ABSTRACT

Title of dissertation: ESSAYS ON UNINTENDED EFFECTS OF GOVERNMENT POLICY

Lucas William Goodman
Doctor of Philosophy, 2018

Dissertation directed by: Professor Melissa Kearney
Department of Economics

Public finance economics is often concerned with the "unintended consequences" of government policies, as an evaluation of these unintended effects is required to obtain a complete picture of the efficacy of a given policy. In my dissertation, I study three government policies, along with their potential unintended effects. I find that, in general, the potential unintended effects that I study are economically small. These findings suggest that policymakers in these areas can focus relatively more on the direct, intended effects of these policies when evaluating their costs and benefits.

In the first chapter, I study an expansion of Medicaid which occurred in 2014 as part of the Affordable Care Act. Because this expansion was only taken up by approximately half of all U.S. states, migration across state lines is a potential unintended consequence of these reforms. I analyze data from the American Community Survey and find that this migration response was not large: I estimate that migration could have increased Medicaid rolls in

expansion states by no more than 2 percent.

In the second chapter, Chris Boone, Arindrajit Dube, Ethan Kaplan, and I study extensions of unemployment insurance (UI) in the U.S. during the Great Recession. The "moral hazard" effect of UI on individual search effort has been well-studied. In this chapter, we broaden these potential unintended effects to include all potential effects (positive and negative) on aggregate employment at the county level. Again, we find little effect: we can rule out negative effects of the extensions in excess of -0.3 percentage points of the employment-to-population ratio.

Finally, in the third chapter, I use administrative tax data to study a tax reform in the state of Kansas which was designed to provide tax relief to businesses. This reform also increased an incentive for owners of a subset of these businesses to reclassify wage income as the profits of the firm – potentially reducing tax revenue. I again find that the effect was small. I show theoretically why this increase in incentive may not have caused a large response, even if the introduction of the incentive caused a much bigger response.

ESSAYS ON UNINTENDED EFFECTS OF GOVERNMENT POLICY

by

Lucas William Goodman

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Advisory Committee:
Professor Melissa Kearney, Chair/Advisor
Professor Judith Hellerstein
Professor Ethan Kaplan
Professor Phillip Swagel
Professor Lesley Turner

ADDITIONAL COPYRIGHT NOTICE

DISCLAIMER

This research embodies work undertaken for the staff of the Joint Committee on Taxation, but as members of both parties and both houses of Congress comprise the Joint Committee on Taxation, this work should not be construed to represent the position of any member of the Committee.

Preface

In my dissertation, I write three essays that study the effects of a different government policy – along a dimension which is not the explicit goal of that policy. Each of the policies and effects are distinct, making each chapter fairly independent of each other chapter. However, these chapters are united by the goal of constructing a credible identification strategy in order to estimate a parameter than can be connected to economic theory and government policy. They also happen to be united in their findings of null effects.

In the first chapter, I estimate the migration effects of the Medicaid expansions that occurred in some states in 2014, in connection with the Affordable Care Act. While there is a deep literature studying the migration responses to means-tested welfare benefits in general, this natural experiment is unique in that it creates larger than usual variation across state lines, with panel variation. I use the American Community Survey and find that migration from non-expansion states to expansion states did not increase in 2014, relative to migration in the opposite direction. This finding has a key implication for the fiscal federalism of Medicaid: it suggests that migration will not cause a "race to the bottom." This chapter was previously published in the *Journal of Policy Analysis and Management* (Goodman (2017)).

In the second chapter, Arindrajit Dube, Chris Boone, Ethan Kaplan, and I study the employment effects of the substantial extension in unemployment insurance (UI) during the Great Recession, with maximum benefit eligibility increasing generally from 26 weeks to up to 99 weeks. This extension could potentially reduce search effort of recipients – an effect that is well-studied in this literature – reducing aggregate employment. On the other hand, UI extensions can also act as a fiscal stabilizer, increasing aggregate demand in slack periods. The total effect is ambiguous. We exploit panel variation and find a null effect: in our most precise specifications, we can rule out effects more negative than 0.3 percentage points of the employment-to-population ratio. Based on these estimates, we conclude that it is either the case that search effects are small during a slack labor market (as predicted by Michaillat (2012), among others) or the fiscal multiplier is large.

In the third chapter, I use administrative tax data to study a tax reform in the state of Kansas, known as HB 2117. Under pre-reform law, owners of certain businesses known as S corporations faced a tax incentive to reduce the wages they pay themselves – policed by an IRS requirement to pay themselves "reasonable compensation." HB 2117 enhanced this incentive. I empirically analyze this reform and find that S corporation shareholders did in fact reduce the wages that they paid themselves – but only by about 3 percent. This effect is much less than the very large effects in the literature (e.g., Auten, Splinter & Nelson (2016)) which studies the *introduction* of this tax incentive. I use standard theories of tax evasion (with a small modification) to show that there are several theoretical "forces" which could cause the magnitude of the effect to fall when a tax incentive already exists in the baseline.

Each paper faces their own empirical difficulties. The empirical setting in the first chapter is perhaps the most straightforward, but nevertheless requires a careful consideration of the different migration flows that could be affected. The main specification is effectively a difference-in-differences: it estimates the increase in migration from non-expansion states to expansion states relative to migration in the opposite direction. Importantly, this specification captures the sum of two potential effects: the "attractive" effect (the increase in migration from non-expansion to expansion states) and the "retentive" effect (the decrease in migration from expansion to non-expansion states). Both of these effects push in the same direction. If I had estimated a nonzero regression result, I would have not have been able to separate the two effects. However, given that I estimate a null effect, I can say with a reasonable amount of confidence that both effects are zero.

In the second chapter, the central empirical challenge is the mechanical endogeneity of the treatment variable – the UI extensions. By law, the benefit length was set at the state-week level to be a function of the state-level unemployment rate. Thus, employment rates would tend to covary with benefit lengths even in the absence of any causal effect of benefit effects on employment. Our empirical strategy tries to address this endogeneity by studying county pairs that straddle state borders. We show that this mitigates, but does not eliminate, the endogeneity problem. We then restrict our sample of pairs to those that appear to have a good match prior to treatment, and show suggestive evidence that this strategy reduces endogeneity even more.

In the third chapter, obtaining the point estimate is more straightforward, though still not without its complications. The reform happened in a given state in a given year, meaning that difference-in-differences methods are an attractive strategy. In the paper, I try to mitigate some of the standard threats to the parallel trend assumption at the heart of the difference-in-differences strategy by making the treatment and control samples look similar based on observable covariates, and by using the synthetic control method. The bigger challenge in this paper – especially given that I estimate a robust null result – is estimating the proper confidence intervals. I show why confidence intervals constructed using analytical standard errors are likely to miss a substantial share of the uncertainty. I use a randomization inference method to compare the estimated effect in the treatment state (Kansas) relative to the estimated effect in each other state which we know not to have been treated. I show empirically that estimators that are more effective at creating matches between treatment and control are also more precise when uncertainty is measured using this method.

Furthermore, there have been developments (either policy changes or data availability) that affect both the first and third chapter. A substantial shortcoming in the first chapter (which is a published paper in the *Journal of Policy Analysis and Management*) is data availability: at the time, I was able to access the relevant data through 2014 only – the first year of "treatment." Thus, I might have been missing longer term effects. In this dissertation, I have left the text of the original published article unchanged (with very few exceptions – e.g., to update a reference to a more recent version of the same paper), and have instead added an addendum. In this addendum, I estimate models which make use of two additional years of data (and which account for additional states that have taken up the Medicaid expansion). These additional data confirm the original result: that the migration effect was small, if anything.

With respect to the third chapter, there has additionally been a major policy development: the passage of the bill colloquially referred to as the Tax Cuts and Jobs Act (TCJA) in late December 2017.[1] The TCJA introduced Section 199A to the Internal Revenue Code; under Section 199A, taxpayers may generally claim a 20 percent deduction for "qualified business income," subject to various restrictions. This provision reduces the effective tax rate on that type of income; this further has the effect of increasing the wedge between wages and profits for owners of S corporations, which is the margin that I study in this chapter. In other words, the TCJA made this chapter substantially more policy-relevant. In this chapter, I describe at a high level the potential implications of my results for what we might expect in response to the TCJA. However, there are some key differences between the variation that I study and the variation created by TCJA. For instance, this deduction under the TCJA will not be available for certain firms, depending on (among other things) the income earned by their owners. Given time constraints (and constraints related to avoiding political controversy), I have not adjusted the empirical strategy to more directly attempt to estimate the expected effects

[1]For procedural reasons, the official name of this bill (Public Law 115-97) was changed to "An act to provide for reconciliation pursuant to titles II and V of the concurrent resolution on the budget for fiscal year 2018."

of the TCJA. This might be a fruitful area for future work.

Acknowledgments

There are many people without whom I would not have been able to begin, let alone complete, my dissertation. I am grateful to my advisors who have given me detailed guidance on all three chapters of my dissertation, and helped guide me through my entire graduate school career. Melissa Kearney always kept me focused on the big picture, even when the minute details seemed overwhelming. Melissa was also instrumental in helping me find an appropriate outlet to publish a version of the first chapter of my dissertation; this turned out be an amazing fit, and I would not have known where to look for it. I am also grateful to Lesley Turner, who always made time to give me detailed feedback on early drafts and under-developed research ideas – and supported me as I proposed in her public finance course a version of the third chapter of my dissertation, even when the probability of my accessing the necessary data was remote. I also appreciate the subtly different "labor economics" perspective that Judy Hellerstein always brought to my papers and presentations.

I am especially indebted to Ethan Kaplan, who hired me as a research assistant beginning in the summer after my first year, continuing on until the end of my third year. On a very surface level, this relationship was productive in that it produced the second chapter of my dissertation. But more deeply than that, this experience introduced me to the "soft" skills needed to produce successful research – and gave me a substantial head start on the "hard" programming and project organization skills that I used in each chapter of my dissertation. Moreover, I am grateful to Ethan for his continued feedback on my other chapters and other research ideas even after the formal “research assistant” relationship came to an end.

I am also grateful to my colleagues at the Joint Committee on Taxation. First and foremost, I owe my job at JCT to Jake Mortenson, who befriended me and connected me to the senior JCT staff. I thank Tim Dowd for being my first “boss” at JCT and introducing me to the Committee and to the type of projects that they work on. I thank the rest of the staff for giving me feedback – both “big picture” and “in the weeds” – on the third chapter of my dissertation. These people include, but are not limited to: Tom Barthold, Nick Bull, Tim Dowd, Heather Harman, Jamie McGuire, Jake Mortenson, and David Splinter. I am also grateful to other members of the staff for whom I worked, helping me gain familiarity with the tax data, including Heidi Schramm, Kathleen Mackie, and Jim Cilke.

I am also indebted to various people outside Maryland and JCT. Along with Ethan Kaplan, I co-authored the second chapter of my dissertation with Arindrajit Dube and Chris Boone. Arin’s vision helped drive the project, and Chris was always there to help me code a complicated estimator if I needed help. Additionally, we received helpful comments from Gabriel Chodorow-Reich, Ioana Marinescu, and Jesse Rothstein with respect to this paper, and Bryan Hardy and Yuting Huang provided research assistance at various stages of this project. With respect to the third chapter of my dissertation, I also received feedback from Owen Thompson and Danny Yagan, and the many attendees of my presentation at the Tax Economists Forum.

I also received substantial support from my peers. Riley Wilson, Thomas Hegland, Heath Witzen, and Ken Coriale were always available to talk about research ideas – no matter how undeveloped or how low the ceiling. I also thank the entire applied micro student body for the very pointed questions and comments I received during brown-bag presentations. These comments always helped me focus the next of round of edits.

I also wish to thank Robert Pozen, for whom I was a research assistant in the two years prior to my admission at the University of Maryland. Bob got me interested in tax policy, by asking me to help him design a tax course to teach to students at Harvard Business School. He also pushed hard for me to apply for a doctoral program in economics in the first place – so, quite literally, this dissertation would not have been written but for Bob. Going back even further than that, I am also grateful for my high school and college rowing coaches, Ed Maxwell and Tony Kilbridge. During my tenure at UMD, my mind would turn to their confidence-inspiring pre-race wisdom whenever I was nervous before a big presentation, exam, or interview.

Last but certainly not least, I am deeply grateful for the unconditional love and support of my husband, Cormac. He is always there to listen to my practice presentations, even if all he could do was smile and nod. He also has the impeccable ability of knowing when I am "spinning my wheels." In these many occurrences, he always had the wisdom to tell me stop working and take a break – even if that's not what I want to hear. I would not have survived

this experience otherwise.

Contents

List of Tables

List of Figures

Chapter 1

The Effect of the Affordable Care Act Medicaid Expansion on Migration

1.1 Introduction

Fiscal federalism, through which some social programs are run and funded by states and lower levels of government, is a key element of the Affordable Care Act. With the help of federal funds, the ACA asks states to operate health insurance marketplaces, as well as to expand Medicaid for a large number of non-disabled adults. As a general matter, fiscal federalism allows local programs to align more closely with local preferences. At the same time, fiscal federalism can lead to suboptimal levels of public services if states do not internalize the effect of their policy on other states. In particular, if state policymakers believe that a decrease in means-tested benefits such as Medicaid will lead to out-migration of program recipients — increasing net expenditures in other states — states will tend to provide benefits that are too low relative to the social optimum.[1] In equilibrium, such underprovision is more severe if the migration responsiveness of potential beneficiaries is larger. Yet, the empirical extent of migration in response to means-tested benefits remains very much an open question.

This paper studies the migration effects of the 2014 Medicaid expansion brought about by the ACA, as "amended" by the Supreme Court. This policy environment generally provides for stark differences in Medicaid eligibility between 2013 and 2014 for some states ("expansion states"), but not for others ("non-expansion states"). The potential gain to migrants from non-expansion states could potentially be quite large. The Department of Health and Human Services (2014) estimates that per-beneficiary Medicaid expenditures were approximately \$5,500 for the newly-eligible in 2014. As a comparison, 138% of the federal poverty level was equal to \$27,310 for a family of three in 2014.

Despite these large potential gains, this paper estimates that the migration effect of Medicaid is very close to zero. Furthermore, in the primary specification, the statistical precision is sufficient to rule out (at 95% confidence) migration effects that would have a meaningful effect on the size of the Medicaid-eligible population in other states. This suggests that any fiscal ex-

[1] Seminal contributions to the study of fiscal federalism and competition between jurisdictions include Brown & Oates (1987), Wildasin (1991), Oates (1999), Brueckner (2000), Saavedra (2000), and Baicker (2005).

ternality (through Medicaid expenditures or expenditures on other programs) from the Medicaid expansion decision is quite small. Additionally, this paper finds no evidence of a migration effect even in subgroups one might expect *a priori* to be more responsive, including those in worse health and those living closer to a border between an expansion state and a non-expansion state.

There are several reasons why it is important to study the migration effects of these Medicaid expansions. First, *these particular* Medicaid expansions are of first-order policy relevance. The Congressional Budget Office (2016) estimates that the average stock of newly eligible enrollees was 6.1 million in 2014 and 9.6 million in 2015. Furthermore, state policymakers continue to debate whether they should expand Medicaid (or undo their previous expansion), perhaps taking migration considerations into account. The magnitude of the migration response is relevant to these decisions. Second, this policy variation provides a unique setting to study migration induced by means-tested benefits more generally. I contribute to this literature by studying policy variation which is large and varies across time and space. Third, there is a lack of rigorous evidence on the effect of means-tested health benefits on migration (Schwartz & Sommers (2014) being a notable exception). Fourth, the migration effects of Medicaid are relevant to a large literature that exploits variation in Medicaid eligibility across states over time, since this literature implicitly assumes that migration effects are small (e.g., Currie & Gruber (1996a, 1996b), Gross & Notowidigdo (2011), Cohodes, et al. (2016), Cutler & Gruber (1996), Lo Sasso & Buchmueller (2004), and Gruber & Simon (2008)).

To study the effect of Medicaid expansion on migration, this paper uses public use microdata from the American Community Survey (ACS). The ACS asks respondants about their current residential location as well as their location 12 months prior to interview, which I observe at a relatively coarse level discussed below. Ultimately, this paper finds that the Medicaid expansion did not appear to affect individuals' migration decisions to an extent that would be economically meaningful. In the baseline specification, the point estimate is very close to zero; the upper bound of the associated 95% confidence interval corresponds to an increase in the migration rate from non-expansion to expansion states of less than 0.18 percentage points from a base of 1.06%. Such a migration response would have only a trivial effect on the Medicaid-eligible population of expansion states, suggesting that migration concerns should not play a large role in policymakers' expansion decisions. Analysis of state-year variation in Medicaid eligibility more generally, including variation in the years prior to the 2014 expansions, is consistent with the possibility that individuals motivated to migrate to obtain Medicaid benefits had already done so prior to 2014. However, these results are not statistically significant.

To my knowledge, there is only one other paper, Schwartz & Sommers (2014), which examines the migration response to public health care expansions. They look at expansions that took place in Arizona, Maine, and New York in the early 2000s, as well as the rollout of health reform in Massachusetts in 2006/2007. Using the (March) Current Population Survey, they use a difference-in-differences approach, examining inflows into and outflows from these states, relative to a set of one or two control states for each treatment state. They find results that are insignificantly different than zero, and in fact of the opposite sign that would be expected if public health care expansions led to increased in-migration and decreased out-migration. The research environment studied in this paper improves on that of Schwartz & Sommers (2014) in that the ACA Medicaid expansion is much more widespread: the Medicaid expansion in 2014 took place in just over half of the 51 states (including the District of Columbia), rather than four in the setting of Schwartz & Sommers (2014). This policy environment allows for a "region-based" analysis which Schwartz & Sommers (2014) could not have performed. Furthermore, the ACS which I use has a substantially larger sample size than the March CPS that Schwartz & Sommers (2014) use.[2] As a result, while none of the specifications of this paper are perfectly analogous to the main specification of Schwartz & Sommers (2014), the standard errors of this paper (when translated appropriately) are roughly 30% smaller than those of Schwartz & Sommers (2014).[3]

More generally, this paper fits into the deeper literature on welfare-induced migration. One subset of this literature studies across-state variation in the generosity of cash welfare payments, mostly Aid for Families with Dependent

[2]Schwartz & Sommers (2014) could not have used the ACS to study (most of) their variation, as the ACS has existed in its current form only since 2005.

[3]Furthermore, while the standard errors of this paper and those of Schwartz & Sommers (2014) are both calculated via the standard "cluster-robust" formula (clustering at the state level), this paper generally uses 51 clusters while Schwartz & Sommers (2014) uses only 10 clusters. As discussed by Cameron & Miller (2015), the cluster-robust formula estimates the standard error with a downward bias which is exacerbated when the number of clusters is "few." While there is no "magic number" defining what it means to be "few," Cameron & Miller (2015) discuss simulation evidence that the cluster-robust formula leads to much less over-rejection with 50 clusters than with 10.

Children (AFDC), often using microdata from the Decennial Censuses in 1980 and/or 1990; see Meyer (2000) for a detailed review of this early literature. More recently, Gelbach (2004) finds suggestive evidence that never-married high school dropouts who move interstate tend to move to states with higher welfare benefits, relative to various control groups. Bailey (2005), using a rich discrete choice conditional logit model, finds somewhat larger effects of welfare generosity on migration. McKinnish (2005) and McKinnish (2007) find that AFDC expenditures are more highly concentrated just on the high side of the border between a high-benefit and a low-benefit state, relative to the rest of the high-benefit state, suggesting that welfare beneficiaries are migrating as short a distance as possible in order to obtain higher welfare benefits. Borjas (1999) finds that immigrants to the United States tend to locate in states with more generous cash welfare systems.

There are also two other papers more closely related to the present paper in that they exploit a policy change. Kaestner, Kaushal & Van Ryzin (2003) examine the effect of state-specific welfare reforms in the 1990s (in the run-up to nationwide major welfare reforms in 1996) which generally subjected cash assistance to various restrictions. They find that intrastate migration increases in response to welfare reform (likely reflecting increased job search) but *interstate* migration falls slightly, which is the opposite sign one might expect if individuals were migrating to take up more generous welfare benefits. Fiva (2009) studies a national welfare reform in Norway in 2001 which made local welfare policies more uniform, causing larger increases in generosity in localities which were previously less generous. Using a difference-in-differences instrumental variables design, he finds a substantial migration response to the change in benefits among low skill unmarried men — enough that a 10% increase in benefits would lead to a 4% increase in the welfare-receiving population.

Results from the literature on welfare-induced migration need not describe the migration effects of Medicaid expansions under the ACA, however. It is possible that the potential gain from acquiring Medicaid coverage is large enough to overcome adjustment costs, in contrast to relatively smaller differences in AFDC generosity, making migration effects larger. On the other hand, migration effects from the Medicaid expansion could be smaller if individuals do not place a high value Medicaid coverage (as suggested by Finkelstein, Hendren, & Luttmer (2015)), if Medicaid is valued only by a small number of sicker individuals, or if residents of non-expansion states are unaware of the expanded eligibility in expansion states. This paper, by contrast, will provide direct evidence on the effect of Medicaid expansions on migration.

The remainder of this paper is organized as follows. Section II describes the policy environment. Section III describes the data source, the American Community Survey. Section IV describes the empirical strategy and presents the results. Section V performs sensitivity analyses. Section VI concludes.

1.2 Medicaid Expansions in the ACA

In March 2010, President Barack Obama signed two pieces of legislation which together are known colloquially as the Affordable Care Act, hereafter the ACA. Among other legislative goals, the ACA was designed to increase health insurance coverage in two main ways, most of which took effect in the year 2014. First, it created insurance marketplaces complete with the "three-legged stool" of guaranteed issue (i.e., no denial of coverage due to pre-existing conditions, combined with community rating of premiums), an individual mandate to purchase health insurance, and subsidies to help low-income individuals purchase insurance. Second, it provided substantial funding for states to expand their Medicaid program to cover all individuals up to 138% of the federal poverty guidelines. The federal government would provide 90% of the incremental funding for this Medicaid expansion, with states covering the remaining 10%.[4]

In the original text of the ACA, the federal government further incentivized states to take up the Medicaid expansion by conditioning *existing* Medicaid payments on the expansion. In the Supreme Court case *National Federation of Independent Business (NFIB) v. Sebelius*, the Court ruled *inter alia* that this type of "coersion" of state governments by the federal government was unconstitutional. To remedy this constitutional deficiency, the Court ruled that federal government could not take away existing Medicaid funding if that state elected not to adopt the Medicaid expansion. As of early 2014, 24 states

[4]For the years 2014-2016, the federal government provides 100% of the incremental funding; this match amount decreases gradually to 90% in 2020.

chose not to adopt the Medicaid expansion. These states are shown in **Figure 1.1**.[5] The decision to expand Medicaid or not seems most closely related to the political beliefs held by the voters and the governor of each state; of the 24 non-expansion states, 17 voted for Senator John McCain in the 2008 presidential election and 21 were headed by Republican governors when *NFIB* was decided. The idiosyncratic role of governors means that expansion status is not a neat divide between states that are traditionally liberal and conservative. Governor Steve Beshear (D, Kentucky) helped push through the Medicaid expansion in his otherwise conservative state, while Governors Paul LePage (R, Maine) and Tom Corbett (R, Pennsylvania) were able to prevent the Medicaid expansions in their more left-leaning states (at least temporarily, in Corbett's case).

