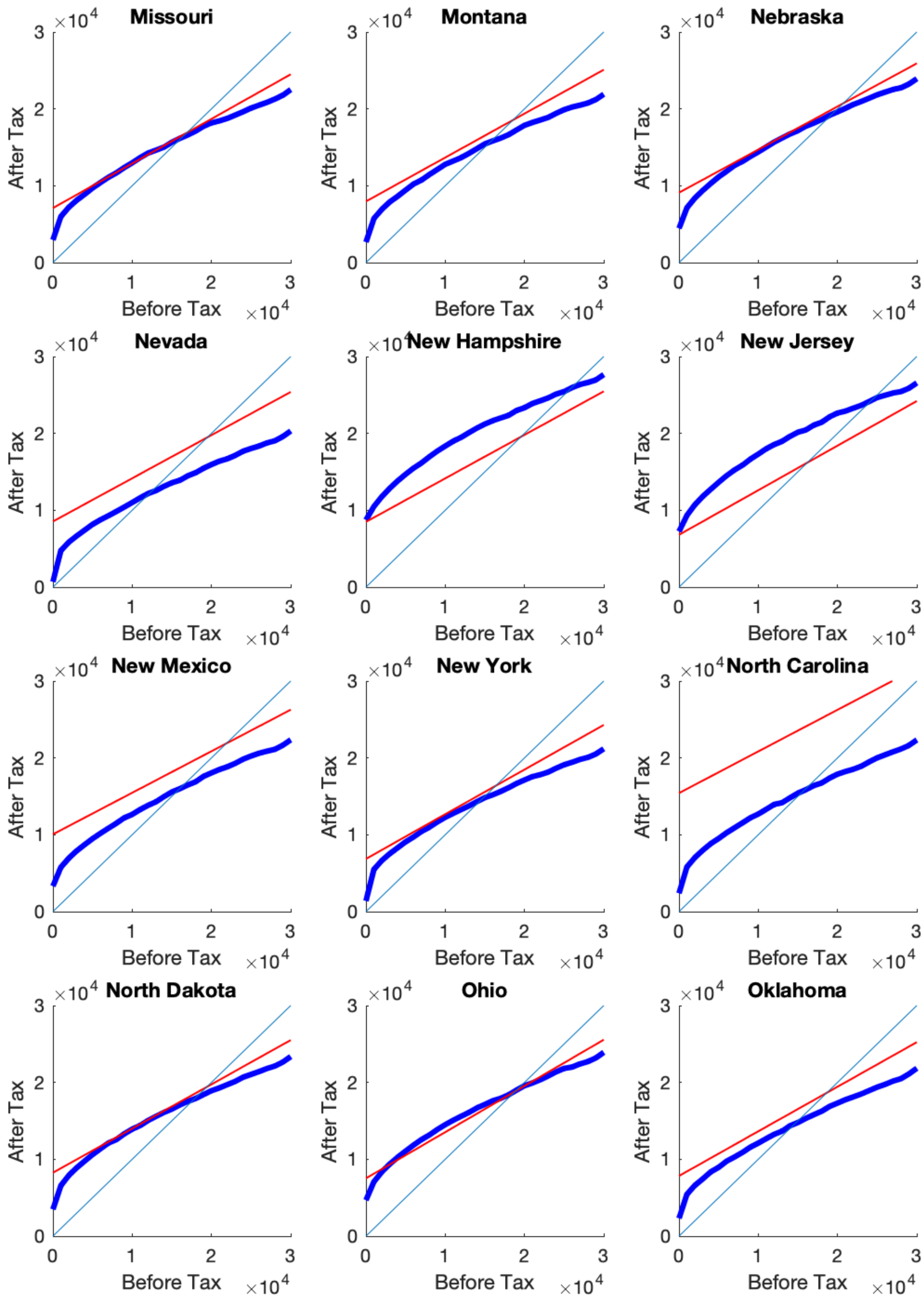
Note: Optimal income tax schedules, given in March 2018 dollars. The thin black line represents no net transfers or taxes, when before- and after-tax incomes are identical. The solid black line represents the optimal federal income tax schedule. The solid grey line represents optimal state tax schedules when the optimal federal income tax schedule is taken as given.

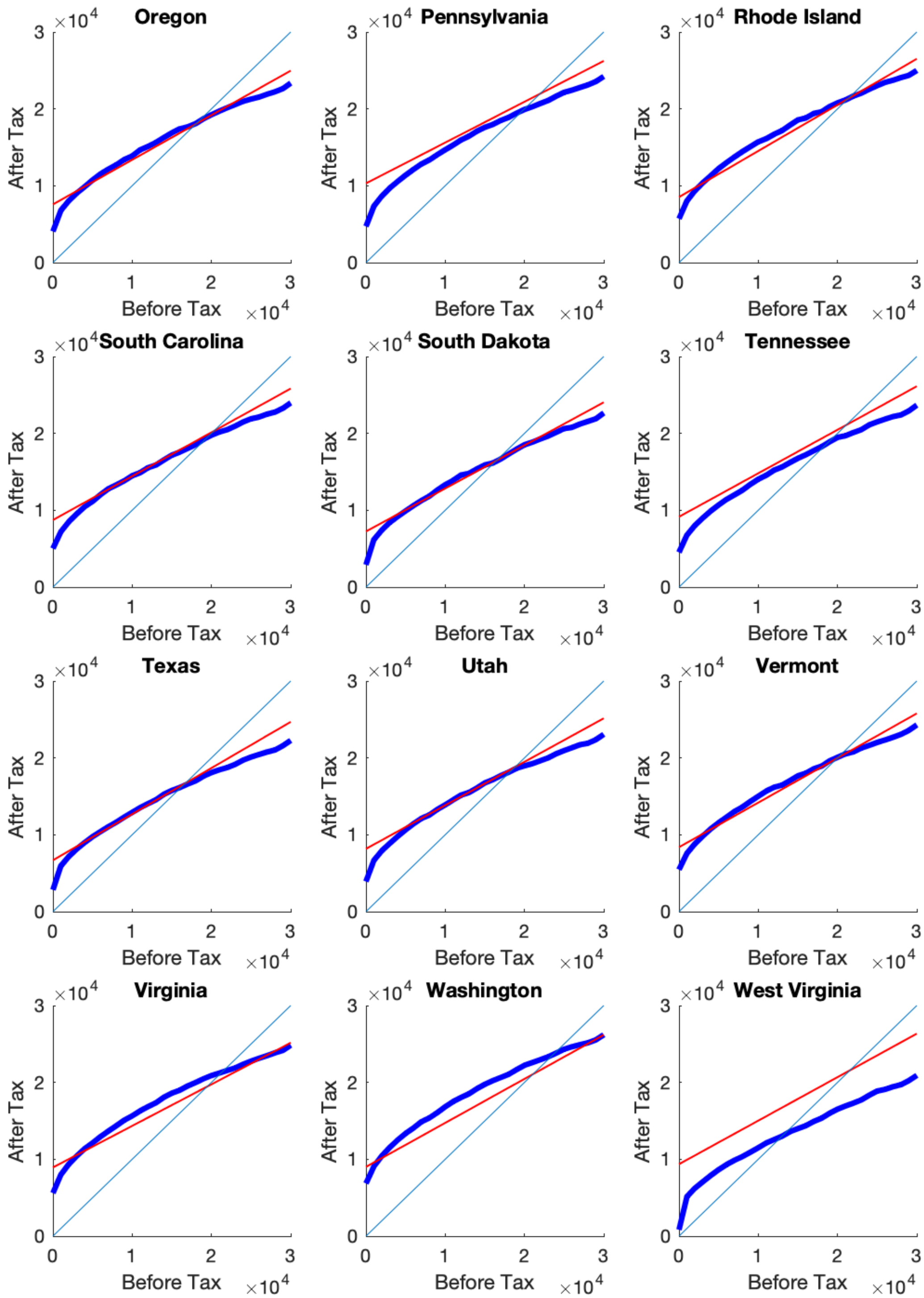
B.3 Additional Figures

Figure B.2: Optimal Transfers by State

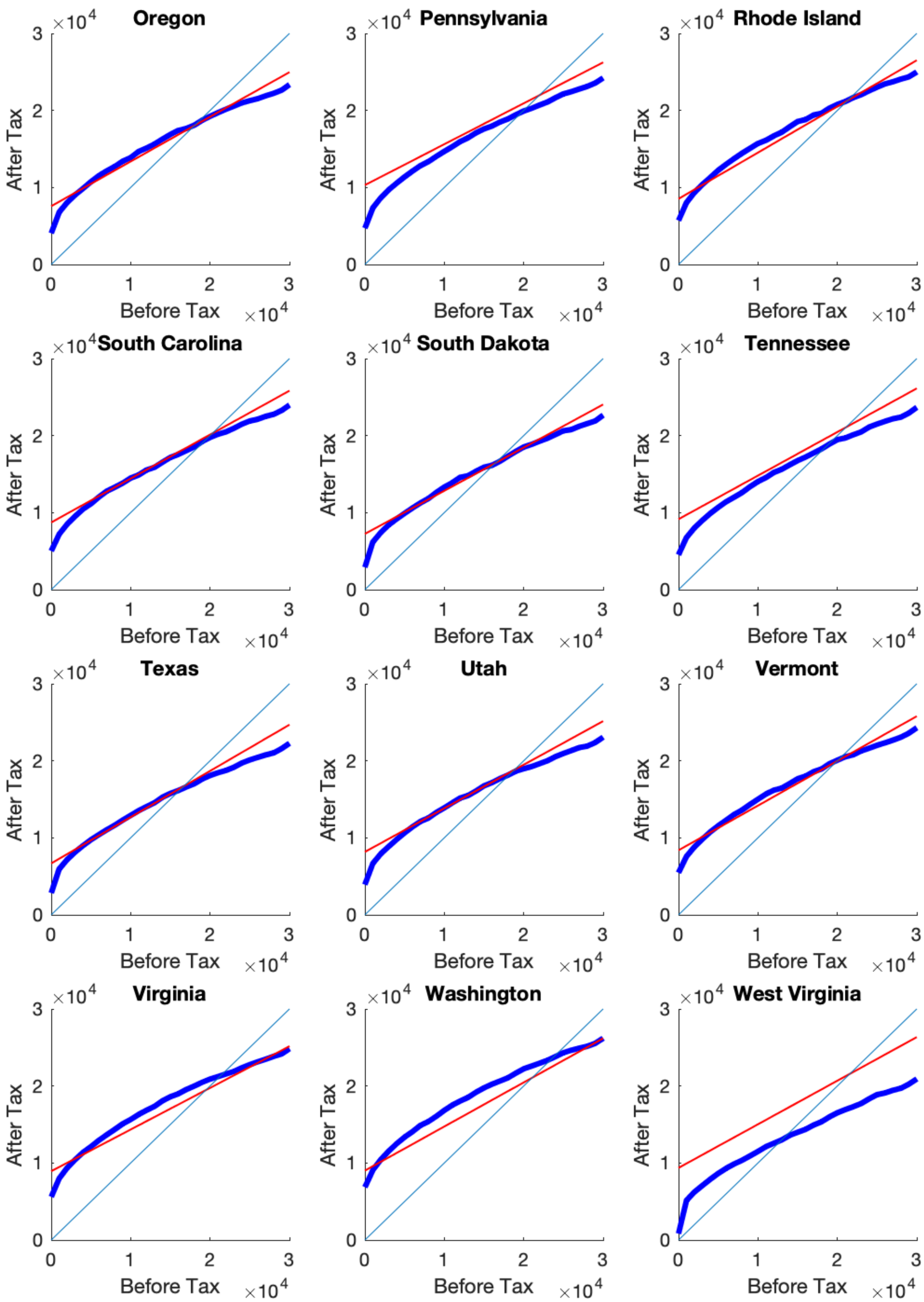



Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
After Tax
Before Tax
$\times 10^4$


Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
After Tax
Before Tax
×10^4
0
1
2
3

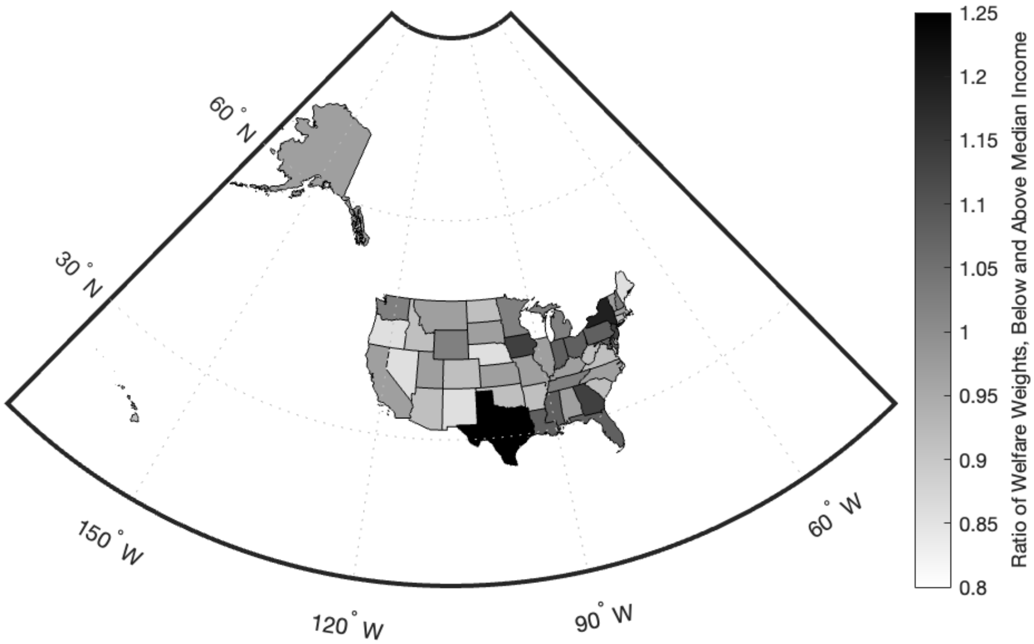


Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
After Tax
Before Tax
×10^4
0
1
2
3



Note: Optimal income tax schedules, given in March 2018 dollars. The thin blue lines represent no net transfers or taxes, when before- and after-tax incomes are identical. Then thin red lines are the default linear approximations of state and federal income tax systems combined. The thick blue line represents the optimal tax schedule.