In non-expansion states, individuals between 100% and 138% of the federal poverty guidelines—who would be eligible for Medicaid had the state expanded Medicaid—are eligible for premium subsidies in the insurance marketplaces. These are somewhat less generous than Medicaid because net premiums paid by the beneficiary are still positive (2% of income for a benchmark plan) and these plans include some element of cost-sharing, such as copays or coinsurance.[6] Individuals earning less than 100% of the federal poverty guidelines are not eligible for premium subsidies, since Congress expected that these individuals would be covered by Medicaid. As a result, this group falls into what has

[5]In 2015 and 2016, several additional states chose to adopt the Medicaid expansion. This paper defines expansion status as it existed in early 2014 because the data describe migration behavior through 2014 only.

[6]Most of these individuals would be eligible for cost-sharing subsidies which reduce, but do not eliminate, copays and coinsurance.

been dubbed the "Medicaid gap." Such individuals were exempted from the individual mandate. Furthermore, the set of expansion states is geographically dispersed. Of all of the non-expansion states, only five states in the southeast — North Carolina, South Carolina, Georgia, Florida, and Alabama — do not share a border with an expansion state. Similarly, of all of the expansion states, only California, Rhode Island, and Vermont do not share a border with a non-expansion state. Of the 831 Metropolitan Statistical Areas defined by the Office of Management and Budget, 25 straddle the border between an expansion state and a non-expansion state (Office of Management and Budget (2013)).[7] Additionally, states are not permitted to condition Medicaid eligibility on length of residence; similar laws in the context of AFDC were found unconstitutional in the Supreme Court case *Saenz v. Roe*.

Complicating matters somewhat, five expansion states (California, Connecticut, Minnesota, New Jersey, and Washington) and the District of Columbia initiated some elements of the Medicaid expansion prior to 2014. Massachusetts also implemented a health care reform in 2006 whereby (*inter alia*) individuals up to 150% of the poverty guidelines were offered fully subsidized health insurance. This suggests that treating these states as implementing the ACA Medicaid expansion in 2014 could lead to attenuation bias. However, as shown in Frean, Gruber, and Sommers (2017), the Medicaid expansion led to large increases in coverage in 2014 even among previously-eligible individuals, suggest-

[7]Examples of such MSAs include New York, Chicago, Philadelphia, Washington, St. Louis, Memphis, and Louisville.

ing a large salience or marketing effect of the 2014 expansions. Furthermore, of these six states and D.C., only D.C. and Massachusetts offered universal coverage throughout the state to virtually the entire population under 138% of poverty prior to 2014 (Kaiser Family Foundation (2016a, 2016b, and 2016c)). For these reasons, these six states and D.C. are treated in the same manner as all other expansion states – i.e., beginning "treatment" in 2014 – in the baseline analysis. However, in sensitivity analyses, the main specifications are re-estimated after dropping these states, with little effect on the results.

1.3 Data

This paper uses public use microdata from the American Community Survey (ACS) conducted in years 2005 through 2014, accessed via the Integrated Public Use Microdata Series (IPUMS) (Flood, et al. (2015)). The survey is a random sample of roughly 1% of all households in the United States in each year, with interviews taking place on a rolling basis throughout the year. Most importantly for the research question at hand, the ACS questionnaire asks about the residential location of each individual one year prior to interview as well as the location of current residence. To prevent disclosure of personal information, this location is reported in the public use samples at the "public use microdata area" (PUMA) level. A PUMA is a region constructed by the Census Bureau fully nested within states, with a population of at least 100,000. The location of prior residence — if the individual migrated — is aggregated at a slightly different level, known as a "PUMA of migration."

The analysis is performed at the individual level, although some variables are defined at the household or family level. A "family" is a set of persons constructed by IPUMS which is a subset of a given household.[8] This is a non-trivial specification choice: among families with at least one migrant, one-third had only certain members of the family migrate. A common example is an adult child moving back in with his or her parents. Another common example is an aging parent moving in with his or her adult children.

The rolling nature of the interviews presents two challenges. First, since the mean (and median) 2014 interview was conducted in June 2014, analysis of the 2014 interviews generally picks up only the very short-run effects, rather than migration effects that occur in the longer-run (e.g., after learning takes place). Second, 2014 interviews will pick up many moves that happen during calendar year 2013. I argue that these moves are rightfully considered part of the treatment effect, since individuals might move in anticipation of the expansion — especially those individuals living in very conservative or liberal states whose expansion decision was not in doubt. An implication of such anticipatory moves is that moves reported in the 2013 survey could *also* be part of the treatment effect, since the potential migration period overlaps with the 2014 survey. For this reason, observations from the 2013 survey are dropped in all regression analyses of the 2014 Medicaid expansion.

[8]Importantly, a cohabitating (non-married) partner will generally be assigned to a separate family unit within the household, which is consistent with Medicaid eligibility, which uses the tax unit as its household concept.

1.3.1 Selection of analysis sample

The analysis sample is selected in order to restrict attention to individuals who could potentially be affected by the Medicaid expansion if they lived in an expansion state. In the baseline specification, the analysis sample is chosen using an income-based selection criterion. In particular, individuals are included in the sample if their reported *family* income over the prior 12 months would have placed them below 138% of the federal poverty level, based on their family size.[9] One substantial concern about selecting the sample in this way is that income could be causally affected by state of residence. Meyer (2000) showed that this can be very problematic when analyzing cross-sectional migration flow. In particular, because the ACS asks about income received over the prior 12 months, the econometrician observes only some convex combination of a migrant's origin income and destination income. When looking at migration from a high-wage state (which may be correlated with being an expansion state), the econometrician would then tend to see outgoing migrants as having lower income than individuals who remained in the high-wage state. In other words, an individual who lived in a high-wage state 12 months prior

[9] While this measure is a reasonable proxy for (potential) Medicaid eligibility, it is not perfect, for several reasons. First, the family concept used for Medicaid eligibility is the tax unit — including the filer and anyone else (including a spouse) for whom he or she claims as a personal exemption. The family concept in the ACS is similar for most households, in that it considers all members who are related to each other as part of the same "family," but there is not enough information to know whether, e.g., an adult child can be claimed as a dependent. Second, Medicaid eligibility is based on monthly income, not annual. Furthermore, in practice, eligibility decisions are usually made with "projected income," which can presumably be manipulated to some degree. It is possible that manipulation of income could substitute for migration, but this is highly unlikely among childless couples, for whom the eligibility thresholds are very low or non-existent in non-expansion states.

to interview would be more likely to be in the analysis sample if he or she migrated to a low-wage state. This could bias upward the estimated migration flow from this high-wage state among the analysis sample and, by an analogous argument, bias downward the estimated migration flow from a low-wage state.

Fortunately, as discussed in more detail below, the empirical strategies of this paper have the spirit of a difference-in-differences. Thus, for the causal effect of residence on income to bias these results, these causal effects must be time-varying within a given state. In **Section 1.4.3**, I consider several possible channels of time-varying bias and present evidence that these channels do not appear to be playing a large role. Ultimately, though, this is an assumption that I must make. Furthermore, **Section 1.5** will explore specifications that select the sample on the basis of education level rather than income, which mitigates this source of potential bias, at the cost of making the sample less well targeted.

Additional sample restrictions are made in order to drop individuals who are mostly unaffected by the Medicaid expansion. First, all families in which all members are over the age of 65 are excluded, as such families would generally be covered by Medicare.[10] For a similar reason, the small number of families in which all members are 17 years of age or younger are also excluded. Furthermore, lawfully present immigrants are generally ineligible for Medi-

[10]While some "dual-eligible" Medicare beneficiaries are also covered by Medicaid, these beneficiaries were not affected by the expansions.

caid until five years after they receive permanent resident status (i.e., a green card); undocumented immigrants are permantly ineligible (except for some emergency services). However, neither the legal status of immigrants, nor the date of permanent resident status, is observable in the data. Thus, I proxy for ineligibility due to immigrant status by dropping all immigrants who arrived to the U.S. within five years of interview.[11]

To target the analysis to groups that might be most likely to respond to the Medicaid expansion, some specifications will restrict attention further to individuals in families without children. Children are generally not part of newly eligible population, because children in families with income less than 138% of poverty would typically be eligibile for Medicaid or CHIP even in non-expansion states. Yet, the presence of children (or an increase in family size, more generally) can substantially increase the costs of migration. Thus, it is plausible that any migration effects of the Medicaid expansions would be concentrated in the group of childless adults.

Table 1.1 presents summary statistics for the analysis sample, separately in expansion and non-expansion states (based on reported residence 12 months prior to interview). Characteristics are fairly well matched between the expansion and non-expansion states, except that expansion-state adults are slightly more likely to be male, have a bachelor's degree, be hispanic, or be an immi-

[11]Borjas (2017), following Passel and Cohn (2014), has developed a mechanism that attempts to identify undocumented immigrants in data sources such as the ACS. However, implementing such a procedure in this context would require selecting on certain observables, including Medicaid receipt, that could introduce compositional issues with implications that are not straightforward. For this reason, I do not implement this procedure here.

grant (although the difference is not statistically significant for the latter two). Furthermore, the sample is almost evenly split between those who live in an expansion state and those who live in a non-expansion state 12 months prior to interview.

To understand broad patterns of migration in the analysis sample, **Figure 1.2** plots the mean interstate migration rate of this sample as well as a sample of middle-income individuals (with earnings between 250% and 400% of poverty). The upper graph uses all observations from these two groups. Immediately, we see that the analysis sample has a higher migration rate than the higher-income sample, by a factor of about 50%. This is not explained by family structure, since this level difference exists in the childless sample (in the lower graph) as well.[12]

1.3.2 Medicaid eligibility thresholds

In most specifications, a state's Medicaid expansion status is treated as binary. Some specifications, however, use eligibility thresholds (expressed as a percentage of the federal poverty level) collected by Kaiser Family Foundation (2016a, 2016b, and 2016c). This data collection performed by Kaiser is no small task: eligibility limits — which also often varied by the working sta-

[12] This relationship between income and migration stands in contrast to literature which finds that individuals with higher education (and presumably higher permanent income) are more likely to migrate (Bound & Holzer (2000), Molloy, Smith, & Wozniak (2011)). One speculative explanation would be that negative income shocks, such as a job separation, tend to lead causally to migration. This channel would lead to larger estimates of the baseline migration rate, relative to an ideal sample which conditioned on permanent income. However, this bias is harmless to the empirical strategy unless the magnitude of this causal effect changes in 2014, differently in expansion and non-expansion states.

tus of the beneficiary — were not reported in a standardized manner. Thus, these thresholds cannot perfectly summarize the eligibility rules for all households. Furthermore, while Kaiser have collected eligibility limits for parents since 2002 (and children since 2000), they have collected eligibility limits for childless adults only since 2011.

1.4 Empirical Strategy and Baseline Results

In a standard setting in which migration is studied, one might estimate the effect of the treatment in question on inflows into and outflows from affected states. In the literature on welfare-induced migration, versions of this approach are undertaken by Fiva (2009), Gelbach (2004), Kaestner, et al. (2003), and Schwartz & Sommers (2014). In the context of the present paper, such an analysis would estimate whether inflows into expansion states increased, and outflows from expansion states decreased, in 2014 relative to inflows to and outflows from non-expansion states in 2014. However, two advantageous features of the 2014 Medicaid expansions allow for a refinement of this approach. First, the 2014 Medicaid expansion took place in roughly half of all states. Second, it is reasonable to model the Medicaid expansion as a uniform treatment in expansion states implemented all at once in those states.[13]

In particular, given this policy environment, it is possible to isolate the migration flows most plausibly affected by the Medicaid expansion: migration

[13]Sensitivity analyses will examine some relaxations of this specification choice, including dropping the six so-called "early expanders."

from the set of non-expansion states to the set of expansion states, as well as migration in the opposite direction.[14] In particular, the expansion of Medicaid could induce some individuals to move from non-expansion states to expansion states in 2014. The expansion could also reduce migration from expansion states to non-expansion states; i.e., some individuals who would have moved from an expansion state to a non-expansion state in the counterfactual of a uniform Medicaid expansion could be induced not to move, or induced to move to some other expansion state. The empirical strategy of this paper aims to test whether either of these effects exist.

To implement this empirical strategy, all 50 states and the District of Columbia are assigned in a time-invariant manner to two "regions," as illustrated in **Figure 1.1**: the set of expansion states and the set of non-expansion states. The dependent variable y_{irt}, measured at the individual level, is defined equal to one if the individual moved from one region to the other — that is, "inter-regionally" — in the 12 months prior to interview in year t. The individual's origin state is indexed by r. The variable $nonexp_r$ is a dummy for having an origin state in the non-expansion region. I.e., $nonexp_r$ equals one for the following types of individuals: (1) an individual who migrates from a non-expansion state to an expansion state, (2) an individual who lives in a non-expansion state and does not migrate (or migrates intrastate), and (3) an individual who migrates from one non-expansion state to another non-

[14]Meyer (2000) also examines inter-regional migration in the cross-section, where regions are defined relative to the median AFDC generosity.

expansion state. The variable $post_t$ is a dummy for 2014. Given these variable definitions, the following linear probability model is estimated:[15]

$$y_{irt} = \lambda_t + \mu_r + \beta \times nonexp_r post_t + \epsilon_{rt} \tag{1.1}$$

The coefficient β captures the extent to which non-expansion-to-expansion migration increased in 2014, relative to the increase in expansion-to-non-expansion migration in 2014. This regression has the spirit of a difference-in-differences regression, with a subtle caveat. In a typical difference-in-differences, one group is "treated" and another group is "untreated" (or "control"). In equation (1.1), non-expansion states are playing the role of "treated" and expansion states are playing the role of "untreated." However, the Medicaid expansion plausibly "treats" the migration from both expansion states and non-expansion states — but in opposite directions. Thus, β measures the sum of both effects: the increase in non-expansion-to-expansion migration and the decrease in expansion-to-non-expansion migration.

Because this regression is isomorphic to a difference-in-differences regression, a causal interpretation of this coefficient requires a standard parallel trend assumption. In particular, in the counterfactual of a uniform Medicaid expansion applying to all states, the assumption requires that non-expansion-to-expansion migration would have followed a trend parallel to expansion-to-

[15]The regressions are weighted by the ACS sample weights. As discussed above, observations from interview year 2013 are dropped. Standard errors are clustered at the level of the origin state. Note that this regression could equivalently be run at the origin state-year level if appropriately weighted.

non-expansion migration. This could be violated if, for instance, expansion states experienced greater economic growth in 2014, causing them to be a more attractive residential location. Further threats to this assumption are considered in **Section 1.4.3**.

Some specifications add control variables to the regression. The first set of control variables are for demographic characteristics of individuals: dummies for being male, non-Hispanic white, Hispanic, and the interaction between male and white, and male and Hispanic; a dummy for being an immigrant; a quadratic in age; dummies for being married; dummies for having 1, 2, 3, or 4 or more children; dummies for having an 8th grade education or less, being a high school dropout, and being a high school graduate (exactly); income as a proportion of poverty; and the log of the distance from the individual's PUMA (or PUMA of migration) to the nearest border between an expansion state and a non-expansion state. Although these demographic variables are roughly orthogonal to $nonexp_r post_t$, conditional on the state and time fixed effects, these controls can improve precision and potentially mitigate any effects from changes in the composition of the sample over time. Second, two unemployment rate controls are added: (1) the own-state unemployment rate and (2) the minimum state unemployment rate in the *opposite region* within 500 kilometers of the origin state.[16] To the extent that the unemployment rate accurately captures local economic conditions relevant to low-income individu-

[16]In the case of Alaska and Hawaii, this latter variable is set equal to the unemployment rate of the closest state in the opposite region (which is further than 500 kilometers).

als, these variables can control for true causal effects of time-varying economic conditions, and they can also control for any time-varying bias in measuring migration rates, as discussed above. Third, some specifications include state-specific trends to control for other forms of time-varying state heterogeneity.

1.4.1 Investigation of pre-existing differences and trends

Figure 1.3 provides a visual summary both of the pre-existing trends in inter-regional migration, as well as the main result of this paper. The Figure shows the mean inter-region migration rates of the analysis sample, where the dashed line and hollow circles represent expansion-to-non-expansion migration (i.e., the mean value of y_{irt} for individuals reporting an origin state in the expansion region in a given interview year) and the solid line and solid circles represent non-expansion-to-expansion migration (i.e., the mean value of y_{irt} for individuals with an origin state in the *non*-expansion region in a given interview year). The top graph, which reports the yearly means for all individuals in the analysis sample, shows that the trends looked different from each other prior to 2008, with a general decline from expansion states and a general increase from non-expansion states prior to 2008. After 2008, however, the trends look fairly parallel. Furthermore, the difference in means in 2014 looks roughly similar to the difference in means in previous years. The pattern among childless individuals (bottom graph) is qualitatively similar. Visually, there appears to be little evidence that Medicaid expansions caused increased migration from non-expansion to expansion states (or reduced migration in the opposite direction).

More formally, to analyze pre-existing trends, an event study regression of the following form is estimated on data from 2005 through 2012:[17]

$$y_{irt} = \lambda_t + \mu_r + \sum_{s \neq 2012} \beta_s nonexp_r * 1(year_t = s) + \epsilon_{rt} \tag{1.2}$$

Each β_s coefficient estimates the difference between non-expansion-to-expansion migration and migration in the opposite direction, relative to that difference as it existed in 2012. If the two sets of states were following parallel trends prior to treatment, the β_s terms would be estimated to be close to zero.

Column 1 of **Table 1.2** reports results from this regression for all individuals; column 4 reports analogous results for the sample of childless adults. Both columns confirm the visual evidence of **Figure 1.3**. In both columns, the coefficient estimates for 2008 through 2011 are close to zero, while the estimates for 2005 through 2007 are negative and (in most cases) significant. Indeed, in both columns, an F test does not reject that the 2008 through 2011 terms are jointly equal to zero. However, the F test does reject that coefficients 2005 through 2011 are equal to zero (or comes close to rejecting, in the case of the sample of childless individuals). Columns 2 and 3, and 5 and 6, add demographic controls and unemployment rate controls, which have essentially

[17] Recall that observations from 2013 are dropped when estimating Equation (1.1) since 2012 is the final interview year which is assumed to be completely untreated.

no effect.[18]

Given the results of the event study, the remainder of this paper focuses attention to the results which start the sample in 2008. Effectively, when the regression includes data from 2005 to 2007, the implicit counterfactual assumes that the difference between the expansion-to-non-expansion migration rate and the non-expansion-to-expansion migration rate would have grown in magnitude 2014 (toward the size of that difference as it existed from 2005 to 2007). There is little reason to have expected this to occur; thus, excluding these three years of data provides for a more plausible counterfactual.

1.4.2 Baseline results

Table 1.3 reports the estimates from Equation (1.1), which I consider to be the main results of this paper. The model is estimated on the analysis sample as described in **Section 1.3.1**, dropping observations from interview-year 2013. The first column shows the results without any controls; the second column adds demographic controls, the third adds unemployment rate controls, and the fourth adds state-specific trends.[19] In brief, the point estimates are quite close to zero and are estimated with sufficient precision to rule out large migration effects. The point estimate in the first column for all individ-

[18] **Table 1.9** reports estimates from a regression of the form $y_{irt} = \lambda_t + \mu_r + \beta nonexp_r * year_t + \epsilon_{rt}$, which estimates the difference in pre-existing trends of inter-region migration from non-expansion states relative to expansion states. The results tell the same story: the difference in pre-existing trends is substantial when starting the sample in 2005 and much smaller when the sample starts in 2008.

[19] **Table 1.3** reports only the coefficient on $nonexp_r post_t$. **Table 1.10** shows the estimates for the control variables as well.

uals is -0.005 – that is, -0.005 percentage points — from a baseline mean of 1.058%. The top of the 95% confidence interval is 0.174 percentage points. For childless adults, the point estimate in the first column is 0.012 percentage points from a mean of 1.716%; the top of the 95% confidence interval is 0.271 percentage points. Including demographic controls and unemployment rate controls makes virtually no difference; adding state-specific trends has little effect for childless adults and makes the estimate somewhat more negative for the estimate using all individuals.[20]

A back-of-the-envelope calculation shows that the fiscal externality of state expanding Medicaid is trivial, even at the largest upper bound reported in this table. In particular, let t_{nat} denote the Medicaid take-up rate of native eligible individuals. Let t_{mig} denote the analogous take-up rate of those individuals induced to migrate to an expansion state (or induced not to migrate *from* an expansion state) by the Medicaid expansion. Let α denote the number of induced migrants as a fraction of the population of non-expansion states, which corresponds to β from equation (1.1). I will let $\alpha = 0.0030$ (i.e., 0.30 percentage points), since 0.0030 is the largest top of the 95% confidence interval reported in **Table 1.3**. Let P^E and P^N denote the population of eligible individuals in expansion and non-expansion states respectively. The total increase in

[20] A triple difference specification, using a control sample (e.g., those with earnings between 250% and 400% of poverty), would be feasible in principle. However, these specifications are poisoned by large differential pre-existing trends. For both childless families and all families, the difference (analysis sample minus control sample) in the difference in pre-existing trends (between non-expansion-to-expansion migration and migration in the opposite direction) is larger in magnitude than the single difference in pre-existing trends in the analysis sample, invalidating the use of the triple difference strategy for these specifications.

the number of Medicaid beneficiaries in expansion states is approximately $\alpha t_{mig} P^N$. The number of native Medicaid beneficiaries in the expansion states is $t_{nat} P^E$. Thus, making the additional approximation that $P^E \approx P^N$ (see **Table 1.1**), the relative increase in Medicaid beneficiaries due to migration is $\alpha \frac{t_{mig}}{t_{nat}}$. So, even if migrants took up Medicaid at a rate *six times* that of native eligibles (so that $\frac{t_{mig}}{t_{nat}} = 6$), migration would increase Medicaid rolls by no more than 2% in the short-run.[21]

1.4.3 Threats to the parallel trend assumption

As mentioned above, a causal interpretation to coefficient estimates from Equation (1.1) requires a parallel trend assumption: in the absence of the Medicaid expansion, expansion-to-non-expansion migration would have followed the same trend as non-expansion-to-expansion migration between the pre-period and the post-period. The most straightforward threat to this assumption is the possibility that economic conditions changed in 2014 in expansion states relative to the analogous change in non-expansion states. For expositional purposes, suppose that expansion states experienced a relative boom. This could potentially bias the results in either direction. On the one hand, the causal effect of economic conditions (better conditions lead to more in-migration) could lead to an increase in migration to expansion states. On the other hand, the

[21]Equation (1.1) could instead be estimated at the origin-state/year level, using the log of the inter-region migration rate as the dependent variable. Without any additional control variables, such a regression yields a coefficient estimate of -.0376 with a standard error of 0.0840. The top of the 95% confidence interval would imply an effect of 12.7 log points, or about 0.14 percentage points from the baseline of 1.07%.

mismeasurement effect discussed in **Section 1.3.1** could lead to an increase in measured migration *from* expansion states: individuals with origin in expansion states might have income low enough to be in the analysis sample if they move away, but not if they remain a resident of an expansion state. In any case, however, the economic improvement between the pre-period (2008 to 2012) and the post-period (2014) appears fairly similar in expansion and non-expansion states as measured by the unemployment rate. In expansion states, the mean unemployment rate (weighted by the appropriate sum of the sample weights in the analysis sample) in expansion states fell from 9.0% to 6.6% between the pre-period and the post-period; the analagous change in non-expansion states was from 8.1% to 5.9%. Thus, changes in economic conditions does not appear a likely candidate to cause a violation of the parallel trend assumpion.[22]

Another threat to the parallel trend assumption is the possibility that Medicaid itself causally affects income, and thus selection into the analysis sample, by reducing labor supply (Garthwaite, Gross, and Notowidigdo (2014)). Such a causal effect would tend to cause individuals with a non-expansion origin state in 2014 to have lower potential earnings if they migrate to expansion states than if they do not. By the mechanism discussed above, this would increase estimated migration, but not true migration, from non-expansion states

[22]The employment-to-population ratio (using employment derived from the Quarterly Census of Employment and Wages) does suggest a slightly better improvement in expansion states. This ratio increased from 51.9 to 52.3 percent in non-expansion states and from 52.5 to 53.3 percent in expansion states. However, the difference between these differences is not statistically significant.

to expansion states in 2014 among the analysis sample, biasing the coefficient estimate from Equation (1.1) upward. An upward bias to the coefficient estimate would indicate that the estimates discussed above (finding a zero effect of Medicaid on migration) are in fact upper bounds.

Finally, despite the fact that observations from interview-year 2013 are dropped, anticipation effects in 2013 remain a threat to the parallel trend assumption. In particular, such anticipation effects would lead to a large increase in non-expansion-to-expansion migration in 2013. These new migrants would appear in the 2014 sample as residents of expansion states, but their rate of inter-region migration in 2014 (that is, back to the region of non-expansion states) might be very low. This would have the effect of reducing the measured expansion-to-non-expansion migration rate in 2014. However, as in the previous paragraph, this would tend to bias the coefficient estimate of Equation (1.1) upward — meaning that such an effect would not threaten the validity of the coefficient estimates as upper bounds of the true migration effect of Medicaid.

1.4.4 Subsample analyses

While estimates using the entire analysis sample find no migration effects from the Medicaid expansion, it is possible that migration effects are concentrated in certain subgroups. First, the Medicaid expansion might be sufficient to induce migration over a short distance but not a long distance. Second, the expansion might induce migration only among those with relatively poor health.

To examine the first possibility, **Table 1.4** estimates versions of Equation (1.1) on subsamples of the analysis sample with origin close to a border between an expansion state and a non-expansion state.[23] The first row restricts the analysis sample to individuals in Commuting Zones that straddle the expansion/non-expansion border.[24] Of the 709 commuting zones, 61 straddle a border between an expansion state and a non-expansion state.[25] In this row, the regression includes fixed effects for the state/Commuting Zone combination (rather than just the state), and the standard errors are clustered twoway at the state and Commuting Zone level. The estimates are generally large and economically meaningful, but the statistical precision drops substantially. For instance, among all individuals in Commuting Zones that straddle the boundary, the coefficient estimate with no state trends is 0.101 percentage points (from, unsurprisingly, a higher baseline interregion migration rate of 1.449%) with a standard error of 0.210. This standard error is more than double than the analogous standard error from estimating Equation (1.1) on the unrestricted analysis sample.

[23]Only the estimates with demographic and unemployment rate controls are reported. The addition of these controls makes essentially no difference; results available upon request.

[24]A Commuting Zone is a set of counties constructed by the United States Department of Agriculture such that commuting ties are strong within each Commuting Zone but weak between different Commuting Zones. Commuting Zones form a partition of the United States.

[25]A given PUMA or PUMA of migration might not map cleanly into a commuting zone. In these cases, individuals are assigned probabilistically to each commuting zone in which their PUMA or PUMA of migration lies. For instance, if an individual's origin is in a PUMA whose population is 75% within Commuting Zone A and 25% within Commuting Zone B, this individual is placed with 0.75 weight in Commuting Zone A and 0.25 weight in Commuting Zone B.

In order to improve power, the remaining rows restrict the analysis sample somewhat less stringently. In particular, the regressions in these rows keep all individuals who live in a PUMA (or PUMA of migration, as the case may be) whose population centroid lies within a certain distance of an expansion/non-expansion border.[26] The results tell a similar story as the sample in straddling Commuting Zone: while the coefficient estimate is sometimes large, the standard errors are even larger.[27] In sum, the advantage of restricting the sample to individuals living near an expansion/non-expansion border — isolating those individuals most likely to respond — appears outweighed by the loss of statistical precision.

Table 1.5 considers the possibility that only individuals in relatively poor health migrate in response to the Medicaid expansions. Two proxies for poor health are considered. In the first row of each panel, the sample is restricted to individuals age 46 or older, as health expenditures (aside from those related to childbirth) tend to increase with age.[28] In the second row, the sample is restricted to a sample of individuals who report a disability.[29] While one estimate in **Table 1.5** — the estimate for childless adults reporting a disability,

[26]One concern about this strategy is the fact that PUMA (and PUMA of migration) borders were revised in 2012. It is thus possible that the composition of individuals in PUMAs (or PUMAs of migration) within a certain distance of an expansion/non-expansion border could have changed substantially between 2011 and 2012. However, in results available upon request, there does not appear to be a break in any series of education, race, or income variables between 2011 and 2012 among this restricted sample.

[27]Note, however, that three of the four coefficients estimated on the sample restricted to being within 75km of the boundary are estimated to be negative.

[28]Note that prior to the ACA, the lowest eligibility threshold in any state for Medicaid coverage for pregnant women was 133% of the federal poverty level; these coverage levels were not substantially affected by the ACA.

[29]Specifically, this restricted sample contains individuals who report a hearing, vision, cognitive, ambulatory, self-care, or independent-living difficulty.

in the specification including state-specific trends — is economically large, all estimates are insignificant and six of the eight are negative. These results are consistent with a null effect of Medicaid on migration, even among relatively sicker populations, though the decline in power means that one cannot rule out modest-sized effects.

1.5 Robustness to sensitivity analyses

A concern with the empirical strategy, as discussed above, is the endogeneity of income with respect to place of residence. For this reason, Equation (1.1) is re-estimated using a sample selected on the basis of education levels. Specifically, all individuals in this revised sample are part of families with (1) at least one individual with less than a high school education and (2) no member with more than a high school education. This ameliorates the endogeneity-of-sampling issue, though it also create attenuation bias, since only 47% of this sample has family income less than 138% of the poverty level. The first row of **Table 1.6** repeats the result from **Table 1.3**, i.e., using the usual income-based sample. The second row reports results using the education-selected sample. The third row restricts the sample to individuals who would qualify for *both* the income-based and education-based sample. First, the results show that baseline migration rate is lower in the education-selected sample (see Column 3), perhaps because this sample excludes many individuals who have transitory low-income which may be correlated with migration decisions, as discussed in

Section 1.3.1. Second, the regression results without state trends are all approximately zero. The specifications with state trends exhibit somewhat more variance, but all estimates are insignificant and half are positive and half are negative, consistent with a null effect.

Next, I analyze the robustness of the results to alternative specification choices regarding certain states for whom either the timing of the expansion, or the assignment to regions, is somewhat tenuous. First, I consider alternative specification choices regarding the seven states (including D.C.) that had expanded health coverage to non-disabled adults prior to 2014. The first two rows of each panel of **Table 1.7** re-estimate the baseline results after dropping these seven states or a subset thereof. The first row drops Massachusetts and the District of Columbia since each jurisdiction expanded coverage to virtually the entire newly-eligible population prior to 2014. Specifically, in this row, (1) individuals with those locations as origin are dropped and (2) the dependent variable y_{irt} is redefined to equal zero for moves into those two locations. The second row drops five more states (CA, CT, MN, NJ, and WA) which implemented more limited elements of the Medicaid expansion prior to 2014. Neither specification yields results that are meaningfully different than the baseline specification.

Additionally, for some states who expanded Medicaid in early 2014, but after January 1, assignment to the set of expansion states or the set of non-expansion states is somewhat arbitrary. For instance, Michigan is assigned to the "expansion" set, since it implemented its expansion decision in April 2014

— later than the implementation date in most other expansion states, but plausibly early enough to affect migration decisions in 2014. But, if individuals were surprised by the implementation, and if most migration decisions in the data were made prior to learning about the implementation, then outmigration from Michigan would be misattributed as coming from an expansion state, while in fact individuals were acting as if Michigan was a non-expansion state. Similarly, some clairvoyant residents in the states of Alaska, Indiana, Louisiana, Pennsylvania, and Montana — states which are assigned to the non-expansion set — could have foreseen that their states would eventually choose to expand Medicaid.

To address this concern, rows 3, 4, and 5 of **Table 1.7** re-estimate the main results in three ways. First, Indiana and Pennsylvania are re-assigned to the "expansion" set. Both of these states expanded in late 2014 or early 2015. Second, New Hampshire and Michigan, which expanded Medicaid in early 2014, are re-assigned to the non-expansion set. Arkansas is also re-assigned to the non-expansion set in this specification; Arkansas implemented a version of the Medicaid expansion which more closely resembled private insurance exchanges and thus may not have been as salient. Third, all five of these states are dropped. None of these changes materially affect the results.

1.5.1 Treating Medicaid thresholds as a continuous variable

It is possible that definition of the Medicaid expansion as a binary variable is too coarse and that a finer classification of Medicaid eligibility would reveal migration effects. In particular, using Medicaid eligibility as a continuous variable exploits the fact that the Medicaid expansion had a bigger eligibility effect in expanding states which had lower prior eligibility thresholds. Furthermore, many states expanded or contracted eligibility prior to the 2014 expansions under the ACA. In the data from Kaiser Family Foundation (2016a, 2016b, and 2016c) used in this section, there are 203 pre-2014 instances of (mostly small) changes to the parent threshold and seven pre-2014 instances of (somewhat larger) changes to the childless adult threshold. Of course, such a finer classification has several costs. First, the interpretation of the results changes somewhat; the results no longer estimate the migration effect of solely the ACA Medicaid expansions, rather than some other Medicaid expansion. Second, using more general variation across states over time makes a region-based analysis impossible, as discussed above. Third, it is difficult to summarize Medicaid eligibility with one eligibility threshold.

To explore the effect of Medicaid eligibility more generally, **Table 1.8** reports estimates of the following two-way fixed effects model run at the state-year (st) level using data derived from the analysis sample.

$$y_{st} = \lambda_t + \mu_s + \beta_1 elig_{st} + \beta_2 childelig_{st} + x'_{st}\gamma + \epsilon_{st} \tag{1.3}$$

In this equation, y_{st} is either the inflow rate from or the outflow rate into a given state among the analysis sample. The inflow rate is the weighted number of newly-arrived migrants in state s at time t divided by the weighted total of observations in the analysis sample living in state s at time t.[30] The outflow rate is the weighted number of migrants who report having left state s in the prior 12 months when interviewed at time t, divided by the weighted total of observations that reported living in state s 12 months before being interviewed at time t. The key explanatory variable is $elig_{st}$: the state-year specific Medicaid eligibility threshold for adults (childless adults or parents, depending on the specification) as a proportion of the poverty level (i.e., so that a value of one corresponds to 100% of the poverty level). If Medicaid were influencing migration decisions of families, the coefficient would be expected to be positive in the inflow equation and negative in the outflow equation.

The model also includes $childelig_{st}$, the state-year specific eligibility threshold for children, since families likely also have preferences over the health coverage for their children, and the child eligibility threshold is correlated with the adult eligibility threshold.[31] However, these eligibility thresholds are generally

[30]This weighted total is computed using the ACS sample weights. Furthermore, Equation (1.3) is run at the state-year level, weighted by the sum of the ACS sample weights in that state in that year.

[31]$childelig_{jt}$ is typically the CHIP threshold.

substantially higher (200 to 300% of poverty) in all states, such that variation in children's eligibility may not be binding for individuals in this sample. This regression is run separately for families with and without children because $elig_{st}$ is available for childless adults only since 2011. In the regression using childless adults, the $childelig_{st}$ variable is not included.

This estimation strategy relies on a parallel trend assumption. In particular, within-state changes in Medicaid eligibility cannot be correlated with state-time specific factors that affect migration. This assumption would be violated if states expand (contract) Medicaid precisely when the state budget finances are in better (worse) shape, since state budget finances may be correlated with economic factors that affect how attractive a state is to potential migrants. While the identification assumption cannot be tested directly, it is addressed by including (in some specifications) state-specific linear trends and the state unemployment rate in x_{st}.

The first two columns of **Table 1.8** present these estimates for parents. The model includes both the parents' eligibility threshold (as a proportion of poverty) as well as the eligibility threshold for the children. In brief, they show weakly suggestive evidence in favor of Medicaid influencing migration decisions, but these findings are not robust. Consider first the upper panel. In both specifications, the top row suggests that a 100 percentage-point increase in the eligibility threshold for parents would increase the inflow rate of individuals earning less than 138% of poverty by about 0.22 percentage points relative to a base of 1.810%, although these estimates are not significantly different than zero. The first row of the bottom panel estimates an outflow effect of about -0.05 percentage points.

Interestingly, three out of the four estimates find a wrong-signed effect of the children's threshold: a negative effect on inflows and a positive effect on outflows. This could be a sign that the within-state variation in the children's threshold is more endogenous, which makes sense since the ACA Medicaid expansion — a plausibly exogenous national policy shock — affected the parents' threshold to a far greater degree than the children's threshold. Indeed, the within-state standard deviation of parents' eligibility (after partialling out time fixed effects) falls by 40% when observations from 2014 are excluded. By contrast, 2014 changes explain very little variation in children's eligibility: the within-state standard deviation is 0.141 using all years and 0.136 excluding 2014.

Columns 3 and 4 of **Table 1.8** report the estimates for a similar model for childless adults using data since 2011, the earliest year for which the childless adult threshold was collected by Kaiser. The specifications without linear trends show small and insignificant effects in the expected direction. Using the estimates from models without state-specific trends, the top of the 95% confidence interval for the inflow rate would be about 0.31 percentage points from a base of 4.15%, while the bottom of the 95% confidence interval for the outflow estimates would yield a reduction of about 0.29 percentage points. The specifications with linear trends make both specifications more positive, though with only four years of data, one should be concerned that a model with linear trends is overfit.

Ultimately, these results show that it is *possible* that Medicaid eligibility changes — including but not limited to those changes brought about by the ACA — could have increased inflows to and reduced outflows from states with more generous Medicaid eligibility thresholds. On the other hand, most estimates are not statistically distinguishable from zero, and the results for the children's threshold provoke some concern regarding policy endogeneity.

1.6 Conclusion

Using various specifications, this paper finds little evidence that the Affordable Care Act's expansion of Medicaid in 2014 affected migration. In the baseline specification, one can rule out (at 95% confidence) an effect on the inter-region migration rate of 0.18 percentage points from a base of 1.06%, or an effect of 0.30 percentage points from a base of 1.72% among childless individuals — exerting a neglible (at best) fiscal externality from expanding Medicaid. This leaves open the question of *why* the expansions do not appear to lead to migration, given that the average newly eligible Medicaid beneficiary received medical care that cost $5,500 to provide on average. First, the cost of moving (including the cost of destroying the idiosyncratic match between the individual and the place of origin) might be very large relative to the value of Medicaid benefits — perhaps because beneficiaries might value Medicaid at far less than cost (Finkelstein, Hendren, and Luttmer (2015)) or because the distribution of anticipated health expenditures is highly skewed. Second, individuals in non-expansion states might have chosen to remain in their origin state, anticipating that the state will choose to expand Medicaid in the near future. Residents of Indiana, Montana, Pennsylvania, Louisiana, and Alaska would have been correct. Although robustness checks show that the null result holds even when excluding those states which expanded Medicaid in late 2014 or early 2015, the same "delaying" effect could be present in other states as well. Third, it is possible that migration effects would not become visibile until 2015, either because it takes time to learn about the Medicaid benefits available in other states, or because the certainty in each state's expansion or non-expansion decision increases over time. The present research design cannot distinguish between these hypotheses. Fourth, it is possible that potential Medicaid beneficiaries had already sorted themselves into high-benefit states prior to the 2014 expansions, in line with some of the suggestive evidence using all within-state variation in Medicaid eligibility thresholds.

This project leaves room for future research. Most simply, data from an additional year of the ACS would be able to uncover any longer-run effects that are not visible in the short term. Additionally, one could use aggregate statistics released by the Statistics of Income (SOI) division of the Internal Revenue Service. Among other aggregate statistics, the SOI releases information on the total (interstate) outflow and inflow from a given county in a given year. Among the set of counties in non-expansion states, one could test whether (1) poorer counties and (2) counties closer to a border with an expansion state see a larger outflow in 2014, which would be consistent with a migration response to Medicaid. The relevant data would measure moves between tax season 2014 and tax season 2015; such data are expected to be released approximately December 2016. Until then, the 2014 ACS remains the best data source with which to analyze migration effects of the ACA Medicaid expansion.

1.7 Tables and Figures

Table 1.1: Summary stats

	Expansion states	Non-expansion states	*Differences*
Age	40.496	40.894	-0.398
(adults)	(19.989)	(20.335)	(0.298)
Male	0.457	0.449	0.007**
(all individuals)	(0.498)	(0.497)	(0.004)
White, non-Hispanic	0.488	0.510	-0.021
(all individuals)	(0.500)	(0.500)	(0.086)
Hispanic	0.281	0.196	0.085
(all individuals)	(0.449)	(0.397)	(0.104)
Immigrant	0.152	0.090	0.062
(all individuals)	(0.359)	(0.287)	(0.044)
Real family income	11813.29	12049.14	-235.85
(families)	(9514.01)	(9365.11)	(297.49)
Married	0.321	0.332	-0.012
(adults)	(0.467)	(0.471)	(0.017)
Family size	2.210	2.246	-0.037
(families)	(1.654)	(1.610)	(0.081)
High school degree	0.430	0.450	-0.020
(adults)	(0.495)	(0.498)	(0.019)
Some college	0.228	0.216	0.012
(adults)	(0.419)	(0.412)	(0.007)
Bachelor's or more	0.082	0.069	0.013***
(adults)	(0.274)	(0.253)	(0.005)
Share of observations	0.505	0.495	
(all individuals)	(0.500)	(0.500)	
Obs. (all individuals)	2398360	2349949	

Notes: Each cell reports the sample mean of the variable indicated, among individuals in the analysis sample with origin state in expansion states (Column 1) or non-expansion states (Column 2). The assignment of a state to "expansion" or "non-expansion" is time-invariant, based on the 2014 Medicaid expansion, as defined in the text. Column 3 reports the difference in means and the standard error thereof (calculated allowing correlation among observations in the same origin state). The mean for male, white, hispanic, and immigrant is taken over all individuals. The means for education outcomes, marriage, and age are taken over all adults age 18 or older. The mean of family size and real family income is taken over all families. Standard deviations are in parentheses. The analysis sample includes individuals in families with income less than 138% of the federal poverty guidelines. See text for further details of sample selection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.2: Event study for all individuals

	All individuals			**Individuals without children**		
	(1) No controls	(2) Demographics	(3) Unemp. rate	(4) No controls	(5) Demographics	(6) Unemp. rate
treatXyr2005	-0.205**	-0.209**	-0.212**	-0.0544	-0.0504	-0.0710
	(0.0883)	(0.0876)	(0.0984)	(0.175)	(0.174)	(0.187)
treatXyr2006	-0.301***	-0.299***	-0.306***	-0.468**	-0.435**	-0.452**
	(0.103)	(0.105)	(0.112)	(0.191)	(0.199)	(0.205)
treatXyr2007	-0.195**	-0.194**	-0.200**	-0.181	-0.178	-0.189
	(0.0916)	(0.0904)	(0.0906)	(0.120)	(0.120)	(0.127)
treatXyr2008	-0.0398	-0.0336	-0.0342	0.0545	0.0749	0.0673
	(0.0931)	(0.0932)	(0.0972)	(0.150)	(0.152)	(0.157)
treatXyr2009	-0.0575	-0.0626	-0.0549	-0.114	-0.140	-0.141
	(0.0937)	(0.0946)	(0.0991)	(0.144)	(0.143)	(0.144)
treatXyr2010	0.0318	0.0298	0.0394	0.00179	-0.00178	-0.00103
	(0.0627)	(0.0627)	(0.0711)	(0.0875)	(0.0887)	(0.101)
treatXyr2011	-0.00667	-0.00676	-0.00401	-0.129	-0.125	-0.128
	(0.0848)	(0.0842)	(0.0887)	(0.126)	(0.128)	(0.131)
P-value: 2005-2011 equal zero	0.010	0.010	0.008	0.083	0.106	0.124
P-value: 2008-2011 equal zero	0.900	0.895	0.870	0.617	0.543	0.536
Non-exp. outmig. rate (2012)	1.058	1.058	1.058	1.716	1.716	1.716
Observations	3707149	3707149	3707149	1319817	1319817	1319817
Clusters	51	51	51	51	51	51
State fixed effects	X	X	X	X	X	X
Time fixed effects	X	X	X	X	X	X
Demographic controls		X	X		X	X
U-rate controls			X			X
State-specific trends						

Notes: This table reports coefficients on the set of $nonexp_r * 1(year_t = s)$ variables from a regresion of the form $y_{irt} = \lambda_t + \mu_r + \sum_{s \neq 2012} \beta_s nonexp_r * 1(year_t = s) + x'_{irt}\gamma + \epsilon_{rt}$ estimated at the individual level on the analysis sample using data from 2005 through 2012, wherewhere $nonexp_r$ is a dummy for being a non-expansion state and y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview in year t. Each column corresponds to a separate regression. Columns 4-6 restrict attention to childless adults. In Columns 1 and 4, x_{irt} is empty. Columns 2 and 5 add demographic variables to x_{irt}. Columns 3 and 6 add the unemployment rate controls to x_{irt}; see text for details. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: Main results: Changes in inter-regional migration rates from non-expansion states relative to expansion states.