Figure B.3: Welfare Weight Ratios by State



Note: Ratio of average welfare weights for incomes below to above median income in March 2018 dollars for US states. Darker shades indicate a higher ratio.

APPENDIX C

Appendix for "A Transparent Look At How Taxes Affect Growth: Evidence from Cross-Country Panel Data"

C.1 Description of the Construction of the Tax Rate Data Series

Our tax rates data come from two different sources. In this Data Appendix, we document how we have reconciled discrepancies between the two data sources.

C.1.1 Tax Rates

We construct our top statutory CIT and PIT rates and VAT rates series using the OECD Tax Database and the American Enterprise Institute's (AEI) International Tax Database.[1]

C.1.1.1 Corporate Tax Rates

Table II.1 of the OECD Tax Database provides the central government and combined (central plus sub-central government) top statutory corporate income tax (CIT) rates.[2] This table provides CIT rates for 34 countries between 2000 and 2017.[3] Historical data are available for 1981-1999, but the OECD notes that these data have not been verified in recent years. Nevertheless, we use these data when available.

[1]We thank Aparna Mathur for sharing this database with us.

[2]These data are available at http://www.oecd.org/tax/tax-policy/tax-database.htm#C_CorporateCaptial.

[3]These countries are Australia, Austria, Belgium, Canada, Chile, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Latvia, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom, and the United States.

The AEI's International Tax Database contain CIT schedules at the central government and the local government level. These data are available for 153 countries between 1981 and 2011. The database references several underlying source materials: (1) PriceWaterhouseCooper's *Corporate Taxes-Worldwide Summaries*; (2) Coopers and Lybrand's *International Tax Summaries*; (3) Ernst and Young's *Worldwide Corporate Tax Guide 2001*; (4) the International Bureau of Fiscal Documentation Loose-leaf Service; (5) Embassies and ministries of taxation in individual countries; (6) and KPMG's *Worldwide Corporate Tax Tables*.

Figure C.1 depicts the central government CIT rate series for the 34 countries that appear in both the OECD and AEI databases. These two data sources agree in a majority of country-year observations. The non-trivial differences occur for the following countries: Estonia, Italy, Norway, and Portugal. We turn to external data sources to reconcile these differences.

- Estonia: Data from the Republic of Estonia Tax and Customs Board match the OECD data. Data available at: https://www.emta.ee/eng/business-client/income-expenses-supply-profits/tax-rates.

- Italy: Table 1 in Caiumi and Di Biagio (2015) provides information on changes to the general corporate income tax rate (initially called IRPEG, and then renamed to IRES in 2004). These rates are consistent with the AEI tax rate series. In addition, the data in the University of Michigan's World Tax Database (WTD), when available, is consistent with the AEI series.

- Norway: The WTD data are consistent with the AEI data. We were unable to locate a primary data source from the Norwegian government.

- Portugal: The WTD data are consistent with the AEI series between 1982 and 1999. Two other sources provide evidence that the AEI series contains the central government rate. First, an *International Tax Review* article reports that for 2015, there was a reduction in the standard corporate income tax rate from 23% to 21%.[4] Second, Figure 1 (page 10) of Bessa (2016) shows that the central government tax rate (taxa nominal) was 25%, 23%, and 21% in 2013, 2014, and 2015, respectively. These figures match the AEI data. The rates that add in state and municipal tax rates appear to match the OECD data.

We construct our top statutory CIT rate series using the OECD Tax Database as our primary source, and supplementing with the AEI International Tax Database, except where

[4] Article available at http://www.internationaltaxreview.com/Article/3421912/Portugal-Portuguese-corporate-tax-changes-for-2015.html.

indicated above. AEI data on CIT rates for Brazil, China and India end in 2011. We fill in these missing observations using KPMG's Worldwide Corporate Tax Tables.[5]

C.1.1.2 Personal Income Tax Rates

We construct a series of the top statutory central government PIT tax rate. Table I.1 of the OECD Tax Database includes the central government rates and thresholds for the PIT schedule. As with the CIT series, the OECD database contains data for 34 countries between 2000-2017, with historical data available from 1981-1999.

From the AEI International Tax Database, we use a spreadsheet of central government rates and thresholds for the PIT schedule. The database contains data for 150 countries between 1981 and 2012. The underlying sources for the AEI database are the same as for CIT rates, with the exception of the PriceWaterhouseCooper's *Individual Taxes-Worldwide Summaries*. AEI data on PIT rates for Brazil, China, India and South Korea end in 2011. We fill in these missing observations using KPMG's Worldwide Individual Tax Tables.[6]

Figure C.2 depicts the central government PIT rate series for the 34 countries that appear in both the OECD and AEI databases. These two data sources agree in a majority of country-year observations. The non-trivial differences occur for the following countries: Canada, France, Germany, Israel, Norway, Spain, and Sweden. We turn to external data sources to reconcile these differences.

- Canada: The Canadian Government's website provides historical information on General Income Tax and Benefit Packages in each year from 1985 through 2017. These documents include Schedule 1, which shows rates and income thresholds for the federal income tax schedule. We compile the top statutory PIT rate for 1985 through 2000. These data are consistent with the OECD series. Data available at: https://www.canada.ca/en/revenue-agency/services/forms-publications/tax-packages-years.html.

- France: The Institut des Politiques Publiques (2014) study, "1914-2014: One Hundred Years of Income Tax in France," provides information on the French PIT schedule in 1983, 1988, 1994, 2006, 2007, and 2014. The top statutory rate reported in these years match the OECD data. In addition, the publication provides a supplementary table of rates between 1981-2013, which match the OECD series.

[5] Available at https://home.kpmg/xx/en/home/services/tax/tax-tools-and-resources/tax-rates-online/corporate-tax-rates-table.html.

[6] Available at https://home.kpmg/xx/en/home/services/tax/tax-tools-and-resources/tax-rates-online/individual-income-tax-rates-table.html.

- Germany: The WTD data, available through 1999, are consistent with the OECD numbers. The Tax Policy Center (TPC) provides tables of top marginal PIT rates in OECD countries. This series is also consistent with the OECD series through 1999. Beginning in 2000, the TPC series is consistent with the AEI series but this is the same year in which the TPC changes its data source. In 2000, the TPC begins to use OECD Table I.7, "Top statutory personal income tax rate and top marginal tax rates for employees," which combines central government and sub-central government rates. To maintain a consistent definition throughout the panel, we use the OECD data.

- Israel: When available, the WTD and TPC series are consistent with the OECD data, so we use that series throughout.

- Norway: The WTD data are closest to the OECD countries in 1984-1996, when it becomes closer to the AEI series. The WTD data are well above both series in 1981 and 1982 (data missing in 1983). We use the OECD data.

- Poland: The observations in 1988 and 1989 appear to be outliers in the series, and are notably unavailable in the OECD Database. The PIT rate reported for these years might be the CIT rate, as the source notes in the AEI International Tax Database reference an "equalization tax." Kierzkowski et al. (1993) report that 1988 and 1989 were years of massive tax reform with many different rates depending on a number of factors, such as whether the employer was socialized or non-socialized, industry, etc. Beginning in 1990, a 40% top rate was uniformly applied. To maintain a consistent definition, we being the Poland series in 1990.

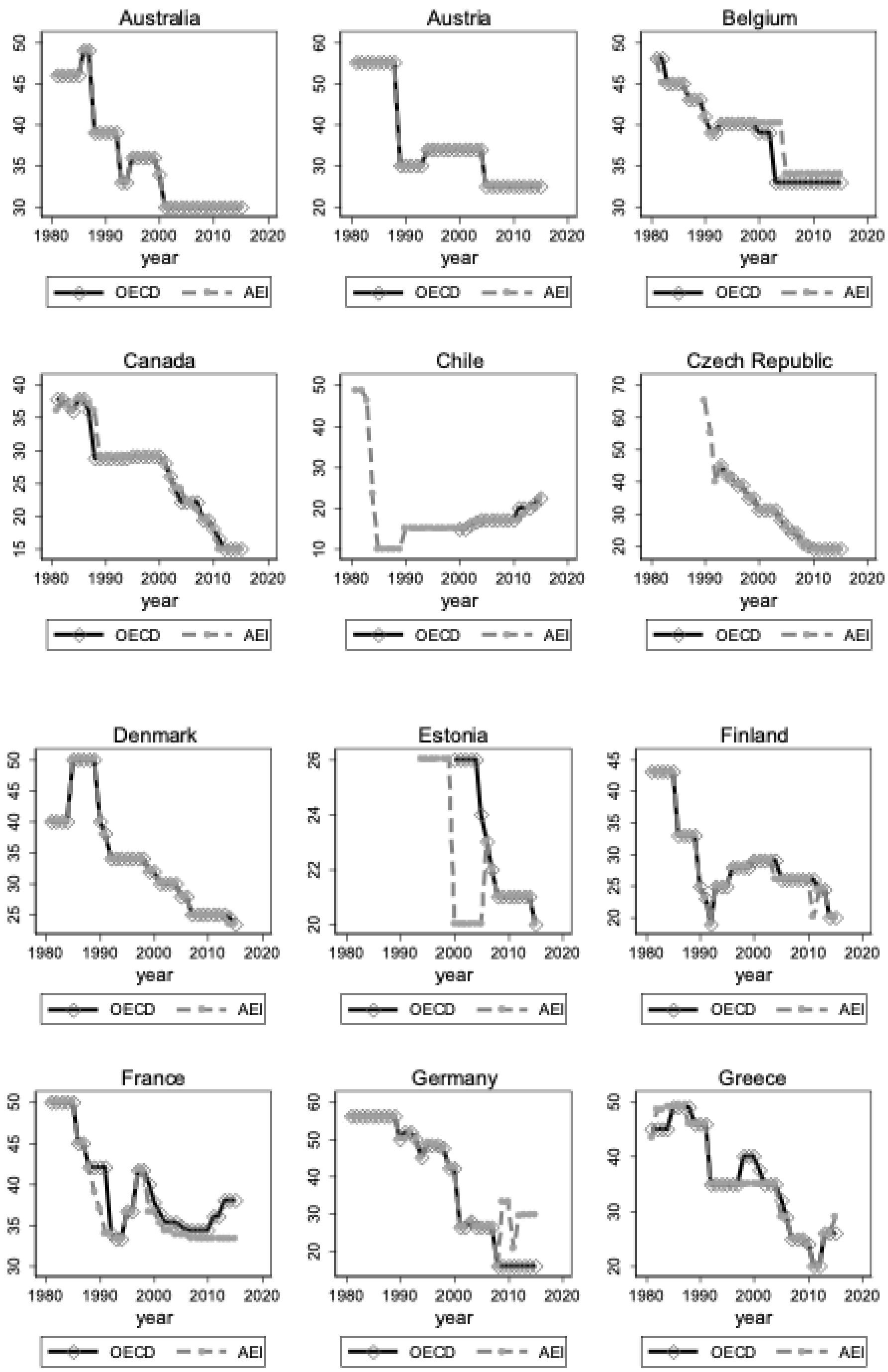
- Spain: The two series diverge in 2008 and 2009. We use the OECD data in these years.

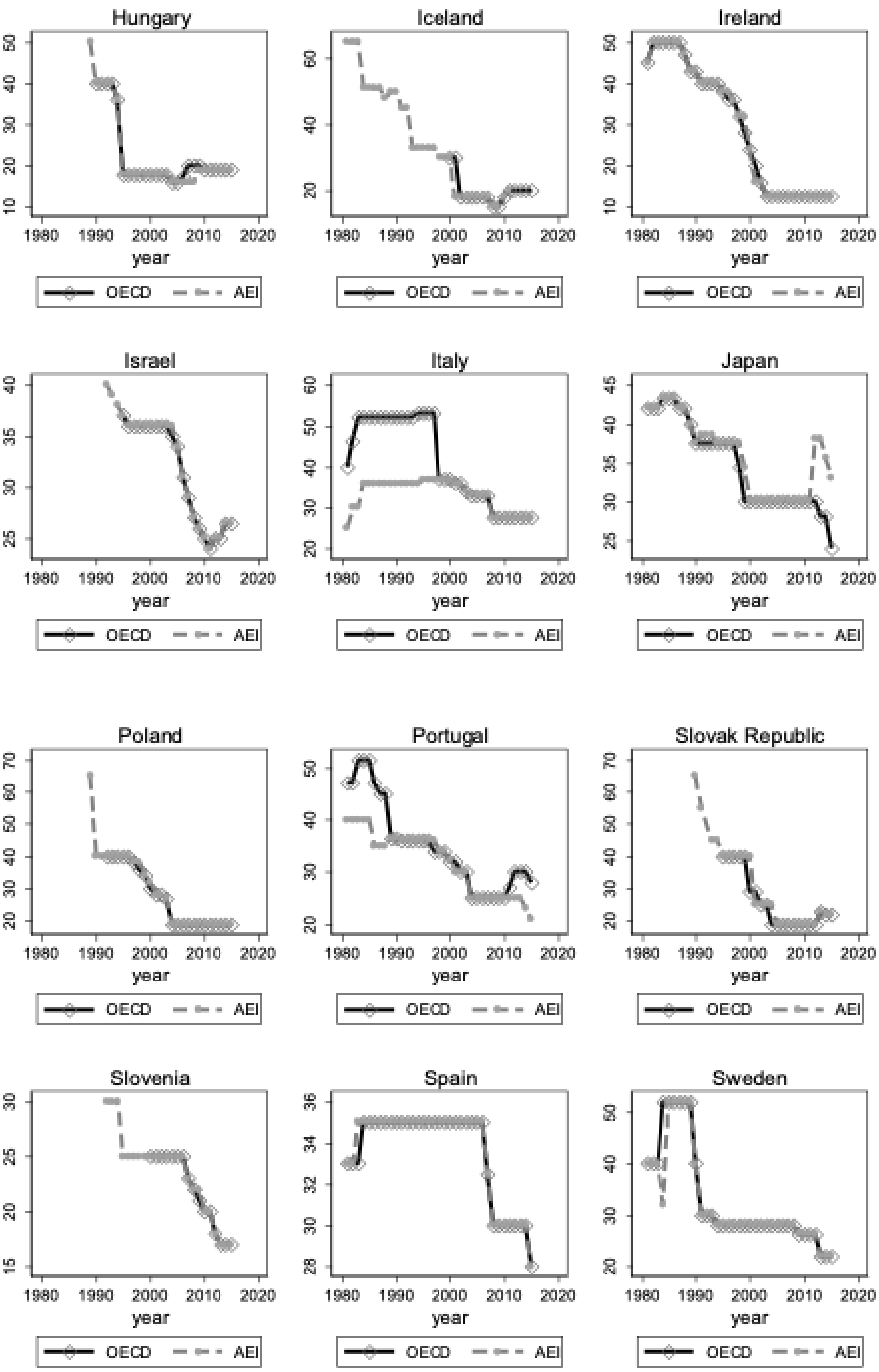
- Sweden: Figure 2 in Stenkula, Johansson and Rietz (2013) depicts the central government PIT rate (called the state rate, as opposed to municipalities). These data are consistent with the AEI series prior to 1991.