	No controls	Demographics	Unemp. rate	State trends
All Individuals	-0.005	-0.004	-0.014	-0.053
	(0.091)	(0.089)	(0.087)	(0.109)
	[-0.184,0.174]	[-0.179,0.171]	[-0.184,0.156]	[-0.268,0.161]
Non-exp. outmig. rate (2012)	1.058	1.058	1.058	1.058
Observations	2985490	2985490	2985490	2985490
Clusters	51	51	51	51
Childless Individuals	0.012	0.006	-0.012	0.010
	(0.132)	(0.130)	(0.125)	(0.146)
	[-0.247,0.271]	[-0.249,0.262]	[-0.258,0.234]	[-0.277,0.296]
Non-exp. outmig. rate (2012)	1.716	1.716	1.716	1.716
Observations	1085956	1085956	1085956	1085956
Clusters	51	51	51	51
State fixed effects	X	X	X	X
Time fixed effects	X	X	X	X
Demographic controls		X	X	X
U-rate controls			X	X
State-specific trends				X

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data from the analysis sample, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014, y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview. In the first column, x_{irt} is empty. The second column adds demographic controls, the third column adds unemployment rate controls, and the fourth column adds state-specific linear trends; see text for details. The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level; numbers in bracket indicate the bounds of the associated 95% confidence interval. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: Changes in inter-regional migration rates from non-expansion states relative to expansion states for individuals with origin near expansion/non-expansion border.

	Coefficient Estimates		Summary Stats	
	No state trends	State trends	*Baseline mean*	*Fraction in sample*
All Individuals				
Commuting Zones that straddle	0.101	0.348	1.449	0.124
	(0.210)	(0.352)		[533060]
Less than 75km from border	-0.022	-0.195	1.459	0.230
	(0.191)	(0.320)		[683668]
Less than 150km from border	0.150	0.111	1.276	0.431
	(0.138)	(0.213)		[1281839]
Less than 250km from border	0.090	0.087	1.183	0.575
	(0.108)	(0.168)		[1721566]
Childless Individuals				
Commuting Zones that straddle	0.257	0.467	2.616	0.127
	(0.347)	(0.562)		[198355]
Less than 75km from border	0.054	-0.074	2.392	0.240
	(0.256)	(0.410)		[258187]
Less than 150km from border	0.216	0.170	2.069	0.454
	(0.197)	(0.271)		[486033]
Less than 250km from border	0.095	0.146	1.888	0.597
	(0.168)	(0.225)		[646143]
Place fixed effects	X	X		
Time fixed effects	X	X		
Demographic controls	X	X		
U-rate controls	X	X		
State-specific trends		X		

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data from subsets of the analysis sample, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014, y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview in year t. In the first row, r indexes the Commuting Zone / state combination; in other rows, it indexes the state. In the first column, x_{irt} includes demographic controls and unemployment rate controls, and the second column adds state-specific linear trends; see text for details. The third column reports the baseline mean migration rate in the relevant sample from non-expansion states in 2012. The fourth column reports the number of observations (in brackets) as well as the size of the restricted sample as a fraction of the unrestricted analysis sample. The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level, except for the first row which is clustered twoway at the state / Commuting Zone level. The number of clusters along the minimum dimension in row 1 (of each panel) is 36. Row 2 has 33 clusters. Row 3 has 42 clusters. Row 4 has 47 clusters. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection and additional restrictions for this table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Changes in inter-regional migration rates from non-expansion states relative to expansion states for individuals with higher expected health expenditure.

	Coefficient Estimates		**Baseline mean**
	No state trends	State trends	
All Individuals			
Age 46 and older	-0.077	-0.114	0.701
	(0.080)	(0.134)	[756624]
Reporting a disability	-0.084	0.017	0.800
	(0.100)	(0.172)	[505586]
Childless Individuals			
Age 46 and older	-0.082	-0.174	0.815
	(0.102)	(0.165)	[563850]
Reporting a disability	-0.080	0.329	0.867
	(0.137)	(0.212)	[308225]
Clusters	51	51	
State fixed effects	X	X	
Time fixed effects	X	X	
Demographic controls	X	X	
U-rate controls	X	X	
State-specific trends		X	

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data from subsets of the analysis sample, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014, y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview. In the first column, x_{irt} includes demographic controls and unemployment rate controls, and the second column adds state-specific linear trends; see text for details. The third column reports the baseline mean migration rate in the relevant sample from non-expansion states in 2012 as well as the number of observations (in brackets). The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection and additional restrictions for this table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: Changes in inter-regional migration rates from non-expansion states relative to expansion states under alternative sample restrictions.

	Coefficient Estimates		Baseline mean
	No state trends	State trends	
All Individuals			
In baseline income sample	-0.014	-0.053	1.058
	(0.087)	(0.109)	[2985490]
In education sample	-0.005	-0.158	0.570
	(0.061)	(0.136)	[1702135]
In both samples	-0.036	-0.161	0.634
	(0.104)	(0.194)	[791914]
Childless Individuals			
In baseline income sample	-0.012	0.010	1.716
	(0.125)	(0.146)	[1085956]
In education sample	0.001	0.023	0.587
	(0.084)	(0.124)	[689352]
In both samples	0.007	0.121	0.711
	(0.114)	(0.213)	[236895]
Clusters	51	51	
State fixed effects	X	X	
Time fixed effects	X	X	
Demographic controls	X	X	
U-rate controls	X	X	
State-specific trends		X	

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data from different samples, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014, y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview. The first row of each panel repeats the results from Columns 3 and 4 of **Table 1.3**, i.e., using the income-based analysis sample. In the second row of each panel, the sample is selected based on education: specifically, the sample includes families in which (1) at least one adult member has not received a high school diploma and (2) no member has attended college. The third row uses a sample that would be selected either under the income or education-based criterion. In the first column, x_{irt} includes demographic controls and unemployment rate controls, and the second column adds state-specific linear trends; see text for details. The third column reports the baseline mean migration rate in the relevant sample from non-expansion states in 2012 as well as the number of observations (in brackets). The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Changes in inter-regional migration rates from non-expansion states relative to expansion states: Robustness to alternative classifications of states.

	Coefficient Estimates		Baseline Mean
	No state trends	State trends	
All Individuals			
Drop DC and MA	-0.005	-0.064	1.015
	(0.082)	(0.107)	[2939532]
Drop MA and early expanders	0.019	-0.119	0.737
	(0.094)	(0.141)	[2396708]
Drop states with marginal expansion date	-0.041	0.010	0.982
	(0.086)	(0.111)	[2745842]
Assign to expansion	-0.036	0.095	1.133
	(0.076)	(0.090)	[2985490]
Assign to non-expansion	-0.009	-0.062	1.058
	(0.083)	(0.114)	[2985490]
Childless Individuals			
Drop DC and MA	-0.044	-0.057	1.651
	(0.117)	(0.136)	[1065485]
Drop MA and early expanders	0.015	-0.030	1.196
	(0.123)	(0.187)	[880343]
Drop states with marginal expansion date	-0.034	0.106	1.562
	(0.124)	(0.144)	[994931]
Assign to expansion	-0.041	0.225	1.808
	(0.124)	(0.157)	[1085956]
Assign to non-expansion	-0.001	0.015	1.716
	(0.121)	(0.147)	[1085956]
Place fixed effects	X	X	
Time fixed effects	X	X	
Demographic controls	X	X	
U-rate controls	X	X	
State-specific trends		X	

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data the analysis sample, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014, y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview in year t. In the first row of each panel, Massachusetts and the District of Columbia are dropped (see text for details). In the second row of each panel, CA, CT, DC, MA, MN, NJ, and WA are dropped. In the third row of each panel, AR, IN, MI, NH, and PA are dropped. In the fourth row of each panel, PA and IN (usually assigned to non-expansion) are assigned to expansion. In the fifth row of each panel, AR, MI, and NH (usually assigned to expansion) are assigned to non-expansion. In the first column, x_{irt} includes demographic controls and unemployment rate controls, and the second column adds state-specific linear trends; see text for details. The third column reports the baseline mean migration rate in the relevant sample from non-expansion states in 2012 as well as the number of observations (in brackets). The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level. The first row in each panel has 49 clusters. The second row has 44 clusters. The third row has 46 clusters. The fourth and fifth rows have 51 clusters.The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection and additional restrictions for this table. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

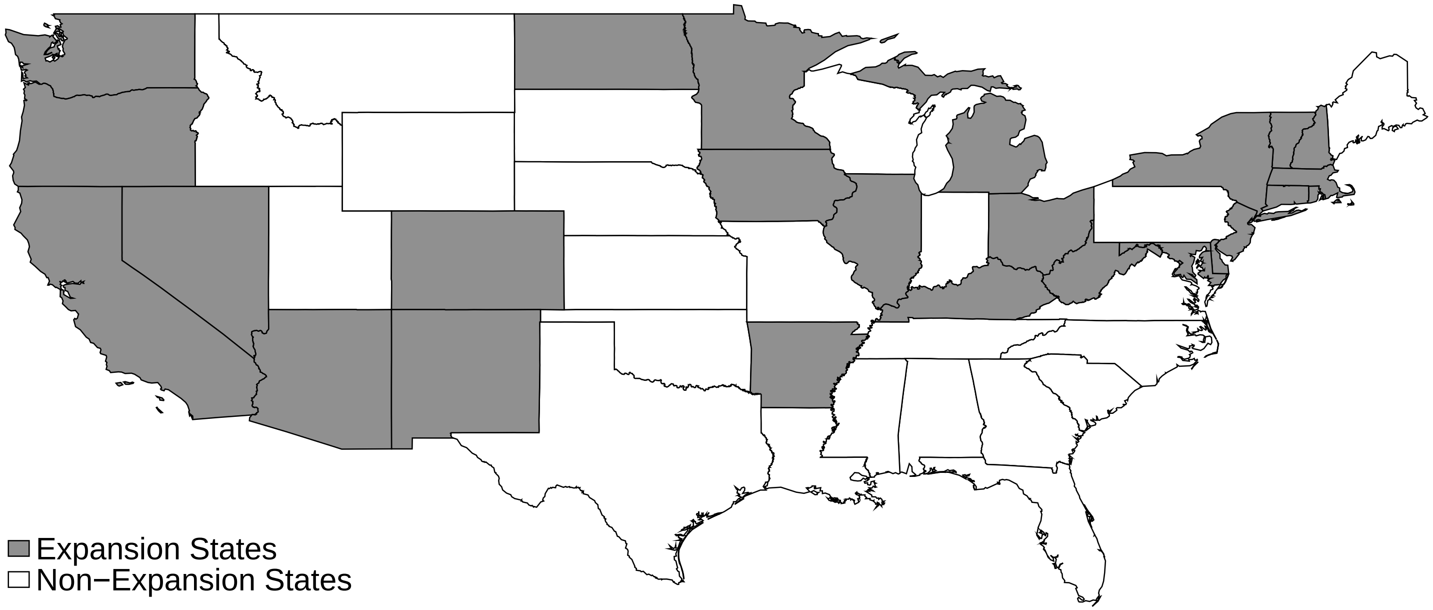
Table 1.8: Using all variation in Medicaid eligibility thresholds

	Individuals with children		**Individuals without children**	
	No state trends	State trends	No state trends	State trends
Inflow rate				
Parents' threshold	0.220	0.238		
	(0.137)	(0.180)		
Children's threshold	-0.169	0.153		
	(0.163)	(0.252)		
Childless adult threshold			0.0916	0.253
			(0.108)	(0.181)
Mean migration rate	1.810	1.810	4.151	4.151
Observations	357	357	204	204
Clusters	51	51	51	51
Outflow rate				
Parents' threshold	-0.0460	-0.0547		
	(0.0944)	(0.139)		
Children's threshold	0.122	0.177		
	(0.115)	(0.175)		
Childless adult threshold			-0.0396	0.0586
			(0.125)	(0.187)
Mean migration rate	1.810	1.810	4.151	4.151
Observations	357	357	204	204
Clusters	51	51	51	51

Notes: Each column of each panel reports the coefficients on $elig_{st}$ and $childelig_{st}$ from a regression of the following form, estimated at the state-year level on data from families with children from the analysis sample, from 2008 to 2014: $\ln(y_{rt}) = \lambda_t + \mu_s + \beta_1 elig_{st} + \beta_2 childelig_{st} + x'_{st}\gamma + \epsilon_{st}$, where y_{rt} represents the year-state specific inflow or outflow migration rate (see text for details), $elig_{st}$ is the eligibility threshold for parents, as a proportion of the poverty level, and $childelig_{st}$ is the eligibilty threshold for children (the maximum of the relevant Medicaid or CHIP threshold). See text for details on sample selection. Standard errors (in parentheses) are clustered at the state level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

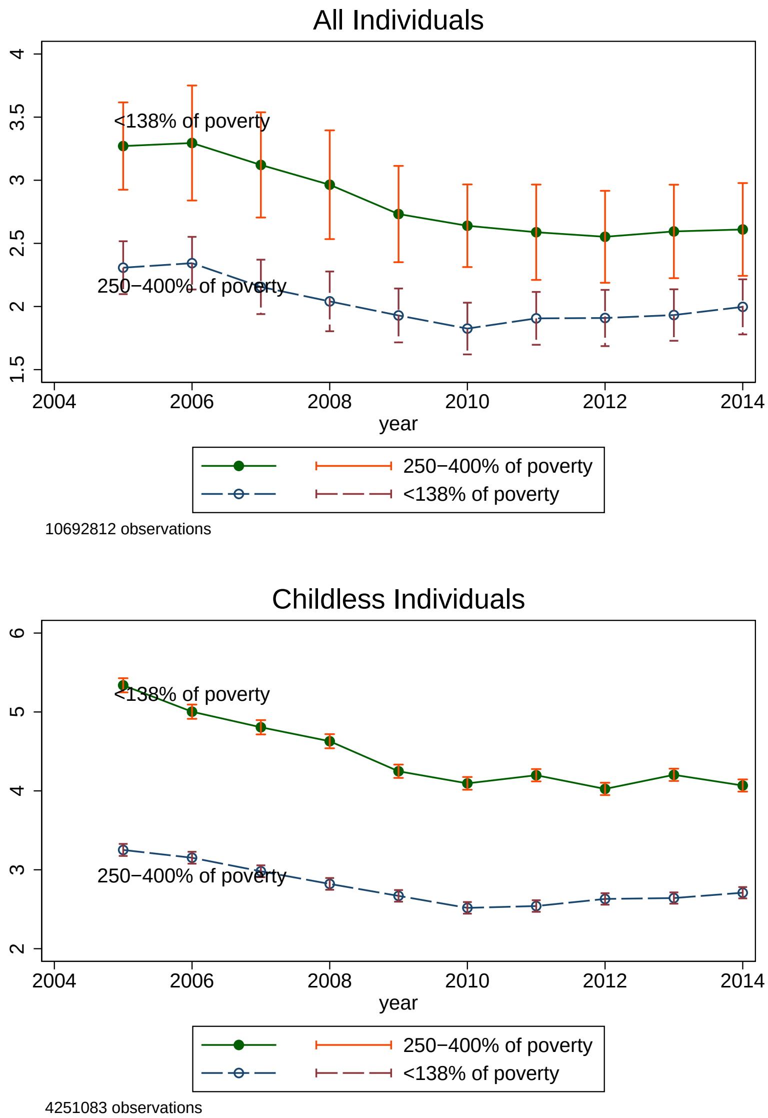
Figures

Figure 1.1: Expansion and non-expansion states, as of early 2014



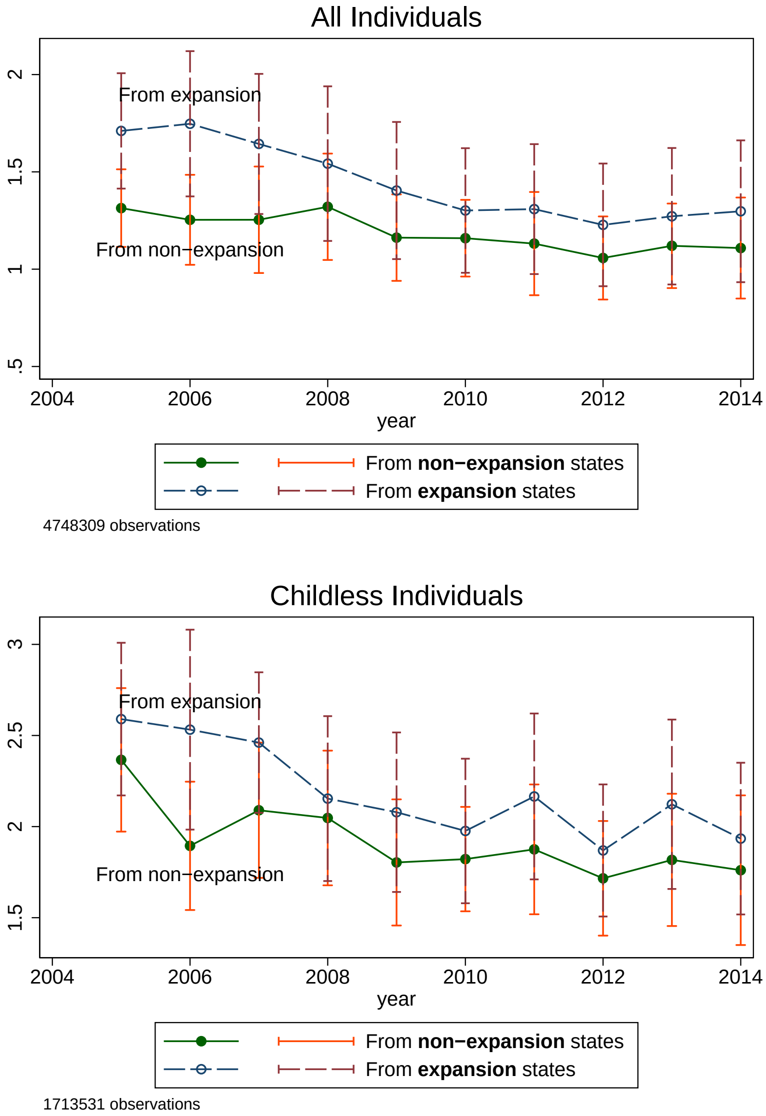
Note: Alaska is an non−expansion state; Hawaii is an expansion state

Figure 1.2: Interstate migration rates for main analysis sample and higher-income sample



Notes: Each point represents the average interstate migration rate (the ratio of interstate outmigrants to total observations) of the group indicated, scaled as a percentage. The analysis sample includes individuals in families with earnings up to 138% of the federal poverty guidelines. The control sample includes individuals in families with earnings between 250% and 400% of the federal poverty guidelines. See text for further details of sample selection. The brackets indicate 95% confidence intervals, robust to clustering at the origin-state level.

Figure 1.3: Summary of main results: Inter-regional migration rates among analysis sample, from expansion and non-expansion states



Notes: Each point represents the average inter-region migration rate (the ratio of inter-region outmigrants to total observations) of the analysis sample, from the region indicated, scaled as a percentage. The two regions are the set of expansion states and the set of non-expansion states. The analysis sample includes individuals in families with earnings up to 138% of the federal poverty guidelines. See text for further details of sample selection. The brackets indicate 95% confidence intervals, robust to clustering at the origin-state level.

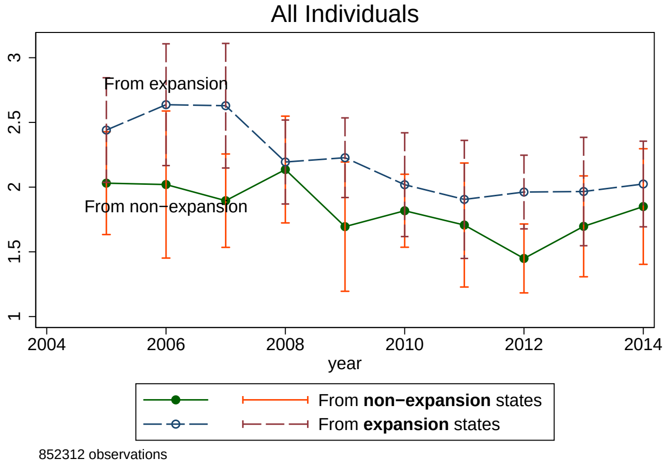
1.8 Appendix

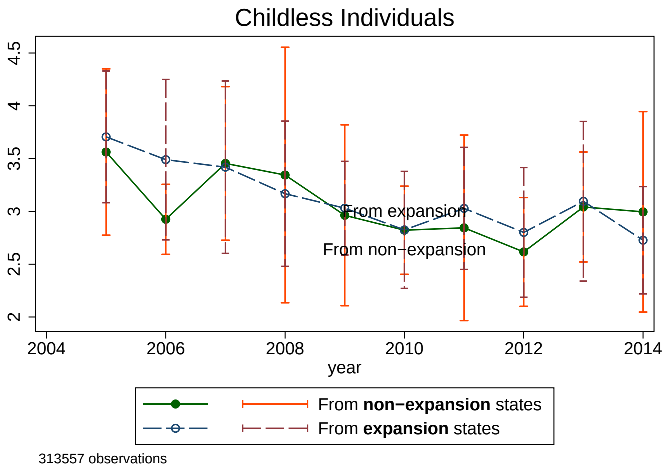
Table 1.9: Pretrends in inter-regional migration rates from non-expansion states relative to expansion states.