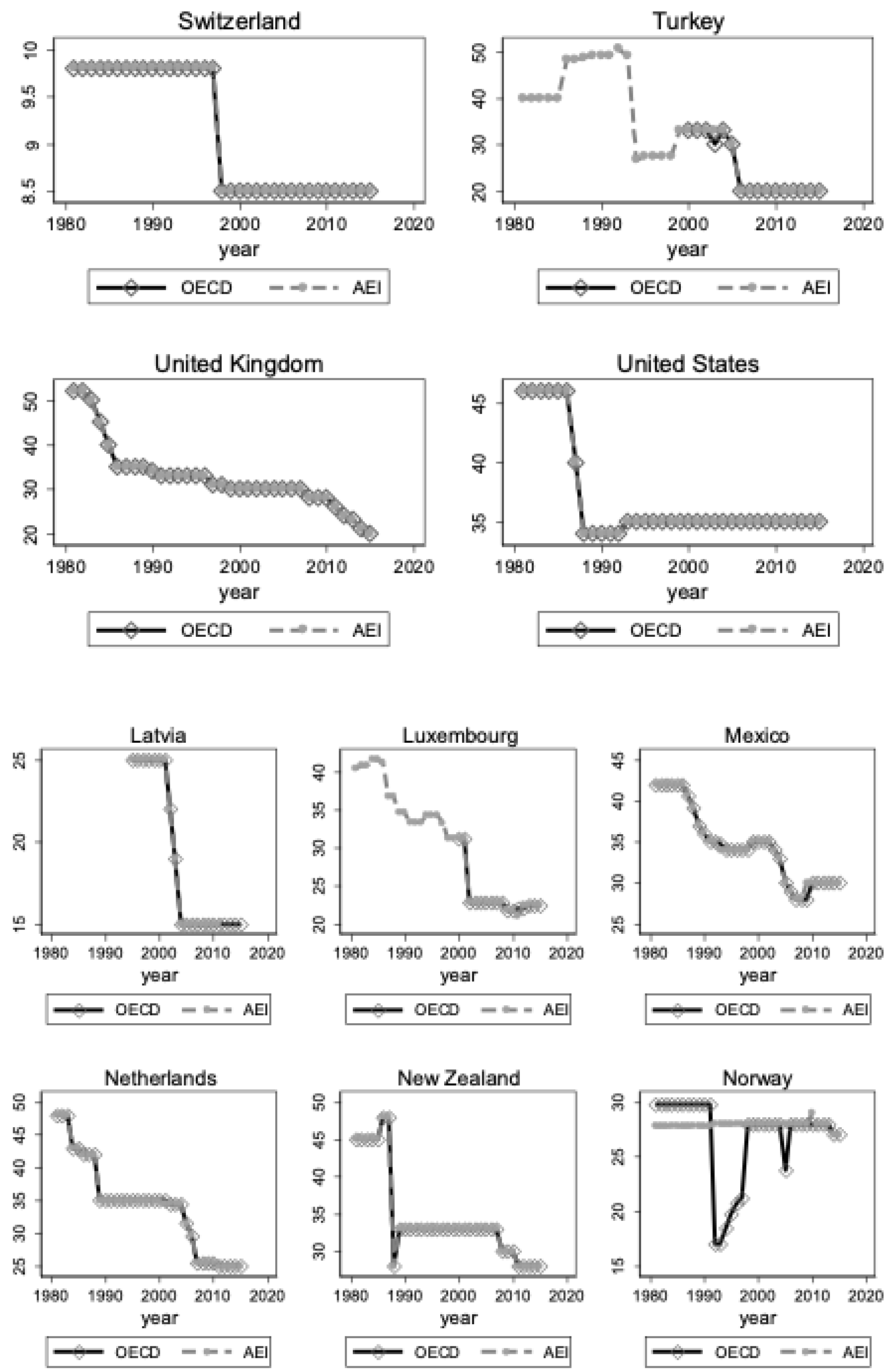
C.1.1.3 Value-Added Tax Rates

The OECD Tax Database contains data on 34 countries in 1985, 1990, 1995, 2000, and each year between 2005 and 2015. The AEI International Tax Database contains data on 147 countries between 1981 and 2008, and then in 2011. Figure C.3 depicts the central government VAT rate series for the 34 countries that appear in both the OECD and AEI databases. These two data sources agree in a majority of country-year observations, and there were no changes to the data made. We fill in the missing years of data using the AEI International Tax Database.

Figure C.1: Comparison of Corporate Income Tax Rates from the OECD Tax Database and the AEI International Tax Database

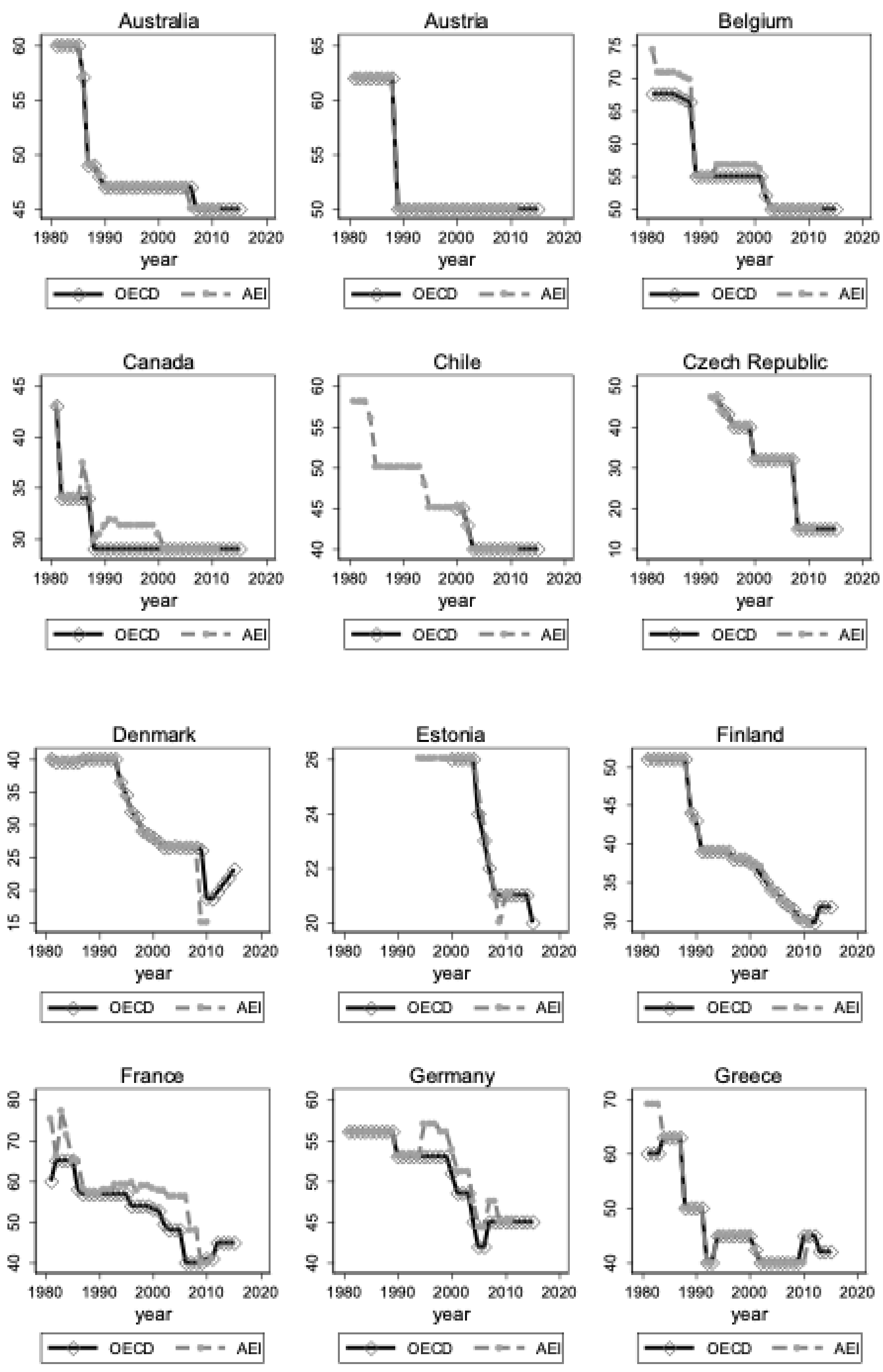


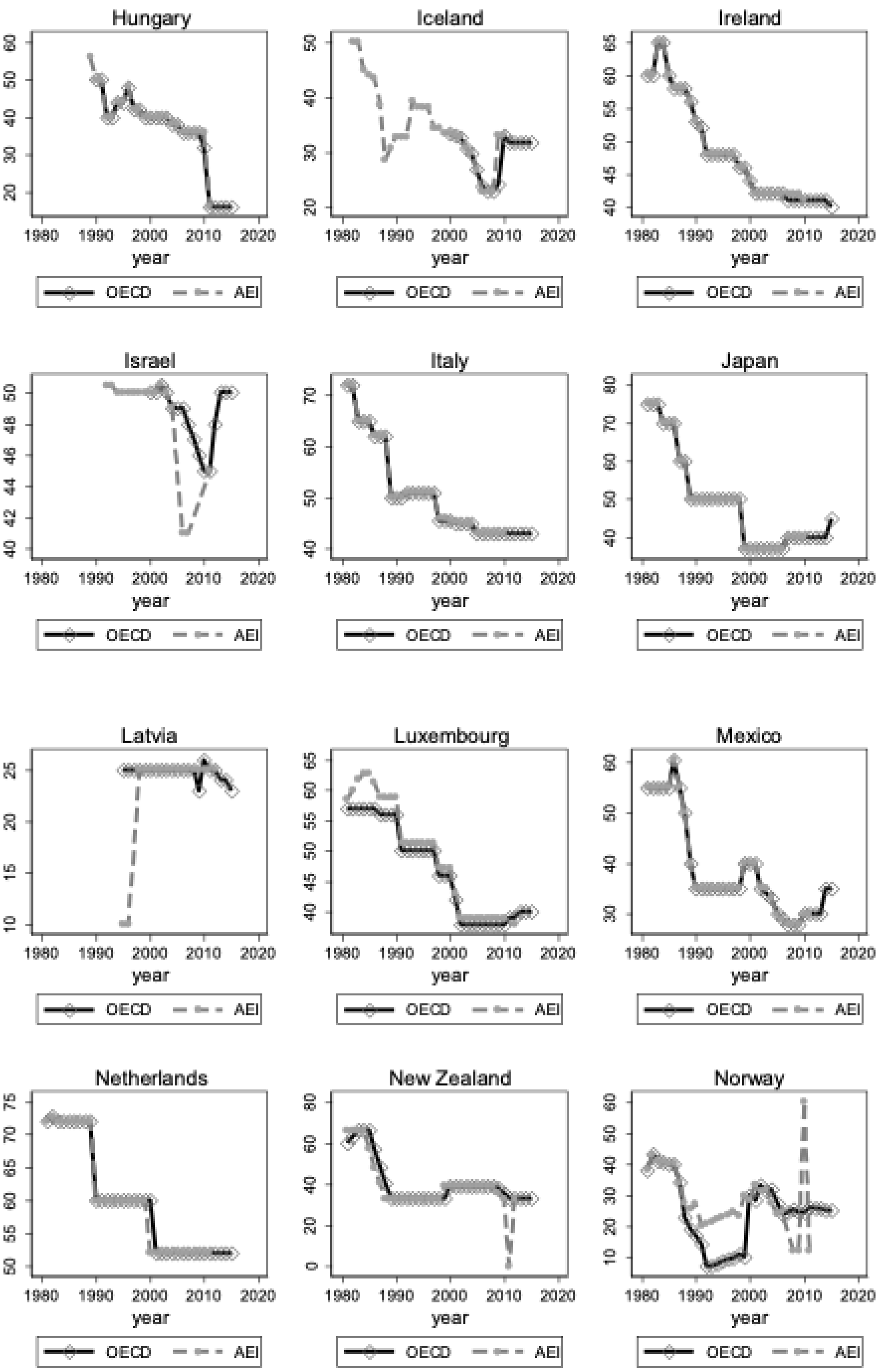

Hungary
Iceland
Ireland
Israel
Italy
Japan
Poland
Portugal
Slovak Republic
Slovenia
Spain
Sweden
year
OECD
AEI

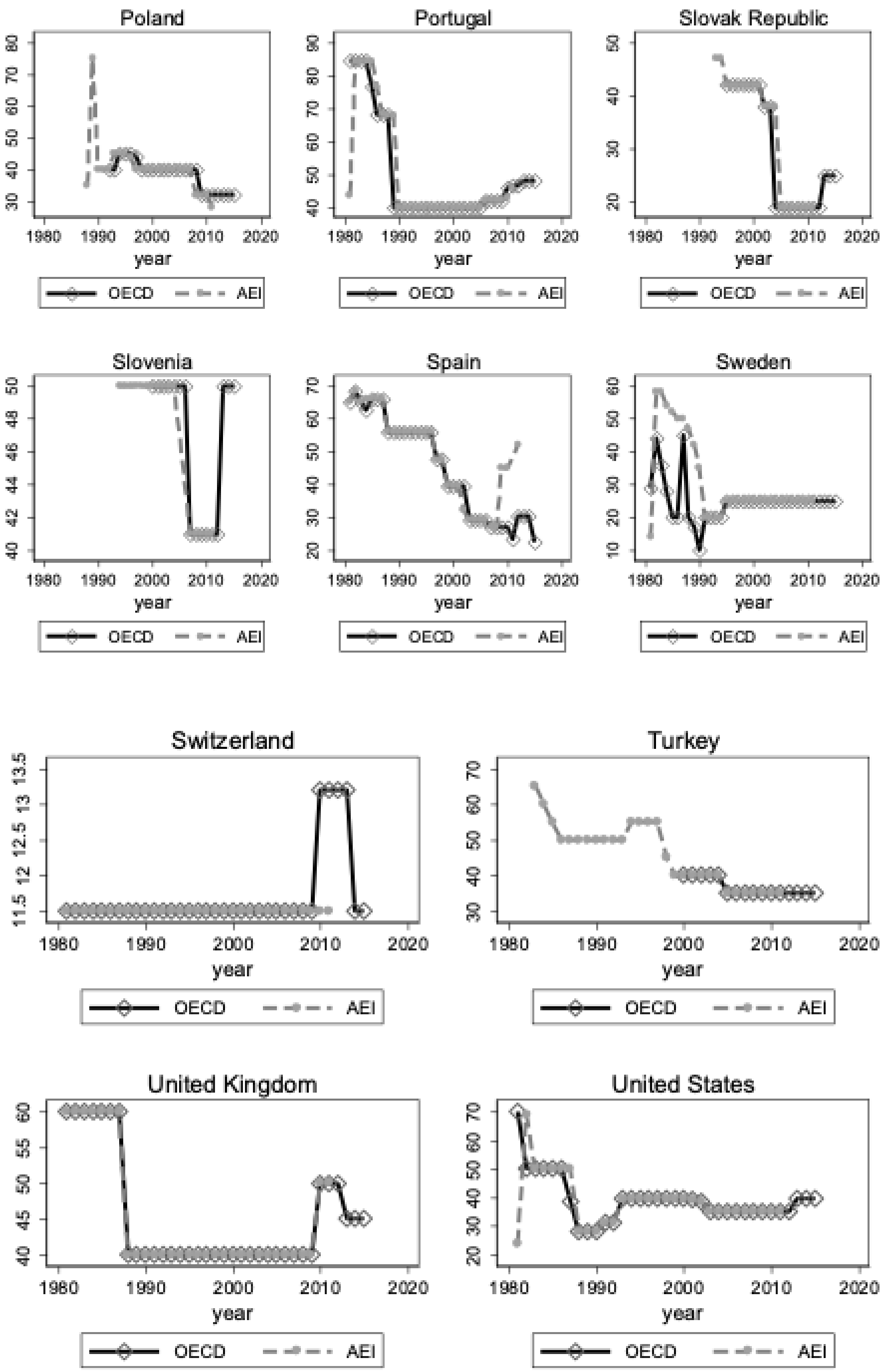


Notes: The central government CIT rate series for the 34 countries that appear in both the OECD and AEI databases. Sources: OECD and AEI tax databases, Republic of Estonia Tax and Customs Board, Caiumi and Di Biaggio (2015), the University of Michigan's World Tax Database (WTD), Bessa (2016), and KPMG's Worldwide Corporate Tax Tables.

Figure C.2: Comparison of Personal Income Tax Rates from the OECD Tax Database and the AEI International Tax Database

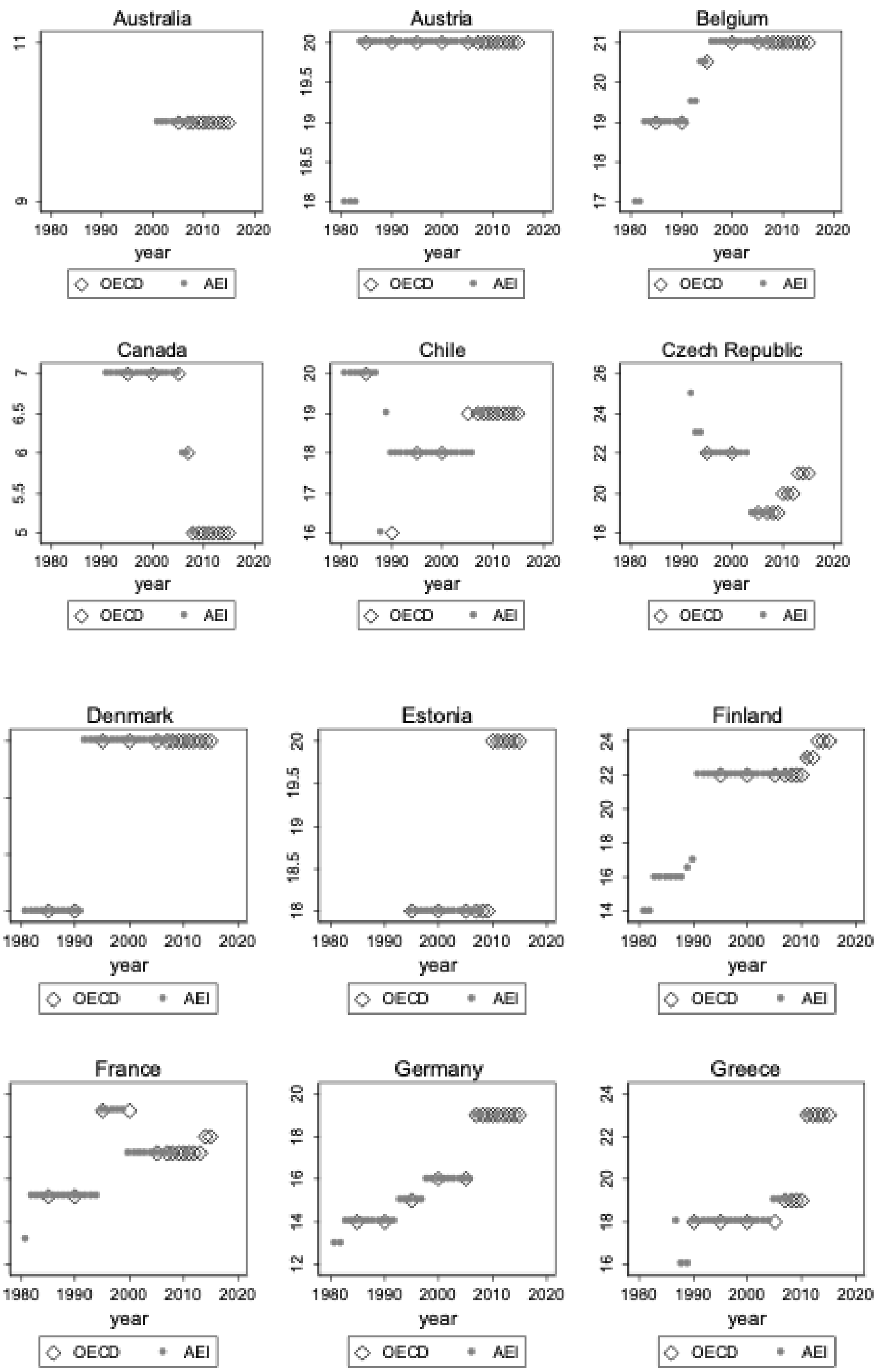


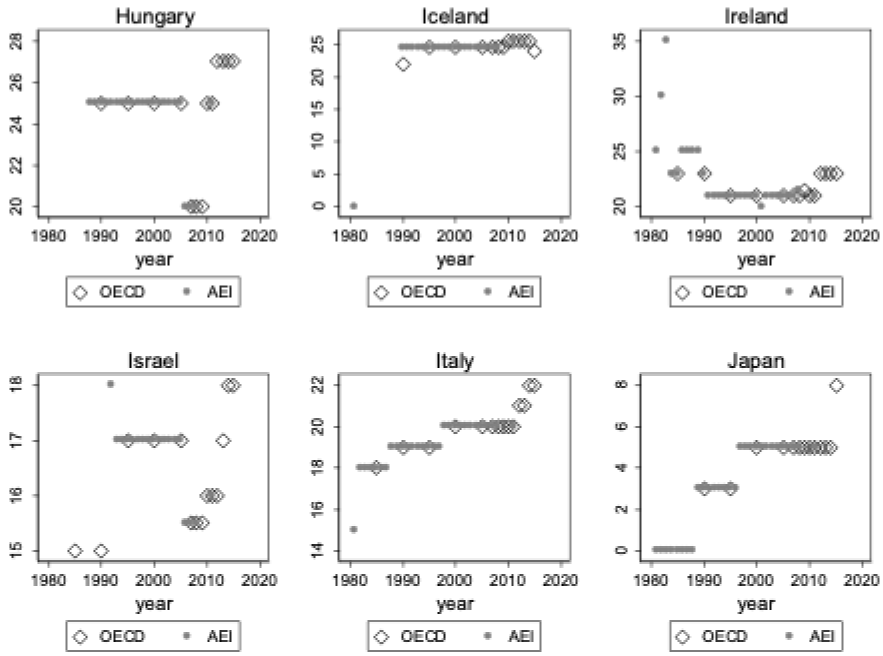

Hungary
Iceland
Ireland
Israel
Italy
Japan
Latvia
Luxembourg
Mexico
Netherlands
New Zealand
Norway
year
OECD
AEI