	No controls	Demographics	Unemp. rate
All Individuals			
Starting 2005	0.416***	0.415***	0.430***
	(0.114)	(0.114)	(0.122)
	[0.192,0.641]	[0.192,0.638]	[0.191,0.669]
Starting 2008	0.124	0.119	0.102
	(0.212)	(0.213)	(0.223)
	[-0.291,0.538]	[-0.299,0.537]	[-0.335,0.539]
Non-exp. outmig. rate (2012)	1.058	1.058	1.058
Observations (2008 row)	2471449	2471449	2471449
Clusters	51	51	51
Childless Individuals			
Starting 2005	0.282	0.256	0.286
	(0.206)	(0.206)	(0.218)
	[-0.121,0.685]	[-0.148,0.659]	[-0.142,0.713]
Starting 2008	-0.100	-0.104	-0.094
	(0.306)	(0.308)	(0.329)
	[-0.701,0.500]	[-0.708,0.500]	[-0.739,0.551]
Non-exp. outmig. rate (2012)	1.716	1.716	1.716
Observations (2008 row)	888895	888895	888895
Clusters	51	51	51
State fixed effects	X	X	X
Time fixed effects	X	X	X
Demographic controls		X	X
U-rate controls			X
State-specific trends			

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data from the analysis sample, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * year_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $year_t$ is the calendar year (divided by 10), y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview. x_{irt} includes demographic controls and unemployment rate controls; see text for details. The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level; numbers in bracket indicate the bounds of the associated 95% confidence interval. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.4: Inter-regional migration rates among analysis sample in commuting zones that straddle expansion / non-expansion border, from expansion and non-expansion states





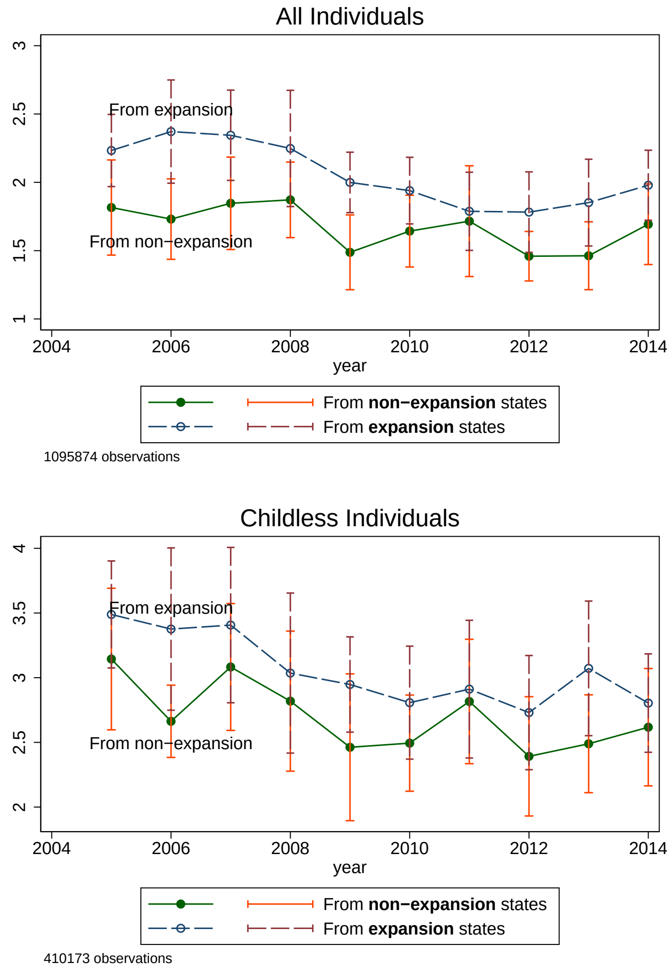
Notes: Each point represents the average inter-region migration rate (the ratio of inter-region outmigrants to total observations) from the region indicated of the analysis sample restricted to Commuting Zones that straddle a border between an expansion state and a non-expansion state. The analysis sample includes individuals in families with earnings up to 138% of the federal poverty guidelines. See text for further details of sample selection and assignment of individuals to Commuting Zones. The brackets indicate 95% confidence intervals, robust to clustering twoway at the origin-state and Commuting Zone level.

Table 1.10: Main regression results including all covariates.

	(1) All Individuals	(2) Childless Individuals
nonexp X post	-0.0534	0.00990
	(0.109)	(0.146)
Non-hispanic white	0.345***	0.514***
	(0.0780)	(0.0636)
Male	-0.00554	-0.108***
	(0.0236)	(0.0419)
White x male	-0.0197	-0.00795
	(0.0419)	(0.0738)
Hispanic	-0.0682	-0.137
	(0.0791)	(0.108)
Hispanic male	0.0213	0.212**
	(0.0451)	(0.107)
Not born in U.S.	-0.0384	-0.180
	(0.0722)	(0.115)
Age	-0.0163***	-0.129***
	(0.00449)	(0.0141)
Age squared	-0.000214***	0.000769***
	(0.0000258)	(0.000113)
Married	0.326***	0.632***
	(0.0520)	(0.0873)
Income / poverty	-0.176***	-0.112**
	(0.0365)	(0.0519)
1 child	-0.819***	
	(0.0880)	
2 children	-1.089***	
	(0.108)	
3 children	-1.111***	
	(0.0929)	
4+ children	-1.154***	
	(0.112)	
Log(distance to border)	-0.320***	-0.475***
	(0.0485)	(0.0839)
8th grade educ or less	-1.106***	-0.795***
	(0.160)	(0.154)
High school dropout	-0.997***	-1.098***
	(0.115)	(0.121)
Exactly high school	-0.587***	-0.668***
	(0.0915)	(0.120)
Own-state unemp. rate	0.0489*	0.132***
	(0.0263)	(0.0418)
Min. nearby unemp. rate	0.0671**	0.0525
	(0.0315)	(0.0481)
Non-exp. outmig. rate (2012)	1.058	1.716
Observations	2985490	1085956
Clusters	51	51
State fixed effects	X	X
Time fixed effects	X	X
Demographic controls	X	X
U-rate controls	X	X
State-specific trends	X	X

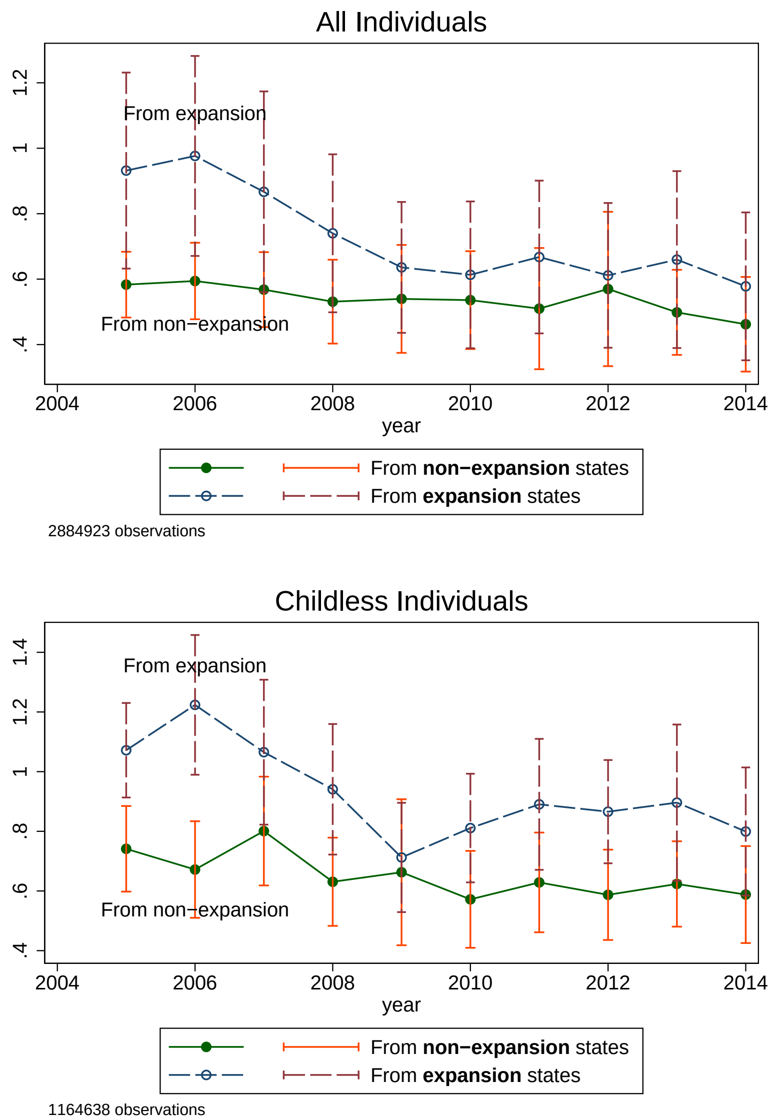
Notes: Each column reports the full results (aside from the fixed effects and the state-specific trends) from a regression of the following form, estimated at the individual level on data from the analysis sample, from 2008 to 2014, dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014, x_{irt} icludes demographic controls and unemployment rate controls (see text for details), and y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview. The bottom panel restricts the sample to childless individuals. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level; numbers in bracket indicate the bounds of the associated 95% confidence interval. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.5: Inter-regional migration rates among main sample from expansion and non-expansion states, among sample with origin in PUMA or PUMA of migration within 75km of an expansion/non-expansion border.



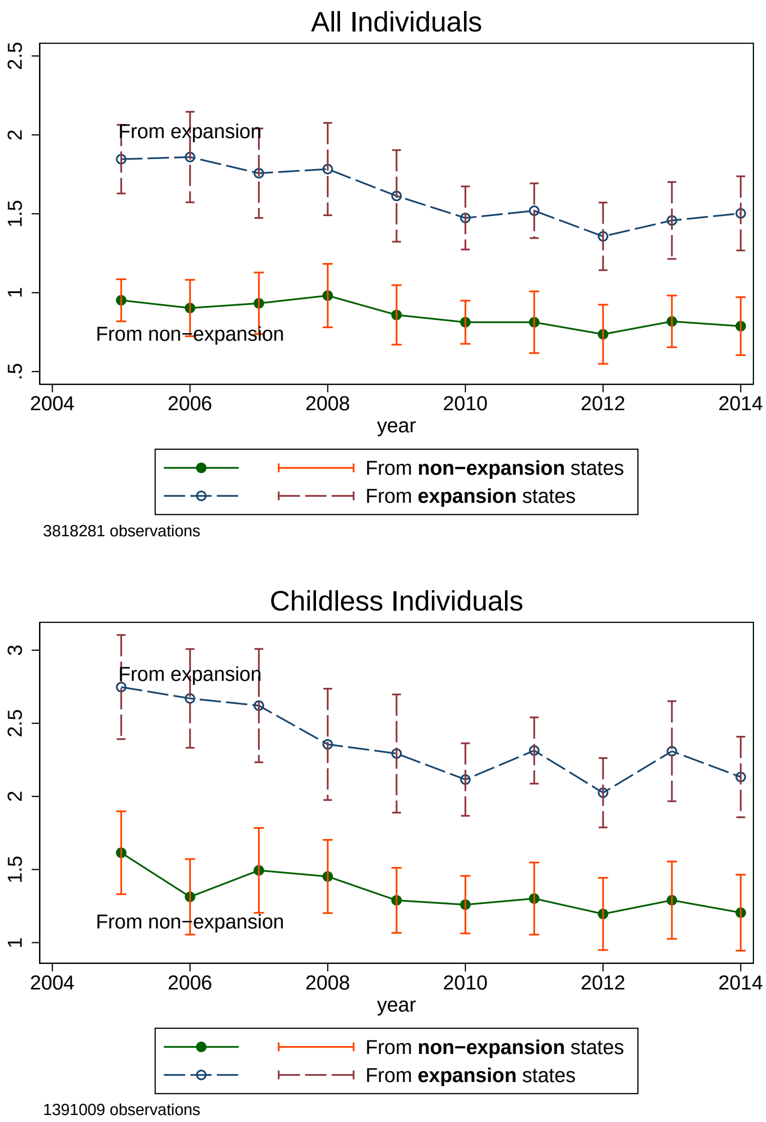
Notes: Each point represents the average inter-region migration rate (the ratio of inter-region outmigrants to total observations, using sample weights) of the analysis sample, from the region indicated. The two regions are the set of expansion states and the set of non-expansion states. The analysis sample includes individuals in families with earnings up to 138% of the federal poverty guidelines. See text for further details of sample selection. The brackets indicate 95% confidence intervals, robust to clustering at the origin-state level.

Figure 1.6: Inter-regional migration rates among education-selected sample, from expansion and non-expansion states



Notes: Each point represents the average inter-region migration rate (the ratio of inter-region outmigrants to total observations) of a sample selected on the basis of education (see text for details), from the region indicated. The two regions are the set of expansion states and the set of non-expansion states. The brackets indicate 95% confidence intervals, robust to clustering at the origin-state level.

Figure 1.7: Inter-regional migration rates, from expansion and non-expansion states, dropping early expanders



Notes: Each point represents the average inter-region migration rate (the ratio of inter-region outmigrants to total observations) of the analysis sample from the region indicated, dropping D.C. and the following six states: CA, CT, MA, MN, NJ, and WA (see text for details). The two regions are the set of expansion states and the set of non-expansion states. The brackets indicate 95% confidence intervals, robust to clustering at the origin-state level.

1.9 Addendum

The material in this chapter is a reproduction of material previously published in the *Journal of Public Analysis and Management*, first published online in November 2016 (Goodman (2017)). At the time of writing (and publication), 2014 was the latest year of American Community Survey (ACS) data availability.[32] For the sake of posterity, I have not amended the text of the prior sections of this chapter (except for updating references in cases when working papers have since been published or revised). In this section, I consider some additional specifications which make use of two additional years of data (2015 and 2016) which have since become available.

These extra years of data address a serious limitation of the original study. In the original study, I would have estimated a non-null result only if the migration response was evident in the 2014 ACS.[33] Yet there are a variety of reasons why we might expect a different result over more years of data. In particular, the migration response might be delayed if it takes time for individuals in non-expansion states to learn about the expansion taking place in expansion states. More simply, even if plans for migration change immediately, it might take time to actually implement those plans (e.g., to find housing, employment, etc.). Lastly, the addition of several years increases statistical

[32] In general, the ACS for year t becomes available on IPUMS in November or December of year $t + 1$.

[33] This is exacerbated by the rolling interview structure of the ACS, in which individuals are interviewed on a rolling basis throughout the year. The date of interview is not observed in the public data which I use.

power by reducing sampling variation.[34]

However, one cannot simply repeat the baseline analysis (e.g., using Equation (1.1)) adding the extra years of data, as five states have switched from being non-expansion states to becoming expansion states. Applying Equation (1.1) literally, one would find that interregion migration from non-expansion states did in fact increase in 2015 and 2016 – but only because the region of expansion states increased in size. I adapt the empirical strategy in two ways. First, I follow the strategy of Section 1.5, where I "drop" the five states whose expansion status changed. Specifically, I drop observations whose origin is in one of those five states; among observations with origins in any other state, I recode y_{it} equal to zero for all individuals whose destination was one of those five states. This is analogous to the strategy used in parts of **Table 1.7**, which considers sensitivity analyses with respect to states whose expansion timing was marginal, or which engaged in a partial Medicaid expansion prior to the ACA.

The top panel of **Figure 1.8** illustrates the results of this first strategy, analogous to **Figure 1.3**. Through 2014, this figure tells the same qualitative story as **Figure 1.3**: the trend in non-expansion-to-expansion migration is relatively parallel to the trend in migration in the reverse direction from 2008 onward, with little apparent effect in 2014. In 2015, there is (if anything) a wrong-signed effect: interregion migration from non-expansion states fell

[34]The additional power will be limited to the degree that state-specific shocks are correlated, since standard errors are clustered to take account of those correlations.

somewhat relative to interregion migration from expansion states. In 2016, this effect diminishes, but remains consistent with a null effect. The figure in the bottom panel, which restricts the sample to childless individuals, tells the same qualitative story.

Table 1.11 shows the regression results of Equation (1.1) after dropping states whose expansion status changed. Using data through 2014 (but still dropping states whose expansion states change), I find that migration from non-expansion states to expansion states fell by a statistically insignificant 0.049 percentage points relative to migration in the reverse direction (i.e., an effect of -0.049), compared to an effect of -0.005 in the baseline specification. This suggests that dropping states whose expansion status changed has only a modest effect on the results. When data from 2016 is added, the magnitude of the drop increases to 0.086; this wrong-signed effect remains insignificant. In the sample of childless individuals in the bottom panel, the estimates are -0.081 (through 2014) and -0.091, which are both insignificant, compared to 0.012 in the baseline specification. Thus, incorporating two extra years of migration data does not uncover a longer-term effect that is qualitatively different than the null effect found in the paper using data only through 2014.

This first strategy is most similar to the baseline strategy in the original study, but it requires dropping a non-trivial share of states. So, I also consider a second strategy which allows Medicaid expansion status to be evolving. In this strategy, I organize the data at the origin state (r), destination state (d), year (t) level. Define n_{rdt} to be the raw number of migrants (multiplied by the sample weight divided by 100) from origin r to destination d at time t (dropping the $r = d$ observations that correspond to not migrating interstate).[35] Further, define m_{dt} and m_{rt} as the binary expansion status of the origin or destination state at time t.[36] I then estimate the following regression, weighted by the mean value of $n_{rdt} + 1$ during the 2005 to 2007 period:[37]

$$\ln(n_{rdt} + 1) = \lambda_{rd} + \theta_t + \gamma_t m_{rt} + \beta_1 m_{dt}(1 - m_{rt}) + \beta_2 m_{rt}(1 - m_{dt}) + \epsilon_{rdt} \quad (1.4)$$

In this expression, the two regressors of interest are $m_{dt}(1 - m_{rt})$ and $m_{rt}(1 - m_{dt})$. The former will equal one if and only if d is an expansion state and r is a non-expansion state (and zero otherwise). The latter will equal one in the opposite case: when d is a non-expansion state and r is an expansion state. Thus, β_1 represents the implied (relative) effect on non-expansion-to-expansion migration and β_2 represents the implied (relative) effect on non-expansion-to-expansion migration. The attractive force would imply $\beta_1 > 0$

[35]Recall that the ACS is approximately a 1 percent random sample of households. The median sample weight in the data is 89 and the mean is 117.5, so dividing the sample weight by 100 means that each individual's adjusted sample weight will be one, to an order of magnitude.

[36]All states are classified as having $m = 0$ prior to 2014 (for simplicity, I ignore partial early expansions in places such as Massachusetts). I classify all expansion states in **Figure 1.1** to have $m = 1$ in 2014 and thereafter. Additionally, I classify Pennsylvania, Indiana, and Alaska as having $m = 1$ in 2015 and thereafter; expansions in these states became effective January 1, February 1, and September 1 of 2015, respectively (Kaiser Family Foundation (2018)). I classify Montana and Louisiana to have $m = 1$ in 2016 only; Montana's expansion went into effect on January 1, 2016, and Louisiana's went into effect on July 1, 2016 (Kaiser Family Foundation (2018)).

[37]It is necessary to add some constant to this weight, and to the argument of the logarithm in Equation (1.4) so that the logarithm is defined in the frequent case when there are zero migrants in the data from a given origin to a given destination in a given year. Adding a number other than one (or using the inverse hyperbolic sine) does not change the results qualitatively, though it does affect the standard errors.

and the retentive force would imply $\beta_2 < 0$. In the baseline specification of Equation (1.1), these two effects are not separately identified; in Equation (1.4), these effects are separately identified at the cost of imposing some additional assumptions.

To see what this regression equation is estimating, let's consider a simple case with two periods, $t = 0$ and $t = 1$; at $t = 1$, region E expands Medicaid and region N does not. Take the difference between $t = 1$ and $t = 0$ and rewrite the regression equation as follows:

$$\Delta \ln(n_{rd}+1) = \Big(\theta_1 - \theta_0\Big) + \gamma_1 m_{r1} + \beta_1 m_{d1}(1 - m_{r1}) + \beta_2 m_{r1}(1 - m_{d1}) + \epsilon_{rdt} \quad (1.5)$$

There are four parameters to estimate in this regression: $(\theta_1 - \theta_0)$, γ_1, β_1, and β_2. Additionally, there are four moments in the data: μ^{NE}, μ^{EN}, μ^{NN}, and μ^{EE}, where μ^{NE} is the mean value of $\Delta \ln(n_{rd} + 1)$ for $r \in N$ and $d \in E$, and the rest are defined analogously. Thus, similar to a standard difference-in-differences, these parameters each correspond to simple functions of the four μ values.

Consider **Table 1.12**. This table shows the four cells: from N or E, to N or E. Each entry corresponds to the implied conditional mean from Equation (1.5). Looking across the top row, we see that β_1 is equal to $\mu^{NE} - \mu^{NN}$, which is the increase in non-expansion-to-expansion migration relative to the increase in interstate migration within non-expansion states. Likewise, looking at the bottom row, β_2 is equal to $\mu^{EN} - \mu^{EE}$, which is the increase in expansion-to-non-expansion migration relative to the increase in interstate migration within expansion states.

Equation (1.4) generalizes Equation (1.5) by allowing more than one pre- and post-period, and allowing the N and E regions to evolve. But the intuition remains: the counterfactual is assumed to be given by the change in interstate migration that does not cross "regional" boundaries. This strategy thus requires that such interstate, intraregional migration is unaffected by treatment. There are several threats to this assumption, all of which will tend to bias the regression toward finding a spurious effect. First, suppose treatment increases μ^{EE} relative to the counterfactual of no Medicaid expansion anywhere. This could be the case if the Medicaid expansion reduces job lock (Garthwaite, Gross, and Notowidigdo (2014), Farooq and Kugler (2016)) and leads to increased mobility more generally. It could also be the case if the expansion induces some interstate migrants from expansion state s to change their destination state from some state in N to some state in E (e.g., an individual from Massachusetts deciding to move to the District of Columbia rather than Virginia). Such an increase in μ^{EE} would cause $\mu^{EN} - \mu^{EE}$ to fall, biasing β_2 downward. Since $\beta_2 < 0$ would be consistent with a retentive effect, this would bias the regression towards finding a spurious retentive effect. Similarly, suppose μ^{NN} falls due to the same destination-switching effect (e.g., individuals from Florida decide to migrate to the District of Columbia rather than Virginia); by the same argument, this will tend to bias β_1 upward, toward finding a spurious attractive force.

The top panel of **Table 1.13** reports these regression coefficients for the full sample.[38] Using data through 2014 (in the first column), I estimate that $\beta_1 = -0.060$, meaning that interregional migration from non-expansion states is estimated to have fallen by 6 percent (not percentage points) relative to interstate migration within the set of expansion states. Similarly, I estimate that $\beta_2 = 0.022$, which can be interpreted similarly as β_1, but applying to interregional migration from expansion states instead. Both of these effects are wrong-signed and insignificant. In the second column, I report estimates adding the two years of data through 2016. These results, -0.075 for β_1 and 0.032 for β_2, are quite similar to those in column 1. This echoes the findings in **Table 1.11** that adding two extra years of data does not uncover an effect that was not present in 2014.