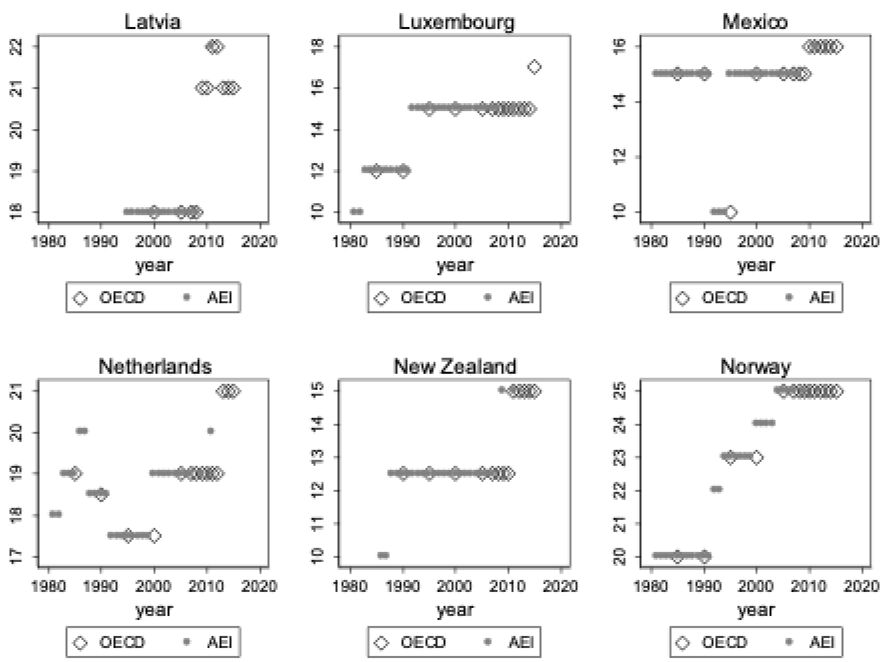


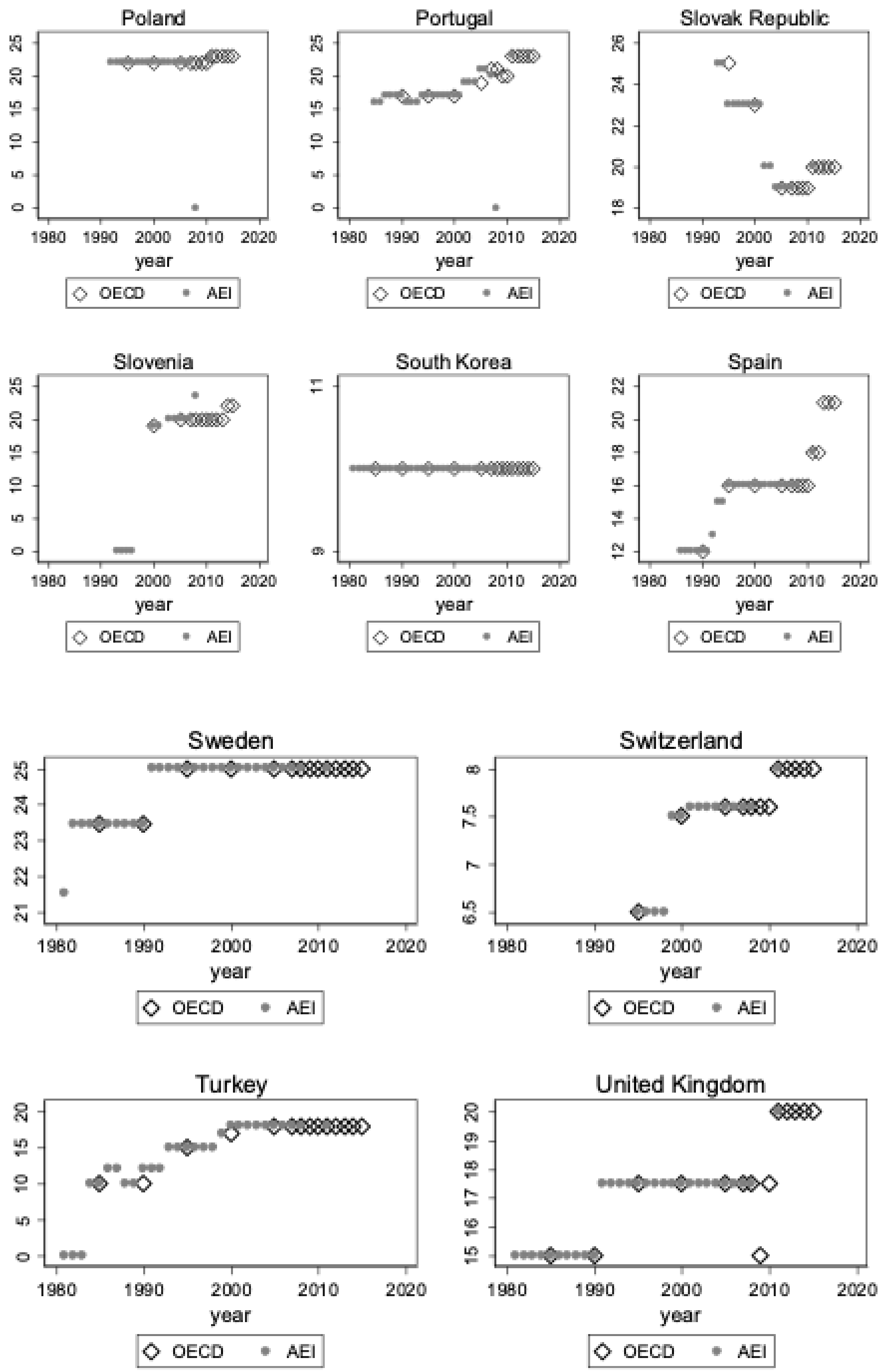
Notes: The central government PIT rate series for the 34 countries that appear in both the OECD and AEI databases. Sources: AEI and OECD tax databases, the Canadian government, the Institut des Politiques Publiques (2014), the Tax Policy Center (TPC), and the University of Michigan's Worldwide Tax Database (WTD).

Figure C.3: Comparison of Value-Added Tax Rates from the OECD Tax Database and the AEI International Tax Database




Hungary
Iceland
Ireland
Israel
Italy
Japan
year
OECD
AEI


Latvia
Luxembourg
Mexico
Netherlands
New Zealand
Norway
year
OECD
AEI



Notes: The central government VAT rate series for the 34 countries that appear in both the OECD and AEI databases. Sources: OECD and AEI tax databases.

C.2 Identifying Exogenous Tax Policy Shocks

Table C.1: What are the Countries and Time Periods Covered?

Country	Years of statutory rate and base data: TPRD (IMF) and Kawano and Slemrod (2016)	Years of exogenous tax changes data: Alesina et al. (2017)	Years of exogenous tax changes data: Dabla-Norris and Lima (2023)
Australia	1981 - 2015	1978 - 2014	1978 - 2015
Austria	1981 - 2015	1978 - 2014	1978 - 2015
Brazil	1981 - 2015		
Belgium	1981 - 2015	1978 - 2014	
Canada	1981 - 2015	1978 - 2014	1978 - 2015
China	1981 - 2015		
Czech Republic	1981 - 2015		
Denmark	1981 - 2015	1978 - 2014	
Finland	1981 - 2015	1978 - 2014	
France	1981 - 2015	1978 - 2014	1978 - 2015
Germany	1981 - 2015	1978 - 2014	1978 - 2015
Greece	1981 - 2015		
India	1981 - 2015		
Ireland	1981 - 2015	1978 - 2014	
Italy	1981- 2015	1978 - 2014	1978 - 2015
Japan	1981 - 2011	1978 - 2014	
Luxemburg	1981 - 2015		
Mexico	1981 - 2015		
Poland	1981 - 2015		
Portugal	1981 - 2015	1978 - 2014	1978 - 2015
South Korea	1981 - 2015		
Spain	1981 - 2015	1978 - 2014	1978 - 2015
Turkey	1981 - 2015		
Sweden	1981 - 2015	1978 - 2014	
United Kingdom	1981 - 2015	1978 - 2014	1978 - 2015
United States	1981 - 2015	1978 - 2014	1978 - 2015

Table C.2: What are "Exogenous" and "Endogenous" Tax Policy Shocks?

Paper	Romer and Romer (2010)	Alesina et al. (2017)	Dabla-Norris and Lima (2023)
Set of tax legislation	All tax legislation in the United States, 1945 -2007	Fiscal consolidations of 16 OECD Countries, 1978—2014	All tax legislation in 10 OECD countries, roughly 1978—2015
Definition of exogenous tax change	"those not taken to offset factors pushing growth away from normal"	"those not implemented with the objective of cyclical stabilization"	"taken in response to existing fiscal imbalances, i.e. the outcome of past rather than contemporaneous shocks"
Criteria for identifying legislation as exogenous	⋆ No motivation to counteract current or forecasted shocks OR ⋆ Motivation is to reduce an inherited deficit	⋆ No motivation to counteract current or forecasted shocks OR ⋆ Motivation is to reduce an inherited deficit	⋆ No motivation to counteract current or forecasted shocks OR ⋆ Motivation is to reduce an inherited deficit
Definition of endogenous tax change	One "taken to offset developments that cause output growth to differ from normal"	Not explicitly defined, but assumed to be all fiscal consolidations that fail to meet the above criteria	One "taken to offset contemporaneous shocks"
Criteria for identifying legislation as endogenous	⋆ Tax changes are designed to be counteract macroeconomic shocks OR ⋆ Tax changes are motivated by paying for new government spending	⋆ Not explicitly defined, but assumed to be failure to pass either of the above criteria	⋆ Tax changes are designed to be counteract macroeconomic shocks OR ⋆ Tax changes are motivated by paying for new government spending
Controls for endogeneity of the tax base?	No, unit of measurement is the change in revenue as a percent of GDP	No, unit of measurement is the change in revenue as a percent of GDP	Yes, unit of measurement is in currency, and measures are categorized into rate and base changes

Table C.3: Granger Causality Tests

	Output	Debt	Gov Purchases	Inflation	Short-term Rate	Tax Revenue
Panel A: Kawano and Slemrod (2016) Dates						
All Rates						
F-stat	0.835	1.165	0.999	2.539	2.644	1.175
P-stat	0.443	0.325	0.38	0.0952	0.0871	0.322
PIT Rates						
F-stat	0.469	0.828	1.44	1.082	2.409	0.714
P-stat	0.63	0.446	0.252	0.351	0.106	0.497
CIT Rates						
F-stat	0.872	2.738	3.364	1.197	1.088	2.905
P-stat	0.428	0.0803	0.0476	0.316	0.349	0.0698
VAT Rates						
F-stat	1.722	0.00309	1.249	0.709	0.0682	0.231
P-stat	0.195	0.997	0.3	0.5	0.934	0.795
Panel B: Alesina et al. (2017) Dates						
All Rates						
F-stat	2.72	0.693	0.107	0.697	2.083	1.689
P-stat	0.0816	0.508	0.899	0.506	0.142	0.201
PIT Rates						
F-stat	0.617	0.432	0.036	0.301	1.361	0.779
P-stat	0.546	0.653	0.965	0.742	0.271	0.468
CIT Rates						
F-stat	4.831	0.469	0.0857	0.284	1.527	0.419
P-stat	0.0149	0.63	0.918	0.754	0.233	0.661
VAT Rates						
F-stat	2.501	0.963	2.607	2.486	1.877	3.091
P-stat	0.0984	0.392	0.0893	0.0992	0.17	0.0593
Panel C: Dabla-Norris and Lima (2023) Dates						
All Rates						
F-stat	0.851	0.254	0.0779	0.773	1.76	0.0888
P-stat	0.437	0.778	0.925	0.47	0.189	0.915
PIT Rates						
F-stat	0.238	0.375	1.863	0.71	2.952	0.0794
P-stat	0.79	0.69	0.172	0.499	0.0666	0.924
CIT Rates						
F-stat	1.233	1.275	2.524	0.634	1.657	1.25
P-stat	0.305	0.294	0.0965	0.537	0.207	0.3
VAT Rates						
F-stat	1.359	0.566	1.2	0.796	0.579	0.942
P-stat	0.272	0.573	0.314	0.46	0.566	0.4

Notes: Joint tests of significance of lagged predictor variables (across columns) on narratively-defined exogenous tax rate shocks for the top personal income tax rate (PIT), top marginal corporate tax rate (CIT), and standard consumption tax rate (VAT).

C.3 The Narrative Approach for Categorizing Tax Changes

C.3.1 Summary

The narrative approach requires judgements regarding the predominant motivations behind each policy change, and these determinations are made at the discretion of the researcher. As a result, not all methodologies produce the same results. In this subsection, we document the discrepancies between our three sources of cross-country data and describe patterns for the series of exogenous and endogenous tax changes.

Dabla-Norris and Lima (2023) start with the set of fiscal consolidation episodes in Alesina et al. (2015), but use additional primary-source information from finance ministries, tax authorities, and legislatures to identify specific changes to tax rates and tax base definitions contained within these consolidation periods. There are three key differences between the resulting datasets. First, seven countries present in Alesina et al. (2015) are excluded due to limitations on the additional sources. Second, Dabla-Norris and Lima (2023) find that revenue changes in consolidation periods identified in Alesina et al. (2015) frequently contain tax changes unrelated to fiscal consolidations, such as reforms intended to stimulate the current economy. These non-consolidation tax changes are then coded as either exogenous or endogenous according to the definitions in Romer and Romer (2010). Finally, with their additional sources, Dabla-Norris and Lima (2023) note when tax consolidations are announced at different dates than recorded in Alesina et al. (2015) and identify additional tax consolidations. In all, Dabla-Norris and Lima (2023) find 38 country-year pairs - or 25% of the country-year pairs - with consolidation measures that are in addition to, or differently-timed than, those for the same countries in Alesina et al. (2015).

We uncovered 104 tax rate changes in our data set that were not included in Dabla-Norris and Lima (2023). We validated 73 of these rate changes directly with the TPRD (version 4.0). The remaining 31 rate changes were found in other national sources, such as data collections of countries' finance ministries. To determine whether the 104 additional tax rate changes we identify are exogenous, we rely on news articles from the International Bureau of Fiscal Documentation (IBFD). If these articles referenced short-run economic conditions, we defined the tax change to be endogenous. If the motivation was to reduce an inherited deficit or simplify the tax code in a manner unrelated to current economic growth, we code the tax change as endogenous. In addition, we analyze GDP growth and if a tax change was announced just preceding or in a midst of a recession, we determined the tax change was likely motivated by current economic conditions. In all, this exercise uncovered 32 additional exogenous rate changes.

C.3.2 The Composition of Exogenous Tax Changes

In Table C.4, we present the composition of exogenous tax rate changes using our measure of exogeneity, relative to the full set of tax rate shocks. Overall, the means and medians across the two series are similar, but exogenous tax changes tend to be much smaller in magnitude. The largest difference between the two datasets are the skewness of the tax rate changes, indicating that exogenous tax changes are much more likely to consist of tax changes greater than their means.

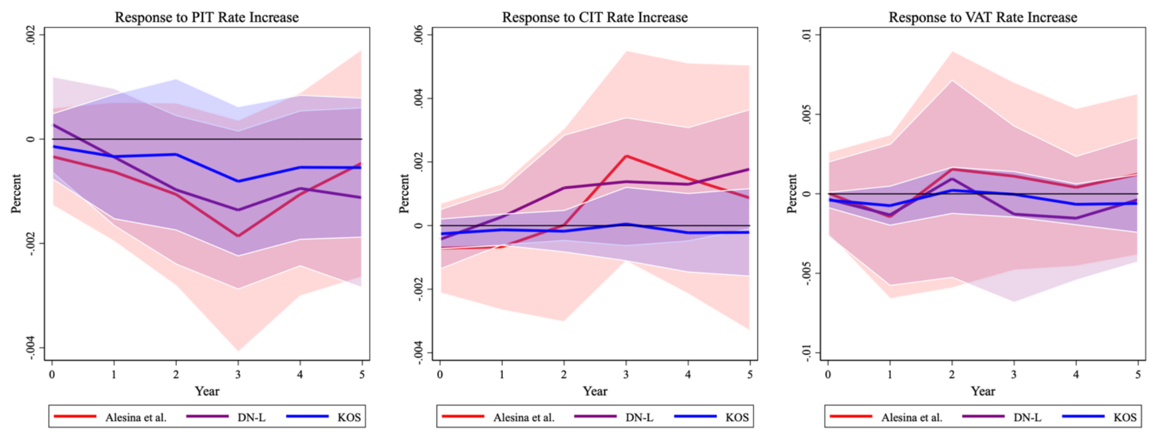
Table C.4: The Composition of Exogenous Tax Rate Changes

	N	Mean	Std. Dev.	Min	Max	Median	Kurtosis	Skewness
PIT rate change	614	-2.846	10.4	-48	55	-2.5	8.517	0.716
CIT rate change	723	-2.43	7.029	-42.5	61.65	-2	23.594	1.018
VAT rate change	308	1.56	5.958	-22	25	1	7.393	0.202
Exogenous PIT rate change	40	-3.161	8.833	-28	10	-2	3.428	-0.903
Exogenous CIT rate change	46	-3.326	6.109	-25	5	-2	5.184	-1.381
Exogenous VAT rate change	24	0.6	4.443	-18.5	5	1.5	15.67	-3.542

Notes: Descriptive statistics for tax rate changes. Observations are at the country-year level. PIT, CIT, and VAT refer to the top marginal personal income tax rate, top marginal corporate income tax rate, and standard VAT tax rate, respectively. Exogenous rate changes are narratively-identified.

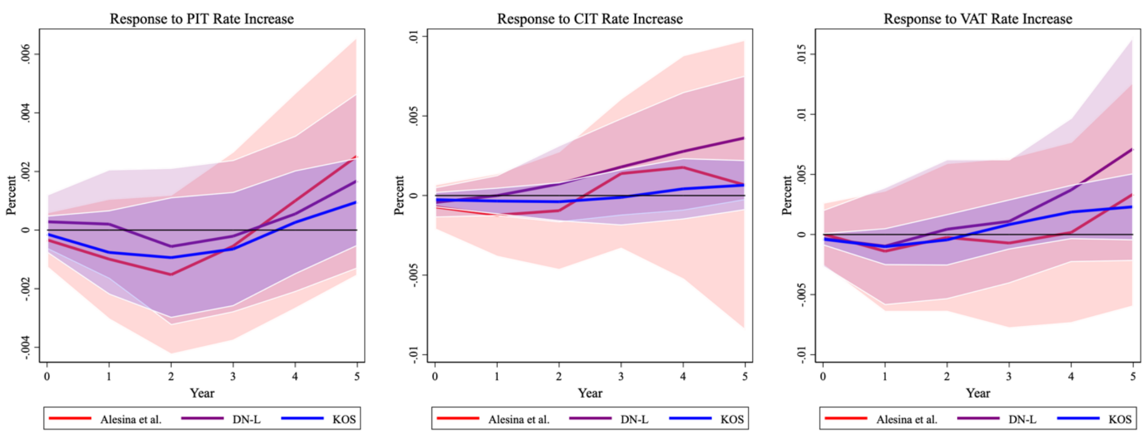
C.4 Additional Figures and Table

Figure C.4: Cumulative Five-Year Effect of Exogenous Tax Rate Changes, Alternative Dates



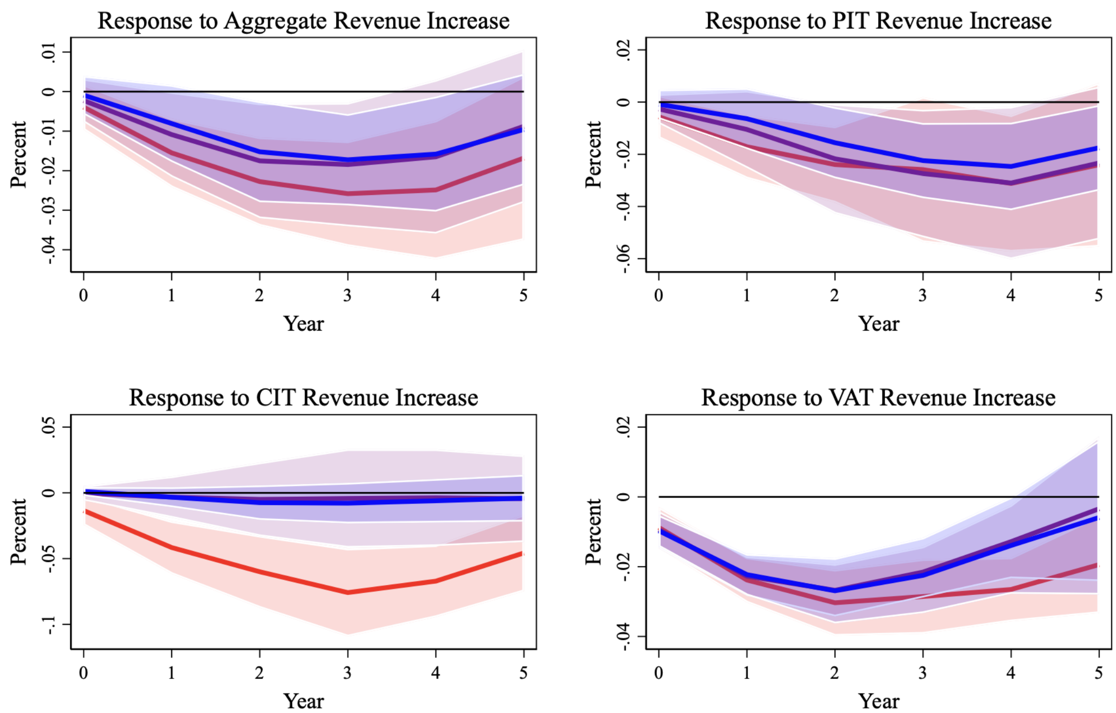
Notes: The cumulative effects of a one-point increase in tax rates on GDP per capita over the subsequent five years from linear regressions. Corresponds to Figure ??, Panel A in the main text. Results using Alesina et al. (2015) exogenous rates are shown in red. Results with Dabla-Norris and Lima (2023) exogenous dates are shown in purple, and denoted DN-L. Results with our own measure of exogenous dates are shown in blue, and denoted KOS. All impulse responses are plotted with 95% confidence intervals constructed from Driscoll-Kraay standard errors.

Figure C.5: Five-Year Effects of Exogenous Tax Rate Changes, Alternative Dates



Notes: The effects of a one-point increase in tax rates on GDP per capita over the subsequent five years from local projections. Corresponds to Figure 3.3, Panel A in the main text. Results using Alesina et al. (2015) exogenous rates are shown in red. Results with Dabla-Norris and Lima (2023) exogenous dates are shown in purple, and denoted DN-L. Results with our own measure of exogenous dates are shown in blue, and denoted KOS. All impulse responses are plotted with 95% confidence intervals constructed from Driscoll-Kraay standard errors.

Figure C.6: Five-Year Effects of Exogenous Tax Revenue Changes, Alternative Dates



Notes: The effects of a one-point increase in tax revenue as a percent of GDP on GDP per capita over the subsequent five years from local projections. Corresponds to Figure 3.4 in the main text. Results using Alesina et al. (2015) exogenous rates are shown in red. Results with Dabla-Norris and Lima (2023) exogenous dates are shown in purple, and denoted DN-L. Results with our own measure of exogenous dates are shown in blue, and denoted KOS. All impulse responses are plotted with 95% confidence intervals constructed from Driscoll-Kraay standard errors.

Table C.5: Point Estimates for Alternative Dates

	KOSH	Alesina et al. (2015)	Dabla-Norris and Lima (2023)
Panel A: Five-Year Point Estimate for Linear Regressions			
PIT rate	-0.0005458	-0.0004545	-0.001123
	[0.0006853]	[0.0011197]	[0.0008847]
CIT rate	-0.0002121	0.0008668	0.0017789
	[0.000711]	[0.0021429]	[0.0009637]
VAT rate	-0.0006055	0.0012373	-0.0003612
	[0.0009471]	[0.0026053]	[0.0020116]
Panel B: Five-Year Point Estimates for Local Projections			
PIT rate	0.0009629	0.0025298	0.0016846
	[0.0007664]	[0.0020812]	[0.0015286]
CIT rate	0.0006581	0.0006582	0.0036278
	[0.000802]	[0.0046618]	[0.0019966]
VAT rate	0.0023065	0.0033642	0.0071424
	[0.0014251]	[0.0047725]	[0.0047671]

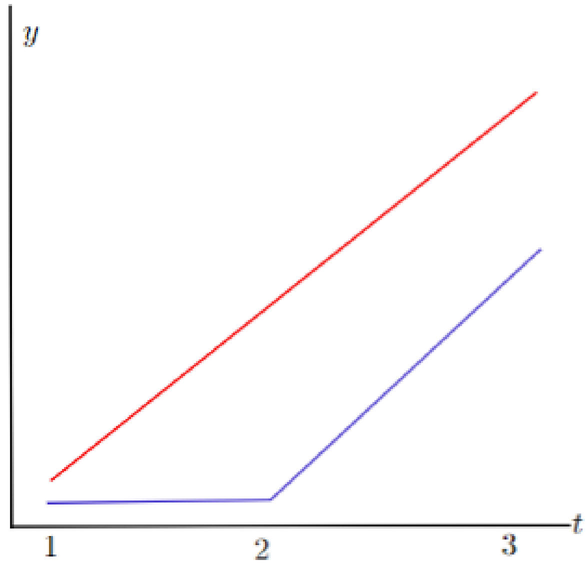
Notes: Point estimates for the impact on GDP per capita five years after the tax rate change from linear regressions (Panel A) and local projections (Panel B). Corresponds to the point estimates in Panel A of both Figures 2 and 3 in the main text, for top marginal personal income tax rates (PIT), top marginal corporate income tax rates (CIT), and standard consumption tax rates (VAT). Results with our own measure of exogenous dates are denoted KOSH. We include results using Alesina et al. (2015) exogenous dates and Dabla-Norris and Lima (2023) exogenous dates. Driscoll-Kraay standard errors are in brackets.

C.5 Accounting for Staggered Treatments in Estimation Strategies

C.5.1 Staggered Treatment in a Simple Setting

To illustrate the issue surrounding local projections, let us consider a toy example consisting of two countries, three periods, and staggered treatment timings. This puts us essentially in the setting of Goodman-Bacon (2021). Schematically, I illustrate the scenario in the diagram below. Let red and blue lines each represent the trajectory of an outcome, for country 1 and country 2, respectively. Country 1 is treated in period 1, and country 2 is treated in time 2.

Figure C.7: Schematic of the Treatment Timing Variations



We analyze the estimand of a local projections regression, defined as

$$Y_{g,t+\ell} = \beta_\ell D_{g,t} + \delta_g + \delta_t + \varepsilon_{g,t} \ , \tag{C.1}$$

where $Y_{g,t+\ell}$ is an outcome of interest in country g at time $t+\ell$, $D_{g,t}$ is treatment indicator of country g at time t, and δ_g, δ_t are country- and time-fixed effects, respectively. The estimated effect of country g being exposed to $\ell \geq 0$ periods of treatement is captured by the coefficient β_ℓ. Note that in the main text, we use h to indicate number of time periods exposed to treatment; here, we use the more widely-used subscript ℓ to indicate the number of time periods exposed to treatment.

The question is what is the estimated parameter when there are two countries and two treatment timings, which is analogous to the setting of Goodman-Bacon (2021). Heuristically, there is a correct comparison at time 1, but at times 2 and 3, there are "forbidden comparisons", in the sense that we compare between two treated countries.

Let's derive the estimand for the general case first, then specialize into our toy example. Suppose the population of country g at time t is given by $N_{g,t}$. By Frisch-Waugh-Lovell, the OLS estimtaor of β_ℓ, denoted $\hat{\beta}_\ell$, is equivalent to

$$\hat{\beta}_\ell = \frac{(T-\ell)^{-1}G^{-1}\sum_{1\leq g\leq G}\sum_{1\leq t\leq T-\ell} Y_{g,t+\ell}N_{g,t}\tilde{D}_{g,t}}{(T-\ell)^{-1}G^{-1}\sum_{1\leq g\leq G}\sum_{1\leq t\leq T-\ell} D_{g,t+\ell}N_{g,t}\tilde{D}_{g,t}} \ . \tag{C.2}$$

Here, $\tilde{D}_{g,t}$ denotes the projection of $D_{g,t}$ onto the country- and time-fixed effects, defined by

$$\tilde{D}_{g,t} = D_{g,t} - \overline{D}_g - \overline{D}_t + \overline{D} \ , \tag{C.3}$$

where

$$\overline{D}_g = \frac{\sum_{1\leq t\leq T-\ell} N_{g,t} D_{g,t}}{\sum_{1\leq t\leq T-\ell} N_{g,t}} \ , \tag{C.4}$$

$$\overline{D}_t = \frac{\sum_{1\leq g\leq G} N_{g,t} D_{g,t}}{\sum_{1\leq g\leq G} N_{g,t}} \ , \tag{C.5}$$

$$\overline{D} = \frac{\sum_{1\leq t\leq T-\ell}\sum_{1\leq g\leq G} N_{g,t} D_{g,t}}{\sum_{1\leq t\leq T-\ell}\sum_{1\leq g\leq G} N_{g,t}} \ . \tag{C.6}$$

In our setting, with only three periods and (in common) one post-treatment period, we can estimate β_ℓ for $\ell = 0$ (with the convention that $\ell = -1$ is the period when the treatment occurs). Furthermore, there are only two countries, so $G = 2$, and—following de Chaisemartin and D'Haultfoeuille (2022), let's assume $N_{g,t} = N_g$ (i.e. constant population in every country over time; this greatly simplifies calculations).