Furthermore, while a point estimate of a 6 or 7.5 percent effect may appear substantial (even if wrong-signed), these cannot be directly compared to the estimates in **Table 1.3**. In the bottom panel, I translate the coefficients β_1 and β_2 into the same scale as **Table 1.3**: the implied increase in interregional migration in 2014 from non-expansion states, relative to the increase in interregional migration from expansion states. For the sake of this rescaling, I consider the effect when the set of expansion and non-expansion states are defined by their 2014 status, which allows me to treat the regions as time-invariant. As above, let μ^{NE} denote the mean interregion migration rate from non-expansion

[38] This strategy yields similar, though slightly more negative (i.e., wrong-signed), results when restricted to the set of childless individuals. These results are available upon request from the author.

states, and let μ^{EN} denote the mean interregion migration rate from expansion states. Both of these have two potential outcomes: when the expansion occurs ($D = 1$) and when it does not ($D = 0$). The coefficient estimated in Equation (1.1) estimates $(\mu^{NE}(1)-\mu^{NE}(0))-(\mu^{EN}(1)-\mu^{EN}(0))$. By contrast, the coefficient β_1 in Equation (1.5) estimates approximately $\frac{n_{rdt}(1)-n_{rdt}(0)}{n_{rdt}(0)}$ for $r \in N$ and $d \in E$. The coefficient β_2 estimates a similar object for $r \in E$ and $d \in N$. Because of linearity, the implied effect on μ^{NE} (i.e., $\mu^{NE}(1) - \mu^{NE}(0)$) is equal to $\beta_1\mu^{NE}(0)$ and the implied effect on μ^{EN} is equal to $\beta_2\mu^{EN}(0)$. I use the observed 2008 to 2012 values as a proxy for $\mu^{NE}(0)$ and $\mu^{EN}(0)$. These baseline interregional migration rates are also shown in the bottom panel.

After this rescaling, I find that these coefficients imply that non-expansion-to-expansion migration fell by 0.090 percentage points relative to migration in the opposite direction using data through 2014, and 0.117 percentage points using data through 2016. These are wrong-signed effects, which is guaranteed by the result that $\beta_1 < 0$ and $\beta_2 > 0$ are each wrong-signed. Furthermore, these two estimates are more negative than the -0.005 effect estimated in **Table 1.3**; however, both of the estimates in this table are insignificantly different than zero, and -0.090 and -0.117 are well within the confidence intervals of the baseline specification.

In sum, analyzing ACS data through 2016 yields results that are largely consistent with the previous analysis of data through 2014 only. The null effect of the Medicaid expansion on migration continues to hold in a specification quite similar to the specification in the original paper, as well as in a separate specification that is designed to exploit further changes in Medicaid availability. If anything, the result becomes more negative (wrong-signed): falling from -0.005 percentage points in the baseline analysis to -0.086 or -0.117 percentage points, depending on the specification. Thus, the addition of two extra years of data does not overturn the result of the original paper: that the 2014 Medicaid expansions did not appear to induce a substantial migration response.

Table 1.11: Baseline result adding 2015 and 2016 data years, dropping states that expanded in 2015 and 2016

	(1) Through 2014	(2) Through 2016
All individuals	-0.049	-0.086
	(0.102)	(0.086)
N	2,593,708	3,415,084
Childless individuals	-0.081	-0.091
	(0.140)	(0.139)
N	875,409	1,174,302

Notes: Each cell reports the coefficient on $nonexp_r * post_t$ from a regression of the following form, estimated at the individual level on data from the analysis sample, beginning in 2008 and ending in 2014 (first column) or 2016 (second column), dropping 2013: $y_{irt}) = \mu_r + \lambda_t + \beta * nonexp_r * post_t + x'_{rt}\gamma + \epsilon_{rt}$, where $nonexp_r$ is a dummy for residing in a non-expansion state 12 months prior to interview in year t, $post_t$ is a dummy for 2014 or later, y_{irt} is a dummy variable that equals one for individuals that report having moved from the set of expansion states to the set of non-expansion states, or vice versa, in the 12 months prior to interview. States that expanded in 2015 or 2016 (Alaska, Indiana, Louisiana, Montana, and Pennsylvania) are dropped from the sample, and moves into these states are disregarded. Regressions are weighted by ACS sample weights. Standard errors (in parentheses) are clustered at the origin-state level. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of sample selection * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table 1.12: Correspondence of sample means to regression parameters in simple model

	To N	To E
From N	$(\theta_1 - \theta_0)$	$(\theta_1 - \theta_0) + \beta_1$
From E	$(\theta_1 - \theta_0) + \gamma_1 + \beta_2$	$(\theta_1 - \theta_0) + \gamma_1$

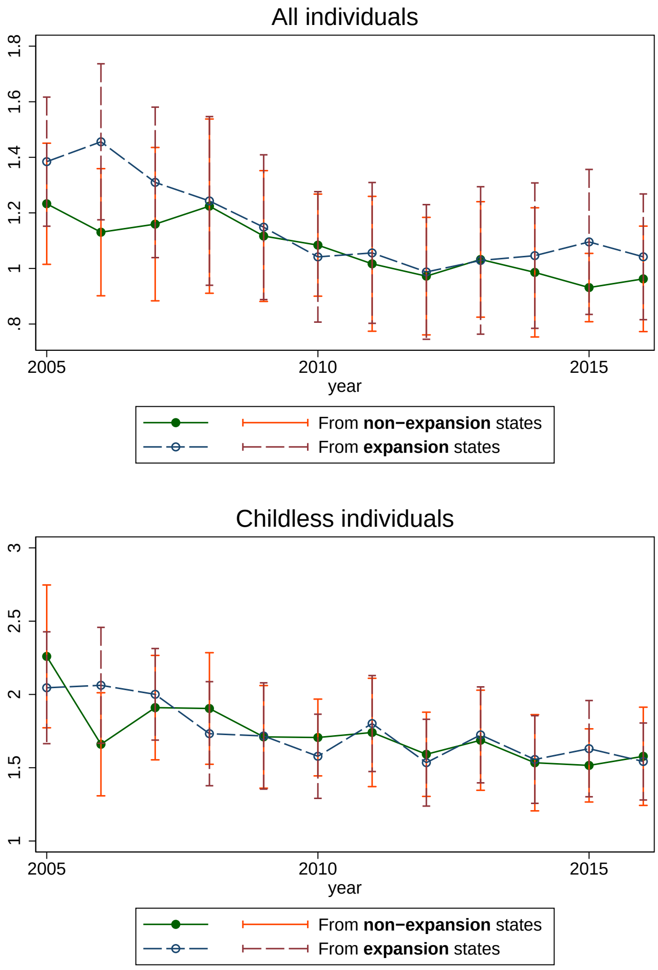
Notes: This table displays the correspondance of sample means in a simplified, two-period model to the parameters of Equation (1.5). "N" and "E" refer to the set of non-expansion and expansion states, respectively. For instance, the sample mean of interstate migration from "N" states to other "N" states will correspond to the regression parameter $\theta_1 - \theta_0$. See the text for further discussion.

Table 1.13: Estimated attractive and retentive forces, and implied effect on change in interregional migration: All individuals

	(1) Through 2014	(2) Through 2016
β_1 (Attractive force)	-0.060	-0.075
	(0.057)	(0.052)
β_2 (Retentive force)	0.022	0.032
	(0.048)	(0.021)
Base mig. rate from N to E	1.025	1.025
Base mig. rate from E to N	1.247	1.247
Rescaled effect (in p.p.)	-0.090	-0.117
	(0.087)	(0.076)
N	17,850	21,801

Notes: The top panel of this table reports estimates and standard errors from the following regression equation, estimated on data collapsed to the origin/destination level from 2008 to 2014 (Column 1) and 2008 to 2016 (Column 2), where m_{rt} and m_{dt} are dummies for a Medicaid expansion being in effect in origin r or destination d at time t, and n_{rdt} is the number of migrants from origin r to destination d at time t: $\ln(n_{rdt}+1) = \lambda_{rd} + \theta_t + \gamma_t m_{rt} + \beta_1 m_{dt}(1-m_{rt}) + \beta_2 m_{rt}(1-m_{dt}) + \epsilon_{rdt}$ Observations with $r = d$ are dropped. The regression is weighted by the average value of $n_{rdt}+1$ in 2005-2007. The coefficient β_1 represents the implied relative increase in migration from non-expansion states to expansion states when the expansion takes effect. The coefficient β_2 represents the implied increase in migration from expansion to non-expansion states when the expansion takes effect. $\beta_1 > 0$ would be consistent with the "attractive force" and $\beta_2 < 0$ would be consistent with the "retentive force." The bottom panel shows the base migration rate (from 2008 to 2012) from non-expansion states N to expansion states E, and vice versa, as defined by their 2014 status. Additionally, the bottom panel shows a rescaled estimate comparable to the estimates in **Table 1.3**. Standard errors (in parentheses) are clustered twoway at the origin and destination levels. The analysis sample includes individuals in families with earnings less than 138% of the federal poverty guidelines. See text for further details of this specification, how the estimates are rescaled, and other sample selection restrictions. * p < 0.10, ** p < 0.05, *** p < 0.01.

Figure 1.8: Inter-regional migration rates among main sample, from expansion and non-expansion states, through 2016 (dropping states which expanded in 2015 and 2016)



Notes: Each point represents the average inter-region migration rate (the ratio of inter-region outmigrants to total observations, using sample weights) of the analysis sample, from the region indicated. The two regions are the set of expansion states and the set of non-expansion states. The analysis sample includes individuals in families with earnings up to 138% of the federal poverty guidelines. This graph drops states that expanded in 2015 or 2016 (Alaska, Indiana, Louisiana, Montana, and Pennsylvania). See text for further details of sample selection. The brackets indicate 95% confidence intervals, robust to clustering at the origin-state level.

Chapter 2

Unemployment Insurance Generosity and Aggregate Employment

2.1 Introduction

During the Great Recession, existing law and new acts of Congress led to the most dramatic expansion in the generosity of unemployment insurance (UI) benefits in U.S. history[1]. In most states, eligible job losers saw their maximum benefit duration rise from the usual 26 weeks to 99 weeks. Continuously from November 2009 through March 2012, the maximum benefit duration exceeded 90 weeks when averaged across states, except for a few small lapses. In com-

[1]The second largest increase provided a temporary increase in unemployment duration of 65 weeks in 1975 following the passage of the Special Unemployment Insurance Extension Act.

parison, during a previous spell of extended benefits in response to the 2001 recession, this average rarely exceeded 40 (Farber & Valletta (2015)).

This unprecedented UI expansion—and its variation across states in magnitude and timing—provides a unique opportunity to study the aggregate employment effects of UI benefit duration. In this paper, we examine the effect of UI duration on aggregate employment during the Great Recession using state-level expansions and contractions in UI generosity. We use county-level monthly employment data from late 2007 until the end of 2014. We provide transparent evidence on employment dynamics around sharp and durable changes in UI benefits across counties that were otherwise very similar, and provide a reconciliation of the differences in findings across existing papers.

While a large body of research has studied the effect of UI duration on the labor supply and job search behavior of individuals, the effects of the benefit extension on aggregate employment may be quite different from the micro-based estimates. Keynesian theory predicts a positive employment effect of UI provision during recessions via stimulating aggregated demand (Summers (2010), Congressional Budget Office (2012)). In contrast, search-and-matching models suggest that extensions could raise reservation wages and lead to lower vacancies and employment (Mitman and Rabinovich (2014)). Finally, if jobs are rationed, the decreased search from increased UI generosity during downturns may have only limited effects on aggregate employment due to increased labor market tightness (the "rat race" phenomenon)—implying a smaller macro effect than micro effect (Michaillat (2012); Landais et al.

(2015); Lalive et al. (2015)). Unfortunately, a small set of recent empirical papers has delivered a mixed verdict on the size of the macro effect of the policy (Chodorow-Reich and Karabarbounis (2016); Coglianese (2015); Hagedorn, Manovskii, & Mitman (2016); Hagedorn, et al (2016); Johnston & Mas (2016)).

We begin by showing that the structure of UI extensions that occurred during the Great Recession makes our task quite difficult: federal policy expanded a state's UI duration automatically when unemployment in that state was high, leading to reverse causality. To address this mechanical endogeneity, we compare neighboring counties located on opposite sides of state boundaries.[2] We show that this border-county-pair (hereafter BCP) strategy substantially reduces the endogeneity problem, mitigating negative pre-existing employment trends in counties that subsequently experienced greater expansions in maximum benefit duration. In addition to OLS specifications that make use of all variation in state-level UI duration over the entire period, we also provide an instrumental variables estimate using variation induced solely by national level policy changes—namely the November 2008 expansion and the December 2013 expiration of the Emergency Unemployment Compensation (EUC) program. These national level policy changes are less endogenous to employment changes between neighboring counties than variation resulting from the move-

[2]This border-county-pair strategy was first used in Dube, et al. (2010) to study minimum wage policies, which change discontinuously at state borders. Note that the same problem of mechanical endogeneity does not arise when studying the effects of the minimum wage, as statutory wage rates are not directly tied to measures of state level unemployment. However, minimum wage policies can also be subject to endogeneity bias through political economic channels, and more generally may be correlated with spatially varying confounders.

ments in state-level unemployment rates. At the same time, the bite of the policy differed across state borders, which allows us to use the BCP strategy in conjunction with the IV approach. We show changes in aggregate employment during the 12 months before and after these expansion and expiration events; we also combine the data for both events to produce a pooled IV estimate.

Our main results are as follows. We find no evidence that UI benefit extensions substantially affected county-level employment. For the full sample OLS regressions, our point estimates for the effect of expanding maximum benefit duration from 26 to 99 weeks range from 0.21 to 0.43 percentage points of the employment-to-population (EPOP) ratio. These estimates are not significantly different than zero, and they allow us to rule out negative effects on EPOP greater than -0.48 percentage points at the 95% confidence level. For comparison, the total change in EPOP over the course of the Great Recession was about -3 percentage points in our sample.

Our IV estimates that specifically use variation from the national level policy changes in 2008 and 2014 reach a similar conclusion. For the 2008 IV estimation, the point estimates also indicate positive impacts on EPOP as a result of the UI expansion, but the standard errors are much larger. For the 2014 IV analysis, however, the impacts are estimated with more precision: the point estimates are -0.02 and -0.18 percentage points of EPOP, suggesting a very small negative impact on employment. When pooled over both events, our point estimates for the effect of increasing the maximum benefit duration from 26 to 99 weeks range between -0.07 and 0.14. While the IV estimates are somewhat less precise (especially for the 2008 expansion event), the most precise pooled estimate rules out effects more negative than -1.31 percentage points of EPOP from a 73-week increase in maximum benefit duration, at the 95 percent confidence level. Similarly, the estimates from the 2014 expiration event rules out effects more negative than -1.20 from the same policy change.

These conclusions are reinforced when evaluating dynamic evidence from our distributed lag specifications. For the full sample, we find that employment remained essentially unchanged over a 36 month window that includes 24 months after treatment. In particular, we see no trends prior to treatment, indicating that neither endogeneity nor policy anticipation confound our estimates. Event studies for the 2008 introduction and 2014 expiration also show qualitatively similar results. Taking into account the micro-econometric estimates of labor supply from other studies, we back out ranges of potential Keynesian multipliers that would be consistent with our macroeconomic estimates. Our macro employment estimates are consistent with a range of positive fiscal multipliers centered near 1 when we consider typical labor supply estimates from the UI benefit expansion—as found in many of the studies using data from the Great Recession.

A number of recent papers have exploited the panel variation across U.S. states over time in benefit duration during the Great Recession to study (micro-level) labor supply behavior. Rothstein (2011) uses data from the Current Population Survey (CPS) and variation from the uneven roll-out of extended benefits across states and finds that UI extensions were responsible for an increase in unemployment of 0.2 percentage points.[3] In concurrent work using similar variation, Farber & Valetta (2015) find that the availability of extended benefits increased the unemployment rate by 0.4 percentage points. Farber et al. (2015) find similar results when they exploit variation in UI generosity that arises due to the phase-out of extended benefits in 2012-2014: the effect of UI on duration to re-employment is small. Evaluating a sudden reduction in benefits in Missouri, Johnston & Mas reach a different conclusion: they find that newly unemployed workers who are eligible for 16 fewer weeks of UI (due to starting their claim shortly after a policy change) were 10 percentage points more likely to be employed starting in the quarter immediately after the policy change took place.

In contrast to the large empirical literature on the micro-level labor supply elasticity, there are relatively fewer papers that have estimated the macro-level impact of unemployment insurance on overall employment. The papers most closely related to ours are Hagedorn, Karahan, Manovskii, & Mitman (2016)—hereafter HKMM—and Hagedorn, Manovskii, & Mitman (2016)—hereafter HMM. Like us, these papers use a BCP strategy; HKMM provide evidence complementary to us that the BCP strategy mitigates the endogeneity problem. However, they both estimate large negative effects of UI on aggregate employment. HKMM find that the expansion of UI during the Great Recession from 26 to 99 weeks increased the unemployment rate by 80%, which is an effect on unemployment that is roughly comparable to the unemployment

[3]This calculation is made for December 2010.

growth that actually occurred during the Great Recession itself; they interpret this result as an explanation for the slow recovery in the unemployment rate in the years after the trough of the Great Recession. HMM study the 2014 expiration of EUC and find that that expiration was responsible for the creation of approximately two million jobs. This effect would translate into a 1.1% decrease in employment as a result of the expansion of UI from 26 to 99 weeks, which corresponds to about one third of the employment decrease of the Great Recession as measured in our data set.

However, our results are quite different from those in HKMM and HMM, despite employing apparently similar strategies. In **Section 2.8.1**, we compare our results to both HKMM and HMM and we discuss in detail what accounts for the substantial differences in our respective estimates. In summary, with respect to HKMM, we have found that a few factors explain the bulk of the difference between our two sets of results. First, our dependent variable is constructed using county-level employment data from the Quarterly Census of Employment and Wages (QCEW), which is derived from administrative filings. HKMM and HMM, in contrast, use as their primary dependent variable the county-level unemployment rate from the Bureau of Labor Statistics LAUS program, which is partially model-based. Second, we handle the dynamics of the treatment effect differently. HKMM quasi-forward difference their dependent variable, and scale up their estimate to deduce the effects of a permanent change in policy. In contrast, we use a less parametric distributed lag framework to document the dynamics of the employment response in a transparent fashion over a window spanning from a year prior to treatment to two years following treatment. This provides clear evidence on endogeneity concerns, policy anticipation, and the actual impact on employment over the two years following the policy change. We also replicate HMM and find that our replication of their estimates for the 2014 expiration of the extended benefits fall close to zero when we use the most recent LAUS data, which were substantially updated in a 2015 redesign of the LAUS estimating procedure. Additionally, in an event study specification, HMM estimate a substantial negative employment effect using QCEW data. These results seem primarily driven by their choice of auxiliary parametric assumptions—namely their use of a county-specific polynomial trend model, estimated over a long time horizon. Instead of relying on a parametric counterfactual, we show that our treatment and control units across the border exhibited parallel trends prior to the expiration, display no jump at expiration and continue in parallel fashion after expiration—implying little employment effect.

More recently, two working papers have estimated the macro effect by exploiting variations in state-level UI extensions coming from measurement error in the total unemployment rate. Coglianese (2015) uses the variation between the CPS-measured unemployment rate and a constructed unemployment rate from UI records as an arguably exogenous shifter in the maximum benefit duration. Using a conceptually similar strategy, Chodorow-Reich & Karabarbounis (2016) use the variation in benefit duration coming from the gap between real-time and subsequently revised official unemployment rates.

Both Chodorow-Reich & Karabarbounis (2016) and Coglianese (2015) find very small effects of UI extensions on aggregate employment. One limitation of the measurement error based approach is that the policy changes they study are less durable than the changes we examine in this paper and thus the external validity may be more limited. However, the very different types of variation leveraged across our two sets of papers makes them complementary. Our findings are also consistent with Marinescu (2017), who finds that UI benefit extensions during the Great Recession decreased job applications but not posted vacancies, implying a modest impact of the extensions on overall job finding and unemployment rates. Finally, in their case study of Missouri, Johnston & Mas (2016) find substantially larger, negative, macro employment effects than we find in this paper. Their macro estimates are similar in size to their micro estimates. Our approach differs from their macro estimates primarily in that we aggregate across many different benefit extensions and reductions and that our analysis uses variation across border counties rather than neighboring or similar states.

The remainder of the paper is structured as follows: In Section 2.2, we discuss important institutional details of the unemployment insurance extensions during the Great Recession that are critical for our identification strategy. In Section 2.3, we discuss our data. In Section 2.4, we discuss the identification challenges we face in our estimation and present our methodological approaches. In Section 2.5, we present our empirical results. In Section 2.6, we compare our macro estimates of UI expansion on employment with micro-level estimates based on labor supply elasticities, and back out an implied fiscal multiplier. Finally, in Section 2.7, we conclude.

2.2 Unemployment Insurance Background

The Great Recession saw a dramatic expansion of unemployment insurance benefits in all states. In part, this expansion occurred due to policies that were put in place prior to the Great Recession. However, Congress also passed legislation extending the maximum duration of unemployment insurance. In a majority of states, maximum benefit duration increased from 26 weeks to a maximum of 99 weeks depending on the state of the local labor market. In this section, we describe these extensions and how they were rolled out across states. It is precisely these differences across states—and in particular neighboring states—which we exploit in our identification of the impact of unemployment insurance benefit duration on employment.

2.2.1 Extended Benefits (EB)

Historically, when not in recession, most U.S. states have provided a maximum of 26 weeks of unemployment insurance to job-losers. At the onset of the Great Recession, in 2008, only two states offered more than 26 weeks of regular benefits. Massachusetts had a maximum of 30 weeks of UI benefits and Montana had a maximum of 28 weeks and no states offered less than 26 weeks.[4]

However, since Congress created the Extended Benefits (EB) program in 1970, maximum benefit lengths increase automatically when unemployment is high and growing. At a minimum, in states where the Insured Unemployment Rate (IUR) exceeds 5%, and the IUR is at least 1.2 times the IUR in the previous two years, claimants are eligible for 13 additional weeks of UI after the expiration of regular benefits.[5] The same law also provides two optional "triggers," which can be adopted by states at their own discretion. The first trigger provides for 13 weeks of EB for states whose IUR exceeds 6% (regardless of the change in the IUR over time). The other optional trigger is based on the Total Unemployment Rate (TUR): the trigger provides for 13 weeks of EB when both (1) the TUR exceeds 6.5% and (2) the current TUR is at least 1.1 times its value in the prior two years. States adopting this second trigger must provide 20 weeks of EB when (1) the TUR exceeds 8%, subject to the same growth-over-time requirement.[6] States can adopt zero, one, or both optional triggers, but no more than one trigger can be "on" at any point in

[4]Not all claimants are eligible for the maximum number of weeks of benefits. In most states, individuals with relatively weak recent labor force attachment are eligible only for a fraction of the maximum weeks of benefits. Throughout this paper, we abstract from this complication by focusing on the maximum UI duration. Our estimates, therefore, can be seen as an intention to treat effect. Johnston & Mas (2016), using micro-data from Missouri, find that approximately 70% of UI claimants had sufficient labor force attachment to be eligible for the full 26 weeks of regular benefits from 2003-2013.