Under these conditions, the numerator of (C.2) (without the "scaling factor" $T^{-1}G^{-1}$) is

$$\sum_{1\leq t\leq T} (Y_{1,t}N_1\tilde{D}_{1,t} + Y_{2,t}N_2\tilde{D}_{2,t}) \ . \tag{C.7}$$

By direct calculations,

$$\begin{aligned}
\sum_{1\leq t\leq T} Y_{1,t}N_1\tilde{D}_{1,t} &= \sum_{1\leq t\leq T} Y_{1,t}N_1(D_{1,t} - \overline{D}_1 - \overline{D}_t + \overline{D}) && \text{(C.8)}\\
&= \sum_{1\leq t\leq T} \big(Y_{1,t}N_1D_{1,t} - Y_{1,t}N_1\overline{D}_1 - Y_{1,t}N_1\overline{D}_t + Y_{1,t}N_1\overline{D}\big) && \text{(C.9)}\\
&= \sum_{1\leq t\leq T} Y_{1,t}N_1D_{1,t} - \frac{1}{T}\sum_{1\leq t\leq T} Y_{1,t}N_1\left(\sum_{1\leq t\leq T} D_{1,t}\right) && \text{(C.10)}\\
&\quad - \frac{1}{2}\sum_{1\leq t\leq T} Y_{1,t}N_1\left(\sum_{1\leq g\leq G} D_{g,t}\right) && \text{(C.11)}\\
&\quad + \frac{1}{2T}\sum_{1\leq t\leq T} Y_{1,t}N_1\left(\sum_{1\leq t\leq T}\sum_{1\leq g\leq G} D_{g,t}\right) && \text{(C.12)}
\end{aligned}$$

$$= \sum_{1\le t\le T} Y_{1,t} N_1 \left(D_{1,t} - \frac{1}{2} \sum_{1\le g\le G} D_{g,t} \right) \tag{C.13}$$

$$+ \frac{1}{2T} \sum_{1\le t\le T} Y_{1,t} N_1 \underbrace{\left(\sum_{1\le t\le T} \sum_{1\le g\le G} D_{g,t} \right)}_{=2} \tag{C.14}$$

$$- \frac{1}{T} \sum_{1\le t\le T} Y_{1,t} N_1 \underbrace{\left(\sum_{1\le t\le T} D_{1,t} \right)}_{=1} \tag{C.15}$$

$$= \sum_{1\le t\le T} Y_{1,t} N_1 \left(D_{1,t} - \frac{1}{2} \sum_{1\le g\le G} D_{g,t} \right) \ . \tag{C.16}$$

This toy example has the feature that $D_{1,t} = 1$ if $t = 0$, $D_{1,t} = 1$ if $t = 1, 2$, $D_{2,t} = 1$ if $t = 1$, $D_{2,t} = 0$ if $t \in \{0, 2\}$. Simplifying, we have that the numerator equals

$$\sum_{1\le t\le T} Y_{1,t} N_1 \tilde{D}_{1,t} = \frac{N_1}{2} \left(Y_{1,1} - Y_{1,2} \right) \ . \tag{C.17}$$

The calculations for country 2 is symmetric; therefore,

$$\sum_{1\le t\le T} Y_{2,t} N_2 \tilde{D}_{2,t} = \frac{N_2}{2} \left(-Y_{2,1} + Y_{2,2} \right) \ . \tag{C.18}$$

Hence, the numerator of (C.2) is equivalent to

$$\frac{1}{2} \left((N_1 Y_{1,1} - N_2 Y_{2,1}) + (N_2 Y_{2,2} - N_1 Y_{1,2}) \right) \tag{C.19}$$

Now, again specializing the estimand to our setting of two countries and three periods, we can rewrite the denominator of (C.2) as

$$\sum_{1\le t\le T} \left(D_{1,t} N_{1,t} \tilde{D}_{1,t} + D_{2,t} N_{2,t} \tilde{D}_{2,t} \right) \tag{C.20}$$

for $\ell = 0$. By direct calculations,

$$\sum_{1\le t\le T} D_{1,t} N_1 \tilde{D}_{1,t} = \sum_{1\le t\le T} D_{1,t} N_1 (D_{1,t} - \overline{D}_1 - \overline{D}_t + \overline{D}) \tag{C.21}$$

$$= \sum_{1\le t\le T} \left(D_{1,t} N_1 D_{1,t} - D_{1,t} N_1 \overline{D}_1 - D_{1,t} N_1 \overline{D}_t + D_{1,t} N_1 \overline{D} \right) \tag{C.22}$$

$$= \sum_{1 \le t \le T} N_1 D_{1,t}^2 - \frac{1}{T} \sum_{1 \le t \le T} D_{1,t} N_1 \left(\sum_{1 \le t \le T} D_{1,t} \right) \tag{C.23}$$

$$- \frac{1}{2} \sum_{1 \le t \le T} D_{1,t} N_1 \left(\sum_{1 \le g \le G} D_{g,t} \right) \tag{C.24}$$

$$+ \frac{1}{2T} \sum_{1 \le t \le T} D_{1,t} N_1 \left(\sum_{1 \le t \le T} \sum_{1 \le g \le G} D_{g,t} \right) \tag{C.25}$$

$$= \sum_{1 \le t \le T} D_{1,t} N_1 \left(D_{1,t} - \frac{1}{2} \sum_{1 \le g \le G} D_{g,t} \right) \tag{C.26}$$

$$+ \frac{1}{2T} \sum_{1 \le t \le T} D_{1,t} N_1 \underbrace{\left(\sum_{1 \le t \le T} \sum_{1 \le g \le G} D_{g,t} \right)}_{=2} \tag{C.27}$$

$$- \frac{1}{T} \sum_{1 \le t \le T} D_{1,t} N_1 \underbrace{\left(\sum_{1 \le t \le T} D_{1,t} \right)}_{=1} \tag{C.28}$$

$$= \sum_{1 \le t \le T} D_{1,t} N_1 \left(D_{1,t} - \frac{1}{2} \sum_{1 \le g \le G} D_{g,t} \right) \ . \tag{C.29}$$

By identical arguments as before,

$$\sum_{1 \le t \le T} D_{1,t} N_1 \tilde{D}_{1,t} = \frac{N_1}{2} \ . \tag{C.30}$$

The calculations for country 2 is symmetric; therefore,

$$\sum_{1 \le t \le T} D_{2,t} N_2 \tilde{D}_{2,t} = \frac{N_2}{2} \ . \tag{C.31}$$

Hence, the denominator of (C.2) (without the "scaling factor" $T^{-1}G^{-1}$) is equivalent to

$$\frac{N_1}{2} + \frac{N_2}{2} \ . \tag{C.32}$$

By the continuous mapping theorem,

$$\hat{\beta}_0 \xrightarrow{\text{P}} \left(\frac{1}{2} \left(\mathbb{E}[N_1 Y_{1,1} - N_2 Y_{2,1}] + \mathbb{E}[N_2 Y_{2,2} - N_1 Y_{1,2}] \right) \right) \times \left(\frac{\mathbb{E}[N_1]}{2} + \frac{\mathbb{E}[N_2]}{2} \right)^{-1} \tag{C.33}$$

$$= \frac{\mathbb{E}[N_1 Y_{1,1} - N_2 Y_{2,1}] + \mathbb{E}[N_2 Y_{2,2} - N_1 Y_{1,2}]}{\mathbb{E}[N_1 + N_2]} \ . \tag{C.34}$$

We note a few things about the estimand, in this simple setting.

One, the linear projections estimator, as shown above, does not make use of information from time period 3 in this setting. That alleviates an a priori concern that the estimator does make comparisons beyond the one period after which the treatment occurs.

Two, the estimand does not represent a causal effect. Even if two countries' populations are constant and identical over time, i.e. $N_1 = N_2$, the estimand becomes

$$\frac{\mathbb{E}[Y_{1,1} - Y_{2,1}] + \mathbb{E}[Y_{2,2} - Y_{1,2}]}{2} \,. \tag{C.35}$$

This is the estimand shown in Section 3.5 of the main text. Note that while $Y_{1,1} - Y_{2,1}$ is a valid comparison (i.e. this is a treated group's outcome minus control group's outcome, where the control group is untreated), $Y_{2,2} - Y_{1,2}$ is not a valid comparison, as group 2 just got treated, but group 1 is already treated. As such, there is a comparison between treated and already-treated units, which parallels the observation made in Goodman-Bacon (2021).

Three, based on analyzing this example, the issues highlighted above will be exacerbated when the setting becomes more complicated. For one, having more countries will introduce a lot more pairwise comparisons in outcomes made by the local projections estimator! When there are more countries with different treatment timings, there will likely be more "forbidden comparisons", i.e. comparing outcomes between treated and already-treated groups Goodman-Bacon (2021). For another, note that the weights, as derived above, are functions of treatment timing. Therefore, in more complicated scenarios—with repeated treatments—these weights will be less interpretable and more complicated-looking. Finally, the same conclusion applies to scenarios with potentially non-monotone treatment, e.g. tax changes that reverse signs over time. As we commented in the main text, deriving the closed-form expressions for these weights, in complicated settings, seems out of reach.

Table C.6: Total Effects From Methodology of de Chaisemartin and D'Haultfoeuille (2022)

	KOSH	Alesina et al. (2015)	Dabla-Norris and Lima (2023)
PIT rate	-0.0031846	0.0069231	0.0081564
	[0.0047588]	[0.0088656]	[0.0073818]
CIT rate	0.0093504	0.0071036	0.0350038
	[0.0050165]	[0.013377]	[0.0134774]
VAT rate	-0.002753	-0.0145792	0.0002338
	[0.0144631]	[0.0459528]	[0.0240164]

Notes: Point estimates for the total effect of tax rate changes using the methodology of de Chaisemartin and D'Haultfoeuille (2022). Corresponds to the point estimates in Panel A of Figure 3.5 in the main text. Results with our own measure of exogenous dates are denoted KOS. We include results using Alesina et al. (2015) exogenous dates and Dabla-Norris and Lima (2023) exogenous dates. Results for the the top marginal personal income tax rate (PIT), top marginal corporate tax rate (CIT), and standard consumption tax rate (VAT) are shown. Standard errors are constructed using the statistical package in de Chaisemartin and D'Haultfoeuille (2022).

BIBLIOGRAPHY

Abraham, K. G., Haltiwanger, J. C., Hou, C., Sandusky, K., and Spletzer, J. R. (2021). Reconciling Survey and Administrative Measures of Self-Employment. *Journal of Labor Economics*, 39(4):825–860.

Acosta-Ormaechea, S., Sola, S., and Yoo, J. (2019). Tax Composition and Growth: A Broad Cross-Country Perspective. *German Economic Review*, 20(4):e70–e106.

Aguiar, M. a., Bils, M., and Boar, C. (2020). Who Are the Hand-to-Mouth? Working Paper 26643, National Bureau of Economic Research.

Alesina, A., Barbiero, O., Favero, C., Giavazzi, F., and Paradisi, M. (2017). The Effects of Fiscal Consolidations: Theory and Evidence. Technical Report w23385, National Bureau of Economic Research, Cambridge, MA.

Alesina, A., Favero, C., and Giavazzi, F. (2015). The Output Effect of Fiscal Consolidation Plans. *Journal of International Economics*, 96:S19–S42.

Altig, D. and Carlstrom, C. T. (1999). Marginal Tax Rates and Income Inequality in a Life-Cycle Model. *American Economic Review*, 89(5):1197–1215.

Altig, D., Ilin, E., Ruder, A., and Terry, E. (2020). Benefits Cliffs and the Financial Incentives For Career Advancement: a Case Study of the Health Care Services Career Pathway. *Frb Atlanta Community and Economic Development Discussion Paper*, (2020-1).

Alumbaugh, T., F. D. H. M. J. M. O. J. and Xu, A. (2019). Tax Calculator. Technical report, Open Source Policy Center.

Amaglobeli, D. (2018). *Tax Policy Measures in Advanced and Emerging Economies: a Novel Database*. International Monetary Fund, Washington, D.C. OCLC: 1107344288.

Angelopoulos, K., Economides, G., and Kammas, P. (2007). Tax-Spending Policies and Economic Growth: Theoretical Predictions and Evidence from the OECD. *European Journal of Political Economy*, 23(4):885–902.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277.

Arnold, J. M., Brys, B., Heady, C., Johansson, Å., Schwellnus, C., and Vartia, L. (2011). Tax Policy For Economic Recovery and Growth. *The Economic Journal*, 121(550):F59–F80.

Baiardi, D., Profeta, P., Puglisi, R., and Scabrosetti, S. (2019). Tax Policy and Economic Growth: Does It Really Matter? *International Tax and Public Finance*, 26(2):282–316.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, 106(2):407.

Barro, R. J. and Redlick, C. J. (2011). Macroeconomic Effects from Government Purchases and Taxes. *The Quarterly Journal of Economics*, 126(1):51–102.

Bart, L. and Horton, E. (2024). Optimal Transfers With Children's Utility. *Working Paper*.

Bastian, J. (2018). Do EITC Expansions Pay For Themselves? Effects on Tax Revenue and Public Assistance Spending. In *111th Annual Conference on Taxation*. NTA.

Bastian, J. and Michelmore, K. (2018). The Long-Term Impact of the Earned Income Tax Credit on Children'S Education and Employment Outcomes. *Journal of Labor Economics*, 36(4):1127–1163.

Bastian, J. E. and Jones, M. R. (2021). Do EITC Expansions Pay For Themselves? Effects on Tax Revenue and Government Transfers. *Journal of Public Economics*, 196:104355.

Benabou, R. (2002). Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency? *Econometrica*, 70(2):481–517.

Benjamin, A. and Summers, L. H. (2018). Saving the Heartland: Place-Based Policies in 21st Century America. *Brookings Papers on Economic Activity, Spring.*

Bessa, A. R. T. (2016). A Taxa Efetiva de Imposto e a Reforma Do Irc de 2013. Master's thesis.

Blanchard, O. and Perotti, R. (2002). An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *The Quarterly Journal of Economics*, 117(4):1329–1368.

Blank, R. M. (2002). Evaluating Welfare Reform in the United States. *Journal of Economic Literature*, 40(4):1105–1166.

Blinder, A. S. and Rosen, H. S. (1985). Notches. *The American Economic Review*, 75(4):736–747.

Blundell, R., Costa Dias, M., Meghir, C., and Shaw, J. (2016). Female Labor Supply, Human Capital, and Welfare Reform. *Econometrica*, 84(5):1705–1753.

Blundell, R. and MaCurdy, T. (1999). Labor Supply: a Review of Alternative Approaches. *Handbook of Labor Economics*, 3:1559–1695.

Blundell, R., Meghir, C., and Neves, P. (1993). Labour Supply and Intertemporal Substitution. *Journal of Econometrics*, 59(1-2):137–160.