[5]The Insured Unemployment Rate (IUR) is, roughly, the ratio of current regular UI claimants to the number of UI-covered jobs. The Total Unemployment Rate (TUR) is the usual "unemployment rate": i.e., the ratio of unemployed persons to persons in the labor force.

[6]From December 2010 through the end of 2013 (a period in which the unemployment rate remained high but was generally not growing), states were allowed to apply a three-year lookback period instead of a two-year lookback period for the purpose of determining growth over time.

time, meaning that the number of weeks of EB is capped at 20.

Normally, the costs of EB are shared equally between the federal and state governments. As a result, many states did not have statutes activating the optional EB triggers at the onset of the Great Recession. However, after the passage of the American Recovery and Reinvestment Act (ARRA), the federal government paid for the full amount of EB extensions. Some states (mostly deeply conservative ones) nonetheless declined to activate the optional triggers. For example, while Mississippi had a TUR of well over 8% continuously from January 2009 through October 2016, peaking at over 11% in 2010, they were never eligible for EB because the insured unemployment rate never went above 5.6% and the state declined to enact the optional triggers. Thus, different states had different numbers of weeks of EB in part due to differences in the state unemployment rates and in part due to state policy differences. The federal government maintained its full support of EB until the end of 2013 when it returned to the default equal cost sharing rule.

2.2.2 Emergency Unemployment Compensation (EUC)

In response to the first signs of a weakening labor market, on June 30, 2008, Congress and President Bush created the Emergency Unemployment Compensation (EUC) program. At first, EUC provided for 13 additional weeks of benefits for all UI-eligible unemployed workers.[7] The Unemployment Compen-

[7]To be more precise, this legislation—and all subsequent legislation related to EUC—provided for increases in benefit lengths equal to the lesser of (1) a specified number of weeks or (2) a fraction of the number of weeks of regular benefits. For the initial legisla-

sation Extension Act of 2008 was then signed into law by President Bush on November 21, 2008. It augmented the EUC program while also creating the first differences across states in their access to the EUC extensions. It authorized 20 weeks of EUC for all states (an increase from 13) and an additional 13 weeks for those with a total unemployment rate exceeding 6%.[8] These additional weeks were organized into "tiers": Tier 1 corresponded to the first 20 weeks of EUC, while Tier 2 corresponded to the baseline 20 weeks plus an additional 13 weeks. During this period, a state with 26 weeks of regular benefits could qualify for up to 79 weeks total of benefits. Then, on November 6, 2009, the Worker, Homeowner, and Business Act of 2009 further increased maximum UI duration. Tier 1 remained in place. However, Tier 2 was increased from 13 to 14 weeks and extended to all 50 states. The law also added Tier 3, providing 13 additional weeks to states with a TUR of greater than 6%, and Tier 4, providing 6 additional weeks for states with a TUR of greater than 8.5%. After the passage of this law, states had access to a maximum of 99 weeks of benefits. This schedule remained in place, with the exception of temporary lapses, until early 2012, when Congress enacted laws that slowly began to phase out EUC.[9]

---

tion in June 2008, the specified number of weeks was 13 and the fraction of the number of weeks of regular benefits was 50%. For the vast majority of states that had regular benefits greater than or equal to 26, the specified number of weeks was the binding factor. For those states with fewer than 26 weeks of regular benefits, the percentage of regular benefits was always binding. In this paper, we code the weeks available under EUC exactly as specified in the law; however, in the discussion that follows, we discuss only the specified number of weeks, which applies to states with at least 26 weeks of regular benefits.

[8]A state could also have become eligible for 33 weeks with a sufficiently high IUR; in practice, the IUR trigger was never binding.

[9]There were four lapses in EUC that occurred in 2010, arising due to political disagreements regarding the extension of the program. The longest such lapse lasted from May

On February 22, 2012, Congress passed and the President signed The Middle Class Tax Relief and Job Creation Act of 2012 which slightly lowered the generosity of the EUC in a gradual way, first starting on May 27, 2012, and then again on September 2, 2012. By September 2, 2012, Tier 1 had been scaled back to 14 weeks and was still available to all states. Tier 2 remained at 14 weeks but again became available only to states with a TUR of greater than 6%. Tier 3 was scaled back from 13 to 9 weeks and the state TUR threshold was raised to 7%. Finally, Tier 4 was increased to provide 10 extra weeks for states with a TUR of above 9%. The program finally came to an end at the end of December 2013.[10] In total, over the Great Recession, individuals in qualifying states received up to 99 weeks of unemployment insurance. Compared to the baseline of 26 weeks, this is an increase of 73 weeks; so the maximum UI benefit duration in some qualifying states increased by almost 300%.

30, 2010 to July 18, 2010. In each of the lapses, beneficiaries were paid retroactively for any weeks of missed payments. Furthermore, during these lapses, the funding rules for EB reverted to their pre-ARRA levels, which led many states to suspend EB payments during these lapses as well.

[10] Upon the expiration of EUC at the end of 2013, EUC beneficiaries immediately stopped receiving benefit payments. Prior to the final expiration, however, the phase-out was more gradual. If a state "triggered-off" a certain tier, people who had already qualified for a given tier were allowed to finish that tier. However, beneficiaries were not allowed to move to the next tier. One exception, discussed in the following subsection, is North Carolina, which lost access to all EUC money as of July 1, 2013. In our econometric specifications, our duration variable is the maximum duration available in a given month for a new entrant into unemployment. Thus, we do not distinguish between gradual phase-outs and sudden benefit cessations.

2.2.3 Changes in State-Level Regular Benefits

In addition to changes in federal policy and changes in state unemployment rates which triggered changes in unemployment benefit generosity, during our sample period, UI duration was also influenced by state-level policy changes. Starting in 2011, some states began to lower maximum duration for regular state-level benefits below the usual 26 weeks. Arkansas reduced its maximum benefit duration to 25 weeks and both Missouri and South Carolina to 20 weeks in 2011. Then, in 2012, Florida, Georgia, Illinois and Michigan reduced their maximum benefit duration. Michigan lowered it to 20 weeks, while the other three made it contingent on the state unemployment rate. North Carolina also reduced its regular benefits to 20 weeks; additionally, North Carolina reduced the weekly benefit amount from \$535 to \$350, which violated its agreement with the Department of Labor. For this reason, all EUC benefits immediately expired in North Carolina, which caused its maximum benefit duration to fall by 53 weeks. The duration of regular benefits fell further in North Carolina in 2014, as it was also set to be contingent on the state unemployment rate.

2.2.4 Variation Between Neighboring States

Importantly, the path of benefit extensions—from regular benefits, EB, or EUC—often differed markedly across neighboring states. These differences across neighboring states were largely a result of differences in state unemployment rates, but also to some degree due to variations in state policy. It is precisely these time-varying differences across neighboring states that we use for our identification strategy. In **Figure 2.1**, we graphically show the evolution of the benefit generosity over time nationally, which strongly (negatively) co-varies with the national employment-to-population ratio.[11] In **Figure 2.2**, we show the differences across neighboring counties in the numbers of weeks of available unemployment insurance, where the reported difference is between "high" and "low" benefit duration counties, defined by comparing the average duration in the treatment period (2008m11-2013m12) versus the the prior 12 months (2007m11-2008m10) when these differences were typically zero or very small. Prior to November 2008, most counties had access to an identical amount of unemployment insurance, with the exception of those in Massachusetts and Montana. Afterwards, however, some neighboring states (and thus neighboring counties across state borders) started offering different lengths of maximum benefit duration. The average gap between states with longer versus shorter total duration within the county pairs rose to nearly 12 weeks by late 2011, before declining to an average gap of near zero with the expiration of EUC in December 2013. This variation over time is used in our full panel estimates. We also use the national level policy variation due to the the November 2008 expansion, and the late 2013 expiration, of the EUC program as instruments for our IV strategy. In **Figure 2.3**, we show a map of the counties that had different generosity levels right before the EUC expiration

[11]Our measure of EPOP is below the US DOL measure. This is largely because our measure is based upon UI employment, and thus excludes those in the informal sector as well as the self-employed. Additionally, we calculate EPOP by dividing employment by the 15+ population in the county, rather than the 16+ population used by the DOL.

in December 2013. **Figure 2.15** shows the analogous map for the variation created by expansion of the EUC program in November 2008.

2.3 Data

We use county-level employment data from the Quarterly Census of Employment and Wages (QCEW). The QCEW data is based on ES-202 filings that nearly all establishments are required to file quarterly with their state government, for the purpose of calculating UI-related payroll taxes. These employment and earnings counts are shared by the states with the Bureau of Labor Statistics, which releases the data at the county-industry-month level. Since 98% of jobs are covered by unemployment insurance, these payroll counts constitute a near census of employment and earnings. There are some limitations: the QCEW does not capture workers in the informal sector or the self-employed, and it misses the small number of workers who participate in their own unemployment insurance system, such as railroad workers and workers at religiously-affiliated schools. Importantly, the QCEW covers both private and public sector employment.[12] The QCEW provides total employment for each month at the county level. In our baseline estimation, we require that each county be in the data set in every month. This excludes four counties for which there is at least one month in the sample where the QCEW does not report data due to confidentiality problems with disclosure. This occurs only in

[12]We focus our analysis on total employment (the sum of private and public sector employment), though we do provide results on private employment as a robustness check.

counties with very low population. In our robustness section, we additionally report estimates using the full unbalanced panel.

We divide employment by population of those 15 and older, which we obtain from the census at the annual level and interpolate log-linearly within each year. Prior to estimation, we seasonally adjust our dependent variables by subtracting off the county-month specific mean of the variable in question, where this mean is calculated over the period 1998-2004.[13] As we show later in the paper, however, our results are robust to using raw rather than seasonally adjusted data.

Our data on the number of weeks of regular benefits comes from Department of Labor reports which are issued biannually.[14] To account for occasional changes in the numbers of weeks of regular benefits that occur during the intervening period, we augment these data with online searches of news media and state government websites. We obtain information on EUC and EB from the trigger reports released by the Department of Labor, available at `http://www.oui.doleta.gov/unemploy/claims_arch.asp`. These reports provide the number of weeks of EB and tiers of EUC available for each state, in each week. When a change in weeks of benefits happens within a month, we assign the time-weighted average of the maximum duration to that month.

As discussed above, there were several lapses in the EUC program during

[13]For the sake of summary statistics and the small number of specifications we estimate without county fixed effects, we add back the overall mean level of EPOP for each county measured over the 1998-2004 period.

[14]`http://www.unemploymentinsurance.doleta.gov/unemploy/statelaws.asp`

2010. In the popular press, expectations were that these lapses would be reversed, and that the original EUC benefit durations would be reinstated. This is in fact what did happen. In our baseline specifications, we treat the lapses as true expirations—that is, those county-by-month observations are coded as having EUC equal to zero. However, we show in robustness checks that our estimates are not substantially affected if we code the benefit durations for these few months as having remained unchanged at their pre-lapse level.

We also use a list of all contiguous county pairs that straddle state borders; this data comes from Dube, et al. (2010). In our baseline specifications, we have a total of 1,161 county-pairs.

In addition, we obtain county level unemployment and employment data at the quarterly level from the Local Area Unemployment Statistics (LAUS) published by the Bureau of Labor Statistics. We obtained the most current data (as of November 10, 2016) via `http://download.bls.gov/pub/time.series/la/`. We additionally obtain a vintage series of county unemployment rates and employment (prior to the March 2015 redesign) via FRED. This is the main data source used by HKMM and HMM, and we use it as part of our reconciliation exercise in **Section 2.8.1**.

2.4 Research Design

2.4.1 The Identification Problem

To credibly estimate the effect of UI extensions on aggregate employment, we need to address a serious problem of reverse causality. Negative employment shocks that raised the unemployment rates were likely to mechanically raise the maximum benefit duration within the policy environment during the Great Recession. **Figure 2.1** illustrates the identification problem facing researchers when estimating the effect of UI extensions on employment. Between 2008 and 2014, we see a U-shaped time path of maximum benefit duration, along with an inverted-U shaped time path for the employment to population ratio.[15] However, it would be naive to assume that this correlation is causal in nature. A closer look confirms that the decline in employment in 2008 preceded the EB and the EUC tier extensions. Similarly, employment was already on the mend well before the 2014 EUC expiration occurred. It is possible that UI extensions were responsible for some of the decline and some of the persistence in the high unemployment rates the U.S. experienced in the 2009-2013 period. However, as **Figure 2.1** highlights, it is likely that some or much of this relationship reflects a mechanical endogeneity of UI maximum benefit duration to the state of the economy.

While the endogeneity problem is most obvious when considering time se-

[15] To be consistent with our baseline regression specifications, this figure shows the time series of EPOP and duration taken as an unweighted average of counties.

ries variation, a differences in differences (or the classic two-way fixed effects) strategy is unlikely to eliminate the endogeneity bias. On the one hand, there was a substantial amount of variation in UI generosity over time and differentially across US states, making it feasible to use panel variation in UI duration. However, the assumption that states which saw larger increases in the maximum benefit duration had parallel employment trends with states which experienced smaller increases is unlikely to hold due to the mechanical endogeneity: the rules of EUC and EB provide for longer benefits in a given state when the unemployment rate in that state is higher. Locations which switch into offering higher benefit duration will likely be locations in decline, and locations that switch into offering lower benefit duration will be locations in recovery—likely causing a bias in the two-way fixed effects estimate.

We explicitly demonstrate the scope of this endogeneity problem by showing how high-treatment counties—i.e., counties that would eventually experience a large increase in the maximum benefit duration—had very different employment trends prior to treatment as compared to other counties. For this exercise, we construct a time-invariant, continuous measure of the average treatment intensity for each county, $treat_c$. This is defined as the difference in time-averaged maximum benefit duration in a given county during the "treatment period" (i.e., between November 2008 and December 2013) versus the 12 months prior (i.e., between November 2007 and October 2008).[16] For exam-

[16]This "non-treatment" value will in general not be equal to 26, since it includes the period from July to October 2008 when all states were eligible for 13 weeks of EUC.

ple, if a state's average maximum UI duration during the treatment period was 90 weeks, and the average maximum benefit length in the 12 non-treatment months was 30 weeks, it would have a value of $treat_c$ equal to 60 weeks. For ease of interpretation, we rescale this variable by dividing it by 10, so that a value of 1 corresponds to a difference of 10 weeks of treatment, which is roughly equal to the mean difference in duration between neighboring counties which straddle state borders during the treatment period. We then estimate the following model over the 2004m11-2008m10 period, i.e., the four years preceding the introduction of differential UI benefits:

$$E_{ct} = \alpha \times treat_c \times t + \lambda_c + \theta_t + \epsilon_{ct} \tag{2.1}$$

where t is time measured in months divided by 48.[17] λ_c is a set of county fixed effects, while θ_t is a set of common period fixed effects. Our estimate of α thus measures the difference in the linear employment trend between high- and low-treatment counties prior to November of 2008. For this specification, we cluster our standard errors at the the state level. The first column of **Table 2.1** shows our estimate for $\hat{\alpha}$. The estimate, significant at the 1% level, implies that EPOP declined by 0.78 percentage points in the four years prior to November 2008 in counties that would subsequently receive an additional 10 weeks of benefits. This result is consistent with the mechanical endogeneity problem discussed above, and casts doubt on the assumption of parallel trends

[17]Note that there are 48 months in this sample, so the date variable equals (essentially) zero at the start of the sample and one at the end.

across counties prior to increases in benefit duration.[18]

2.4.2 Border county pair strategy

The failure of the two-way fixed effects strategy motivates us to restrict our sample to contiguous county pairs which straddle state borders (Dube et al. (2010, 2016)) and estimate the effects within border county pairs. The main idea behind this strategy is that neighboring counties in adjacent states are reasonably well matched. Dube et al. (2016) show that adjacent county pairs straddling state borders are much more alike in terms of levels and trends in covariates than are randomly matched pairs of counties. However, while adjacent counties are likely to face similar economic shocks as each other, their UI maximum benefit durations will be driven by their respective states' unemployment rates and policy choices—which may be quite different. Therefore, by focusing on comparisons between border counties, we are able to account for all confounders that vary smoothly geographically, and better account for the mechanical endogeneity problem that plagues the two-way fixed effects approach. **Table 2.2** shows that the treated and control counties were quite similar: pre-existing characteristics seem relatively balanced between the high-treatment and low-treatment counties within pairs.

For each month t, our border county pairs (BCP) data is organized to have two observations in each pair p—one for each county c of the pair. Note that this also means that a given county c appears in the data k times (for

[18]We show results from a two way fixed effects model in **Table 2.12.**

each month t) if it borders k counties in adjacent states. Before describing in detail our key empirical specifications, we first use this BCP data to show that within-pair variation dramatically reduces the problem of pre-existing trends. We re-estimate a regression of EPOP on the time-invariant average treatment intensity, $treat_c$, and county fixed effects, similar to **Equation (2.1)**. But now, instead of a single set of period effects, we include a full set of pair-period fixed effects, ν_{pt}. This sweeps out the variation between pairs, and only uses within-pair variation to identify α.[19]

$$E_{cpt} = \alpha \times treat_c \times t + \lambda_c + \nu_{pt} + \epsilon_{cpt} \tag{2.2}$$

As before, the estimation period runs from November 2004 to October 2008. The coefficient α has a similar interpretation as in the prior strategy, but now measures the differential pre-existing employment trends by treatment status within each adjacent county pair. The results in Column 3 of **Table 2.1** show that for the sample of border counties, the differential pre-existing trend within county pairs (-0.24) is much closer to zero and statistically insignificant, in contrast to the estimates from the two-way fixed effects model using the same sample (-0.98). This constitutes very clear evidence that the estimates using neighboring counties as controls are likely to exhibit less bias than those from the two-way fixed effects model. Moreover, the standard error from the BCP

[19] With two observations within each pair-period group, this approach gives the identical coefficients as if we dropped the pair-period fixed effects and instead (1) took the spatial difference of the dependent variable and main independent variable across each county pair p at each time t, and (2) replaced county fixed effects by pair fixed effects.

model (0.29) is not dramatically larger than that of the two-way fixed effects model (0.21), suggesting that it is a reduction in bias and not statistical power that drives the changes in statistical significance in **Table 2.1**.[20]

While the evidence on pre-existing trends from **Table 2.1** show that the BCP strategy is a very important improvement over the two-way fixed effects model, we may worry about remaining endogeneity bias, especially given the explicit reverse causality in this context. This motivates us to implement an additional data-driven refinement to the BCP strategy. In particular, we drop the quartile of pairs with the largest absolute differences in pre-existing EPOP trends over the 2004m11-2008m10 period. These BCPs appear to be more poorly matched in that the counties in these pairs exhibit qualitatively different trajectories prior to the UI extensions, and these trajectories may be mechanically correlated with subsequent UI duration.[21] Hereafter, we refer to this specification as trimming our sample based on pre-treatment trends, or PTT-trimming. Column 4 of **Table 2.1** shows $\hat{\alpha}$ for the PTT-trimmed sample and confirms that removing the worst-fitting quartile further reduces the extent of pre-existing trends to -0.11.

In this paper, we report estimates using several different types of regres-

[20] This evidence is complementary with the evidence provided in Section 4.3 of HKMM. HKMM find substantially larger estimates of the effect of UI on unemployment when their border pair sample is replaced by a "scrambled border pair" sample, in which pairs are formed randomly (rather than by reason of geographical adjacency). HKMM argue (and we agree) that this is indicative of the role played by the BCP strategy in reducing mechanical endogeneity.

[21] Even if economic conditions evolve continuously across state borders, the statistics for a given border county will measure an average of economic conditions some positive distance away from the border. This might be a concern for geographically large counties in the western United States. In our robustness section, we show that dropping pairs whose centroids are more than 100 km apart has little effect on our estimates.

sions. First, to visually show how employment evolves on the high-treatment versus low-treatment sides of the border, we estimate a model using the same time-invariant average treatment intensity, $treat_c$, that we used above for the assessment of pre-existing trends. We regress EPOP on a a set of interactions $treat_c \times \mathbb{1}\{t = s\}$ variables, where $\mathbb{1}\{t = s\}$ is an indicator for date s. In the full sample, we omit the variable corresponding to October 2008. We additionally control for county fixed effects λ_c and pair-period effects ν_{pt}. The estimating equation is as follows:

$$E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt} \qquad (2.3)$$

Since $treat_c$ is a continuous, time-invariant measure, the coefficients β_s trace out how EPOP evolves in the treated versus control sides over time, as compared to a base period of October 2008, the month before the first cross-state variation in federal UI benefits in our sample.

While the time-invariant treatment measure is useful for a qualitative, visual assessment of how employment evolved on the two sides of the border, it does not use the timing of policy changes with any precision. Our baseline BCP-FE specification equation uses a normalized maximum benefit duration (in weeks), D_{ct}, to estimate the following equation:

$$E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt} \qquad (2.4)$$

We normalize D_{ct} by dividing the maximum benefit duration by 73, to make

β interpretable as the change in EPOP from the median expansion in duration that took place in the Great Recession.[22] Again, we include county fixed effects λ_c to account for persistent differences between the two members of the pair,[23] and pair-period effects ν_{pt} to sweep out between-pair variation. Clearly, this strategy still relies on D_{ct} being uncorrelated with η_{cpt}, i.e., $E(D_{ct}\eta_{cpt}) = 0$, but now this assumption needs to hold only within a local area that is likely to be experiencing more similar economic shocks. The third column of **Table 2.1** shows why we believe this assumption is closer to the truth in the county-pair setting relative to the two-way fixed effects setting. **Equation (2.4)** is estimated for both the baseline sample of all border county pairs, as well as the PTT-trimmed sample of county pairs. The baseline regression is estimated over the period from November 2007 to December 2014, which includes the period of differential EUC (November 2008 - December 2013) as well as 12 months prior and 12 months after.

We also present the dynamics of employment around the time of the policy change. There are two specific aims that underlie this analysis. First, we wish to use the leading coefficients to detect pre-existing trends and assess the validity of the research design. Second, we wish to assess possible anticipation or lagged effects of the policy. To this end, we utilize a first-differenced distributed lags specification with a set of 11 monthly leads and 24 monthly

[22] All but two states had 26 weeks of benefits prior to the onset of the Great Recession, and the median as well as mode for state UI duration was 99 weeks from November 2009 until April 2011.

[23] We replace county fixed effects with county-cross-county-pair fixed effects in the small number of specifications in which the panel is unbalanced.

lags, along with the contemporaneous benefit duration, D_{ct}. This specification allows us to focus on employment changes within the 36 month window around the time of treatment.