Borusyak, K., Jaravel, X., and Spiess, J. (2023). Revisiting Event Study Designs: Robust and Efficient Estimation. arXiv:2108.12419 [econ].

Brendler, P. (2023). Rising Earnings Inequality and Optimal Income Tax and Social Security Policies. *Journal of Monetary Economics*, 134:35–52.

Bureau of Labor Statistics (2018). State Unemployment Rates, March 2018, Seasonally Adjusted. Technical report, United States Department of Labor.

Caiumi, A. and Di Biagio, L. (2015). Corporate Effective Taxation in Italy Using a New Microsimulation Model: Istat–Matis. Technical report, Istat Working Papers, 13.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with Multiple Time Periods. *Journal of Econometrics*, 225(2):200–230.

Campbell, J. Y. and Mankiw, N. G. (1989). Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence. *Nber Macroeconomics Annual*, 4:185–216.

Carroll, C., Slacalek, J., Tokuoka, K., and White, M. N. (2017). The Distribution of Wealth and the Marginal Propensity To Consume. *Quantitative Economics*, 8(3):977–1020.

Chetty, R. (2006). A New Method of Estimating Risk Aversion. *American Economic Review*, 96(5):1821–1834.

Congressional Budget Office (2019). The Budget and Economic Outlook: 2019 to 2029. Technical report, Congressional Budget Office.

Conley, T. G. and Topa, G. (2002). Socio-Economic Distance and Spatial Patterns in Unemployment. *Journal of Applied Econometrics*, 17(4):303–327.

Czerwinski, S. J. and McCool, T. J. (2011). State and Local Governments: Knowledge of Past Recessions Can Inform Future Federal Fiscal Assistance. *U.S. Government Accountability Office*.

Dabla-Norris, E. and Lima, F. (2023). Macroeconomic Effects of Tax Rate and Base Changes: Evidence from Fiscal Consolidations. *European Economic Review*, 153:104399.

de Chaisemartin, C. and D'Haultfoeuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

de Chaisemartin, C. and D'Haultfoeuille, X. (2022). Difference-in-Differences Estimators of Intertemporal Treatment Effects. arXiv:2007.04267 [econ].

DeBacker, J., Evans, R. W., and Phillips, K. L. (2019). Integrating microsimulation models of tax policy into a DGE macroeconomic model. *Public Finance Review*, 47(2):207–275.

Easterly, W. and Rebelo, S. (1993). Fiscal Policy and Economic Growth. *Journal of Monetary Economics*, 32(3):417–458.

Embree, R. (2023). Revealing Values: Applying the Inverse-Optimum Method To Us State Taxes. *National Tax Journal*, 76(1):37–61.

Employment and Training Administration (2018). Comparison of State Unemployment Laws - Monetary Entitlement. Technical report, United States Department of Labor.

Fernald, J. (2014). A Quarterly, Utilization-Adjusted Series on Total Factor Productivity.

Flood, L., Hansen, J., and Wahlberg, R. (2004). Household Labor Supply and Welfare Participation in Sweden. *Journal of Human Resources*, 39(4):1008–1032.

Flood, Sarah, M. K. R. R. S. R. and Warren, J. R. (2018). Integrated Public Use Microdata Series, Current Population Survey: Version 6.0. Technical report, Institute For Social Research and Data Innovation.

Food and Nutrition Service (2018a). Supplemental Nutrition Assistance Program (SNAP) National and/or State Level Monthly and/or Annual Data. Technical report, United States Department of Agriculture.

Food and Nutrition Service (2018b). WIC Program Monthly Data – State Level Participation by Category and Program Costs. Technical report, United States Department of Agriculture.

Georgia Center for Opportunity (2022). Benefits Cliffs Workbook. Technical report, www.benefitscliffs.org.

Goodman-Bacon, A. (2021). Difference-in-Differences With Variation in Treatment Timing. *Journal of Econometrics*, 225(2):254–277.

Gordon, R. H. (1983). An Optimal Taxation Approach To Fiscal Federalism. *The Quarterly Journal of Economics*, 98(4):567–586.

Gordon, R. H. (2023). Fiscal Federalism and Role of the Income Tax. Technical report, National Bureau of Economic Research.

Gordon, R. H. and Cullen, J. B. (2012). Income Redistribution in a Federal System of Governments. *Journal of Public Economics*, 96(11-12):1100–1109.

Gouveia, M. and Strauss, R. P. (1994). Effective Federal Individual Income Tax Functions: An Exploratory Empirical Analysis. *National Tax Journal*, 47(2):317–339.

Gramlich, E. (1997). Migration and Income Redistribution Responsibilities. In *Financing Federal Systems*, pages 286–308. Edward Elgar Publishing.

Gruber, J. and Saez, E. (2002). The Elasticity of Taxable Income: Evidence and Implications. *Journal of Public Economics*, 84(1):1–32.

Haider, S. J. and Loughran, D. S. (2008). The Effect of the Social Security Earnings Test on Male Labor Supply New Evidence from Survey and Administrative Data. *Journal of Human Resources*, 43(1):57–87.

Hamersma, S. (2013). The Effects of Medicaid Earnings Limits on Earnings Growth Among Poor Workers. *The Be Journal of Economic Analysis & Policy*, 13(2):887–919.

Hassett, K. a. and Mathur, A. (2015). A Spatial Model of Corporate Tax Incidence. *Applied Economics*, 47(13):1350–1365.

Heathcote, J., Storesletten, K., and Violante, G. L. (2008). Insurance and Opportunities: a Welfare Analysis of Labor Market Risk. *Journal of Monetary Economics*, 55(3):501–525.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal Tax Progressivity: An Analytical Framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Heer, B. and Rohrbacher, S. (2021). Endogenous Longevity and Optimal Tax Progressivity. *Journal of Health Economics*, 79:102515.

Hendren, N. (2020). Measuring Economic Efficiency Using Inverse-Optimum Weights. *Journal of Public Economics*, 187:104198.

Hernanz, V., Malherbet, F., and Pellizzari, M. (2004). Take-Up of Welfare Benefits in OECD Countries: a Review of the Evidence.

Internal Revenue Service (2024).

Jappelli, T. and Pistaferri, L. (2014). Fiscal Policy and Mpc Heterogeneity. *American Economic Journal: Macroeconomics*, 6(4):107–136.

Jordà, O. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182.

Kaplan, G. and Violante, G. L. (2018). Microeconomic Heterogeneity and Macroeconomic Shocks. *Journal of Economic Perspectives*, 32(3):167–194.

Kawano, L. and Slemrod, J. (2016). How do Corporate Tax Bases Change When Corporate Tax Rates Change? With Implications for the Tax Rate Elasticity of Corporate Tax Revenues. *International Tax and Public Finance*, 23(3):401–433.

Kennan, J. and Walker, J. R. (2011). The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, 79(1):211–251.

Kierzkowski, H., Okólski, M., and Wellisz, S. H. (1993). *Stabilization and Structural Adjustment in Poland*. Routledge.

Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Themes in modern econometrics. Cambridge university press, Cambridge.

Killewald, A. (2016). Money, Work, and Marital Stability: Assessing Change in the Gendered Determinants of Divorce. *American Sociological Review*, 81(4):696–719.

Kimball, M. S. and Shapiro, M. D. (2008). Labor supply: Are the income and substitution effects both large or both small? Working Paper 14208, National Bureau of Economic Research.

Kleven, H. (2019). The EITC and the Extensive Margin: a Reappraisal. Technical report, National Bureau of Economic Research.

Kleven, H., Landais, C., Munoz, M., and Stantcheva, S. (2020). Taxation and Migration: Evidence and Policy Implications. *Journal of Economic Perspectives*, 34(2):119–142.

Kleven, H. J. and Waseem, M. (2013). Using Notches To Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *The Quarterly Journal of Economics*, 128(2):669–723.

Ko, W. and Moffitt, R. A. (2024). Take-Up of Social Benefits. *Handbook of Labor, Human Resources and Population Economics*, pages 1–42.

Krueger, D., Mitman, K., and Perri, F. (2016). Macroeconomics and Household Heterogeneity. In *Handbook of Macroeconomics*, volume 2, pages 843–921. Elsevier.

Lee, Y. and Gordon, R. H. (2005). Tax Structure and Economic Growth. *Journal of Public Economics*, 89(5-6):1027–1043.

Li, W. and Sarte, P.-D. (2004). Progressive Taxation and Long-Run Growth. *American Economic Review*, 94(5):1705–1716.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., and Sanghvi, S. (2017). Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation. *Mckinsey Global Institute*, 150(1):1–148.

McClelland, R. and Mok, S. (2012). A Review of Recent Research on Labor Supply Elasticities.

McKay, A. and Reis, R. (2021). Optimal Automatic Stabilizers. *The Review of Economic Studies*, 88(5):2375–2406.

Mendoza, E. G., Milesi-Ferretti, G. M., and Asea, P. (1997). On the Ineffectiveness of Tax Policy in Altering Long-Run Growth: Harberger'S Superneutrality Conjecture. *Journal of Public Economics*, 66(1):99–126.

Mertens, K. and Montiel Olea, J. L. (2018). Marginal Tax Tates and Income: New Time Series Evidence. *The Quarterly Journal of Economics*, 133(4):1803–1884.

Mertens, K. and Ravn, M. O. (2012). Empirical Evidence on the Aggregate Effects of Anticipated and Unanticipated Us Tax Policy Shocks. *American Economic Journal: Economic Policy*, 4(2):145–181.

Mertens, K. and Ravn, M. O. (2013). The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States. *American Economic Review*, 103(4):1212–1247.

Mertens, K. and Ravn, M. O. (2014). A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers. *Journal of Monetary Economics*, 68:S1–S19.

Mieszkowski, P., Bolkowiak, I., Lubick, D., and Sochacka-Krysiak, H. (2013). Tax Reform. In *Stabilization and Structural Adjustment in Poland*, pages 92–111. Routledge.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies*, 38(2):175–208.

Moffitt, R. A. (2002). Welfare Programs and Labor Supply. *Handbook of Public Economics*, 4:2393–2430.

Moore, R. and Pecoraro, B. (2020a). Dynamic Scoring: An Assessment of Fiscal Closing Assumptions. *Public Finance Review*, 48(3):340–353.

Moore, R. and Pecoraro, B. (2020b). Macroeconomic Implications of Modeling the Internal Revenue Code in a Heterogeneous-Agent Framework. *Economic Modelling*, 87:72–91.

National Conference of State Legislatures (2019). Moving on Up: Helping Families Climb the Economic Ladder By Addressing Benefits Cliffs.

Nichols, A. and Rothstein, J. (2015). The Earned Income Tax Credit. In *Economics of Means-Tested Transfer Programs in the United States, Volume 1*, pages 137–218. University of Chicago Press.

Pancrazi, R. and Vukotic, M. (2011). Overlooking the Great Moderation: Consequences For Monetary Policy. *Aea 2012 Annual Meetings, Chicago, Il*.

Plagborg-Møller, M. and Wolf, C. K. (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica*, 89(2):955–980.

Ramey, V. A. (2019). Ten Years After the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research? *Journal of Economic Perspectives*, 33(2):89–114.

Reichling, F. and Whalen, C. (2012). Review of estimates of the Frisch elasticity of labor supply.

Robins, P. K. (1985). A Comparison of the Labor Supply Findings from the Four Negative Income Tax Experiments. *Journal of Human Resources*, pages 567–582.

Roll, S. and East, J. (2014). Financially Vulnerable Families and the Child Care Cliff Effect. *Journal of Poverty*, 18(2):169–187.

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.

Ruh, P. and Staubli, S. (2019). Financial Incentives and Earnings of Disability Insurance Recipients: Evidence from a Notch Design. *American Economic Journal: Economic Policy*, 11(2):269–300.

Saez, E. (2001). Using Elasticities To Derive Optimal Income Tax Rates. *The Review of Economic Studies*, 68(1):205–229.

Saez, E. (2002). Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses. *The Quarterly Journal of Economics*, 117(3):1039–1073.

Saez, E., Slemrod, J., and Giertz, S. H. (2012). The Elasticity of Taxable Income With Respect To Marginal Tax Rates: a Critical Review. *Journal of Economic Literature*, 50(1):3–50.

Sallee, J. M. and Slemrod, J. (2012). Car Notches: Strategic Automaker Responses To Fuel Economy Policy. *Journal of Public Economics*, 96(11-12):981–999.

Schmidheiny, K. and Siegloch, S. (2023). On Event Studies and Distributed-Lags in Two-Way Fixed Effects Models: Identification, Equivalence, and Generalization. *Journal of Applied Econometrics*, page jae.2971.

Slemrod, J. (2013). Buenas Notches: Lines and Notches in Tax System Design. *Ejtr*, 11:259.

Stenkula, M., Johansson, D., and Du Rietz, G. (2014). Marginal Taxation on Labour Income in Sweden from 1862 To 2010. *Scandinavian Economic History Review*, 62(2):163–187.

Stock, J. H. and Watson, M. W. (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit and Banking*, 39:3–33.

Sun, L. and Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies With Heterogeneous Treatment Effects. *Journal of Econometrics*, 225(2):175–199.

Suárez Serrato, J. C. and Zidar, O. (2016). Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms. *American Economic Review*, 106(9):2582–2624.

Tax Foundation (2018). State Individual Income Tax Rates and Brackets. Technical report, www.taxfoundation.org.

United States Census Bureau (2018). Quarterly Summary of State Local Tax Revenue. Technical report.

Ventura, G. (1999). Flat Tax Reform: a Quantitative Exploration. *Journal of Economic Dynamics and Control*, 23(9-10):1425–1458.

Widmalm, F. (2001). Tax Structure and Growth: Are Some Taxes Better than Others? *Public Choice*, 107(3/4):199–219.

Xing, J. (2012). Tax Structure and Growth: How Robust is the Empirical Evidence? *Economics Letters*, 117(1):379–382.