Our estimating equation for the dynamic specification is:

$$\Delta E_{ct} = \sum_{k=-11}^{24} \beta_k \Delta D_{c,t-k} + \nu_{pt} + \epsilon_{cpt} \tag{2.5}$$

Successively summing the coefficients traces out the cumulative response to a one-time, permanent unit change in D: $\rho_\tau = \sum_{k=-11}^{\tau} \beta_k$ represents the cumulative response at event time, τ.[24] For ease of interpretation, we center the cumulative responses around a baseline of the month just prior to treatment, $\tilde{\rho}_\tau = \rho_\tau - \rho_{-1}$, which imposes that $\tilde{\rho}_{-1} = 0$. We plot the centered cumulative response $\tilde{\rho}_\tau$ by event time, along with the associated confidence intervals below.

While the border county pairs strategy provides greater internal validity, one potential concern is about the representativeness of border counties. Summary statistics in **Table 2.11** confirm that border counties are relatively comparable to the full set of counties, indicating that the sample restriction for purposes of internal validity comes at minimal sacrifice of external validity.

[24]Note that β_k is the response associated with D_{t-k}. This indexation convention allows us to index the coefficients by event time.

2.4.3 Instrumental variables estimation: EUC Policy Changes

Estimating **Equation (2.4)** by OLS exploits all of the variation in maximum benefit duration induced by both policy changes (EUC, state adoption of optional EB triggers, and state changes to regular benefits) and endogenous movements in state unemployment rates across various thresholds (from EUC and EB triggers). That is, our OLS specification has the undesirable feature that it exploits variation in benefit duration in a given month which was caused by a change in contemporaneous state-level unemployment. By only comparing adjacent border counties, we are likely to reduce the scope of the endogeneity problem, since the employment shocks affecting policy are from the state as a whole, while we are accounting for the county's employment shock by comparing it to its cross-state neighbor. Nonetheless, to the extent that endogeneity bias may remain, we can further reduce it by restricting the variation we use to national-level policy changes. Counties within a border pair are less likely to have systematically different employment trends when UI duration changes due to national policy than when one county's state is triggering on or off of EB or an EUC tier. We therefore develop an instrumental variables approach that isolates the effects of cross-border changes in benefit duration that are triggered by persistent changes in national policy, and not by contemporaneous economic shocks.

The first policy change that we use is the passage of the Unemployment

Compensation Extension Act (UCEA) in November of 2008, which granted states 20 weeks of federally funded benefits, or 33 if the total unemployment rate at the time exceeded 6%. This led to an increase in UI benefit durations which varied across states, introducing the first across-state variation in EUC availability in our sample.[25] The second national policy change we use is the expiration of the EUC program in December 2013, which led to a larger reduction in UI duration which also varied across states.

Of course, the change in national policy creates variation precisely because there were differences in the *level* of unemployment across states. For the 2008 policy change, states that had a TUR exceeding 6% saw a bigger increase in benefit duration than states with a lower TUR. Similarly, for the 2014 expiration, states with higher unemployment rates experienced larger reductions in benefits. While high and low unemployment states very well may have been on different trajectories around these two events, the BCP strategy is arguably better able to account for such trends compared to times when the policy change is directly induced by changes in state unemployment rates.

For our IV specifications, we use a two year window—one year on each side of the national policy change. We regress EPOP on weeks of benefits, controlling for pair-period fixed effects and county fixed effects. We then instrument benefit duration with a variable that reflects only the change in duration caused by the EUC policy change. The instrument does not exploit variation caused

[25]Prior to UCEA, variation in federally provided benefits existed in two states: North Carolina and Rhode Island were eligible for 13 and 20 weeks of EB, respectively, at the time of the policy change. No other state was eligible for EB at that time.

by EB triggerings, EUC triggerings, and state-level policy changes. Our two stage least squares estimation strategy is thus given by the set of equations:

$$E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt} \tag{2.6}$$

$$D_{ct} = \beta_z z_{ct} + \rho_c + \gamma_{pt} + \epsilon_{cpt} \tag{2.7}$$

where the instrument z_{ct} reflects the instantaneous change in the maximum UI duration available in the county due to the national EUC policy change. The instrument z_{ct} is defined as follows:

$$z_{ct} = \begin{cases} D_c^{08} & \text{Nov. 2007 - Oct. 2008} \\ D_c^{08} + \delta_c^{08} & \text{Nov. 2008 - Oct. 2009} \\ D_c^{13} & \text{Jan. 2013 - Dec. 2013} \\ D_c^{13} - \delta_c^{13} & \text{Jan. 2014 - Dec. 2014} \end{cases}$$

For the 12 months prior to the 2014 policy change, we set the value of z_{ct} to equal the number of weeks of UI available in the last week of December 2013 (immediately prior to the EUC expiration), D_c^{13}. For the remaining 12 months in the sample, we subtract from D_c^{13} the number of weeks of benefits lost as a result of the EUC expiration (δ_c^{13}), and set z_{ct} equal to this value.[26] For the two year window around the 2008 policy change, the instrument is

[26]Therefore, the change in the instrument z_{ct} between December 2013 and January 2014 takes into account the decline in duration explicitly resulting from the EUC expiration, but not any contemporaneous changes in state-level regular benefits. In our robustness section, we show results from a specification where the instrument also takes into account the five state-level policy changes that occurred at the same time as the national policy change.

defined analogously, using the maximum UI duration available just before (D_c^{08}) and just after the introduction of the new EUC program. Therefore, the jump in z_{ct} that occurs in November 2008 (δ_c^{08}) exactly equals the differential number of weeks made available by the onset of the UCEA. We also pool both events together, and estimate this model using the 24 months of data around the 2008 onset along with the 24 months of data around the 2014 expiration.[27] For all of these specifications, we estimate the results using the complete baseline BCP sample as well as the refined (PTT-trimmed) sample. Because the EUC program in North Carolina expired at the end of June 2013 (rather than December), we drop county pairs that include a North Carolina county from the 2014 subsample in the baseline analysis.[28]

We additionally show reduced form and first stage estimates underlying the IV regressions by month relative to the event. As with OLS, the dynamic specification is estimated in first differences:[29]

$$\Delta E_{cpt} = \sum_{\tau=-12}^{11} \beta_\tau \delta_{ct} \mathbb{1}\{eventdate_t = \tau\} + \nu_{pt} + \eta_{cpt} \tag{2.8}$$

[27]For this pooled specification, we allow the county fixed effects to vary across the two subsamples (that is, the county fixed effects are replaced with county-by-subsample fixed effects).

[28]In the robustness section, we report results from specifications which keep North Carolina as well as others which redefine the instrument for North Carolina to exploit variation from its earlier benefit cut.

[29]We note that estimating this model in levels (i.e., using E_{cpt} and D_{ct} and mean differencing) versus first-differences is immaterial in this case where we are estimating monthly coefficients, β_τ, over a fixed 24 month sample. Estimating the model in levels yields numerically identical estimates.

$$\Delta D_{ct} = \sum_{\tau=-12}^{11} \beta_{z\tau}\delta_{ct}\mathbb{1}\{eventdate_t = \tau\} + \gamma_{pt} + \epsilon_{cpt} \tag{2.9}$$

We define $\delta_{ct} = \delta_c^{08}$ for the 2007-2008 sample and $-\delta_c^{13}$ for the 2013-2014 sample, each divided by 10 for the ease of interpretation. As with OLS, the sum of coefficients $\rho_\tau = \sum_{k=-11}^{\tau} \beta_k$ and $\rho_{z\tau} = \sum_{k=-11}^{\tau} \beta_{zk}$ represent the cumulative response by event time. These represent the average within-pair differences in employment and the prevailing maximum benefit duration—over a 24 month window around the national policy change—for a pair in which the difference in the instantaneous increase in maximum benefit duration (due to the policy change) was 10 weeks. We omit the variable corresponding to $eventdate_t = -1$ (which corresponds to October 2008 and December 2013), meaning that the plotted coefficients are centered relative to date -1 leading values.[30]

It is useful to consider where our policy variation is coming from when using this IV approach along with BCP sample. Consider two adjacent counties, A and B, which followed similar employment trends prior to October 2008, but where side A saw a larger increase in benefit duration in October 2008 because it happened to be in a state with an already high state unemployment rate. Variation in policy, then, is coming largely from more negative *past* employment shocks in *other* counties in A's state—as compared to past employment shocks in other counties in county B's state. The same logic applies to EUC

[30]For ease of interpretation, we omit January 2014 instead of December 2013 in the first stage when constructing the graph that analyzes only the 2014 expiration event. This allows the graph to show a drop in relative benefits roughly from 10 to 0 rather than 0 to -10. As we do not report standard errors for this specification, this amounts to a simple vertical shift of the graph.

expiration in December 2014. The combination of more plausibly exogenous variation due to national policy changes with local cross-state comparisons guards against endogeneity bias by putting both geographic and temporal distance between shocks in employment in a county and the shocks that drive the policy.

2.4.4 Standard errors

Except where noted, our standard errors are clustered two-way at the state-pair level and at the state level. Clustering at the state-pair level is designed to account for common, serially correlated shocks to local economies. We also cluster at the state level to account for the mechanical correlation in error terms that is introduced when one county borders counties in at least two states (and thus appears in multiple state-pairs) as well as any state level shocks. Note that our clustering strategy fully accounts for the appearance of a single state multiple times in the border county pair sample.

2.5 Empirical Findings

2.5.1 Motivating graphical evidence

Figure 2.4 plots the regression coefficients for the time-invariant average treatment intensity measure, $treat_c$, period by period, using **Equation (2.3)**. The figure plots two sets of coefficients: one with EPOP as the dependent variable, and the other with maximum UI benefit duration as the outcome. This figure shows that the side of the border receiving a larger treatment (averaged over the full treatment period) experienced a slight decline in employment starting several years prior to treatment, though this pre-existing trend is not statistically significant. The differential employment trend greatly accelerated between 2009 and 2012—at a time when the UI extensions are implemented. This might indicate a causal effect of the UI extensions. However, contrary to that interpretation, employment continued to fall at a similar rate on the side receiving a larger treatment in the post-2011 period when UI generosity difference within the pair was in decline.

Figure 2.5 shows the results of the same analysis using our refined PTT-trimmed set of border county pairs, where we exclude the pairs with the largest differences in pre-existing trends. The findings are reinforced when we consider this refined BCP strategy. Over the 2004-2014 period, employment on the side of the border receiving greater treatment remained essentially unchanged, even as benefit duration rose sharply in late 2008, and then dropped sharply in late 2013. This figure provides compelling visual evidence of the validity of the refined BCP design (no pre-existing trends), and that any causal employment effect of of the policy is likely to be quite small.

Together, the two figures convey several important features of the data and the research design. First, when using the baseline BCP sample, the monotonic decline in employment on the high-treatment side of the border throughout the entire period—both when UI benefit duration difference within the pair is increasing and when it is decreasing—previews our regression results that overall employment effects are likely to be modest in that specification as well. Second, trimming on pre-treatment trends eliminates not only trends prior to treatment but also the secular decline in EPOP post-treatment. These findings suggest that the secular employment decline was due to poor match quality in a minority of observations rather than a causal effect of treatment.

2.5.2 Main Estimates

We present our full-sample OLS estimates for the time period from November 2007 to December 2014 in the top panel of **Table 2.3**. This panel reports two columns of regressions estimating **Equation (2.4)**. The first column reports results using the baseline (i.e., untrimmed) BCP sample and the second column reports results using the sample that we refined based on pre-existing trends (the PTT-trimmed sample). The point estimate for the baseline BCP sample is 0.430. Recall that we normalized D by dividing the maximum benefit duration by 73 weeks, so this allows us to interpret the coefficient as the estimated impact on EPOP from an increase in maximum benefit duration from 26 to 99 weeks. Consequently, the baseline BCP estimate suggests that the 73 week increase in maximum benefit duration raised the EPOP ratio by 0.430 percentage points. The standard error is 0.466 and thus the estimate is not statistically distinguishable from zero. When we restrict the analysis to the PTT-trimmed sample in Column 2, the coefficient falls to 0.213. Even though the PTT-trimmed sample size is 25% smaller than the baseline BCP sample, the standard error for the PTT-trimmed estimate is smaller at 0.270: trimming on PTT rids the sample of poorly matched county pairs and thereby reduces residual variance. As a consequence, when moving from the baseline BCP to the refined BCP estimates, the maximal *negative* impact of expanding UI from 26 to 99 weeks which can be rejected at the 95% level of confidence falls in magnitude from -0.483 to -0.316.[31]

Figure 2.6 visually displays the employment dynamics around the treatment event in a transparent manner using the first-differenced distributed lag specification of **Equation (2.5)**. These estimates are useful for assessing policy anticipation and lagged effects of the policy, as well as possible biases in the research design arising from pre-existing trends. The figure shows the cumulative response in employment ($\tilde{\rho}_\tau$) starting 12 months before treatment, and extending up to 24 months after. Recall that these cumulative responses are centered at event time $\tau = -1$, so the estimates of confidence intervals for $\tilde{\rho}_\tau$ are expressed relative to the month before treatment. The top panel displays the coefficients for the full sample of BCPs, while the bottom panel displays them for the PTT-trimmed sample. For both specifications, during the twelve months prior to treatment, i.e., between $\tau = -12$ and -1, there is little change in employment. The leading values of the cumulative responses

[31] **Table 2.12** presents results from the two-way fixed effects model for the all-counties sample and the border county pair sample. The point estimates are somewhat more negative, consistent with the problem of pre-existing trends documented in **Table 2.1**. Nonetheless, the unweighted estimates which are most comparable to **Table 2.3** are modest in magnitude: -0.385 (with a standard error of 0.355) for the all-counties sample, and -0.382 (with a standard error of 0.361) for the border counties sample.

range between -0.321 and 0.403, and are never statistically distinguishable from zero. Overall, the distributed lag specifications produce little evidence to indicate reduced hiring in anticipation of the policy change.

Following treatment, both the baseline BCP specification and the PTT-trimmed specification show no change in employment over the 24 months following the policy change. The cumulative responses are typically positive and not statistically significantly different from zero. Even as the precision declines for longer lags, 12 months after the policy change, we can nonetheless still rule out employment effects more negative than -0.6 with 95 percent confidence for both specifications. Overall, the dynamic evidence from the OLS model suggests little employment change in the year prior to treatment (e.g., through anticipation), or during the two years following the policy change.

The instrumental variables estimates from **Equations (2.6) and (2.7)** are presented in the bottom three panels of **Table 2.3**. In panel 2 of **Table 2.3**, we report our pooled results using both the 2008 introduction (i.e., a positive treatment) and the 2014 expiration of the EUC (i.e., a negative treatment). For our preferred PTT-trimmed specification, the first stage F-statistic for the excluded instrument is 262.3, indicating that the instantaneous changes due to the national policy changes were responsible for a sizable fraction of the variation in benefit duration over the event window; the first stage coefficient is 0.842.[32] Our preferred PTT-trimmed second stage estimate is close to zero

[32]If the only changes in duration in the year before and the year after policy change were due to the policy change itself, the first stage coefficient would be 1.

(-0.069), with a standard error of 0.635. While less precise than the OLS estimate, these estimates using only national level policy changes in the PTT-trimmed sample can rule out employment reductions of -1.31 percentage points from the 73 week expansion of maximum benefit duration during the Great Recession. The point estimate from the untrimmed BCP sample is similar (0.143), though less precise with a standard error of 0.964.

To assess the employment dynamics around the national policy changes, **Figure 2.7** shows the first stage and reduced form estimates period by period around the event date, as compared to the values from the month just prior to treatment (i.e., -1). The EPOP difference between the two sides of the border is plotted on the left hand Y-axis, with the difference in maximum benefit duration plotted using the right hand Y-axis. The top graph uses the baseline BCP-FE sample while the bottom graph uses the refined PTT-trimmed sample. The dynamic evidence mirrors the numerical results in **Table 2.3**. At date 0, there is (by construction) a clear increase of approximately 10 weeks in the maximum benefit duration relative to the neighboring county.[33] Much of this increase in benefits persists over the following 12 months. There is little indication of a differential trend in employment prior to the national level policy changes, which provides additional validation for the IV coupled with the border county design. Importantly, employment remains fairly stable over the 12 months following treatment and we see little indication of job loss following

[33]The increase is not exactly 10 weeks because the policy changes in question did not occur precisely at the end of a calendar month.

the national level policy changes. Furthermore, the results are visually similar both in the baseline BCP-FE and the refined PTT-trimmed sample.[34]

The pooled estimates combine both the positive treatment in 2008 and the negative treatment in 2014. We also show the disaggregated effects from each of these treatments. The 2008 results using the 2007m11 to 2009m10 period are reported in the third panel of **Table 2.3**, and we show the corresponding graphical evidence in **Figure 2.8**. Again, there is a strong first stage (the F-statistic for the excluded instrument is over 40), though this first stage is substantially weaker than the pooled first stage or the 2014 first stage discussed below. As **Figure 2.8** shows, the duration differences created by the implementation of UCEA in 2008 were somewhat less persistent. The more limited persistence is also reflected in the first stage coefficient of 0.729, as shown in **Table 2.3**. This is unsurprising given the economic turbulence and resulting triggering that followed the UCEA of November 2008.

In general, the second stage estimates from the 2008 event study are fairly noisy. The estimate on the baseline BCP sample is 0.549, with a very large standard error of 2.515. The large standard error is likely because (1) there was a lot of variability in the drop in EPOP across counties during the early part of the Great Recession, substantially increasing error variance (reduced form), and (2) the duration differences created by UCEA were less persistent

[34]We also estimate the model using a sample trimmed based on trends estimated over the 2004m11-2007m10 period in order to address any concerns that PTT-trimming is mechanically eliminating anticipation effects. The graphical results, presented in **Figure 2.8**, are quite similar to the results presented in **Figure 2.16** Regression results are presented in the robustness section, below.

(first stage). Turning to our preferred PTT-trimmed sample, the coefficient falls to 0.198, while the standard error also halves to 1.265. While the standard error remains large, the PTT-trimmed sample is somewhat more precise due to a smaller residual variance. As shown in **Figure 2.8**, however, there is little indication of systematic employment changes—either in the year prior to the 2008 UCEA implementation, or during the subsequent year. Overall, while noisy, the estimates from the 2008 event (especially from our preferred, more precise trimmed sample) are broadly consistent with those from the pooled estimates and do not indicate substantial losses in employment from this policy change.

Panel 4 of **Table 2.3** reports our IV results from the 2014 elimination of EUC. The EUC program expired at the end of December 2013, leading to large reductions in UI generosity in almost every state. Importantly, some states experienced substantially larger reductions in benefits than others. For example, benefits were reduced by 47 weeks in Illinois, Nevada, and Rhode Island, but only by 14 weeks in Virginia, Iowa, New Hampshire, Minnesota and 10 other states. **Figure 2.3** shows a map of the reduction of UI duration at the end of 2013. As discussed above, North Carolina lost all EUC benefits and the maximum benefit fell to 20 weeks a full six months before the national EUC expiration. As a result, we remove North Carolina from our 2014 event study sample.[35] However, in the robustness section below, we show the results

[35]To be clear, in the pooled estimates reported above, we include North Carolina in the 2007-2009 portion of the sample but exclude it from the 2013-2014 portion.

from specifications in which North Carolina is included in the sample.

We show our results graphically in **Figure 2.9**. The figure does not show much of an effect on EPOP from the program expiration. Of note, the duration differences between county pairs were much more persistent (looking backward in time), mostly exceeding 80% of their immediate pre-expiration duration during the entirety of 2013. This explains why the first stage coefficient is much closer to unity: 0.915 for the baseline BCP sample and 0.903 for the PTT-trimmed sample. The first stage F-statistics are very high: 393 for the baseline sample and 424 for the PTT-trimmed sample. The point estimates are slightly lower than the pooled sample, at -0.024 and -0.182, respectively. However, the standard errors are substantially smaller than the 2008 analysis: 0.562 for the baseline BCP sample, and 0.521 for the PTT-trimmed sample. These estimates suggest a relatively precise null estimate of the effect of UI extensions on employment.

Although not statistically distinguishable from zero, the point estimates for the 2008 analysis are somewhat more positive than the 2014 estimates or the full sample OLS results. If these differences are real, and not merely noise, one speculative possibility is that the estimates from 2014 are less positive because they are estimated at a point in time when aggregate demand multipliers are lower.

Overall, both the OLS and IV estimates suggest that there was no sizable positive or negative employment effect of the 73 week increase in UI maximum duration during the Great Recession. This is true when we use all policy variation in our OLS specifications, or when we instrument the policy variation using national level changes. Our dynamic evidence suggests no employment changes for the first year and a half following the policy innovations. And when we consider our preferred refined BCP strategy that excludes some of the more poorly matched pairs, we find no evidence of employment changes up to 24 months following treatment.

2.5.3 Robustness of estimates

In this subsection, we perform a number of robustness checks. First, we show how our estimates vary with the sample period used in our estimation, and why we believe this validates our use of the refined BCP sample that trims on match quality. In the second subsection, we show how our refined PTT-trimmed results vary as we alter the trimming threshold. In the third subsection, we consider our results' robustness to a wide range of other specification choices and controls.

Choice of sample period

Table 2.4 shows results from the full sample OLS specification for alternative samples beginning in 2007m11, 2006m11, 2005m11, and 2004m11. The first column shows results for the baseline BCP sample and the second column shows results for the PTT-trimmed sample. Overall, the baseline BCP estimates range between 0.430 and -0.330, while the PTT-trimmed estimates range between 0.213 and 0.064. Importantly, while the estimates differ in size, we stress that none of the eight estimates shown in **Table 2.4** is statistically significant at conventional levels, and six of the eight are positive in sign.

At the same time, the baseline BCP estimates vary somewhat by sample, and these estimates decrease monotonically in the length of the window: the earlier the sample start date, the more negative the estimate. The gap between the estimate for the sample starting in November 2007 to the sample starting in November 2004 is non-trivial; it represents a differential impact of roughly 0.75 percentage points of EPOP from a 73-week increase in UI duration. Note that the pattern in the estimated effect is consistent with the presence of a downward trend in EPOP in treatment counties relative to control. As we discussed above, and as shown in **Figure 2.4**, between 2004 and 2008 we see a relative decline in EPOP on the side of the border that would eventually have higher UI duration. By pushing the start date further back in time, we are only adding data from the pre-treatment period; there is essentially no variation in UI benefits between 2004 and 2007. Adding observations from a time period when EPOP was relatively higher on the high-treatment side and when treatment was low makes the estimated treatment effect more negative. The fact that the estimated effect varies across the different sample periods leads us to believe that the baseline specification with pair-period fixed effects may reflect a degree of residual endogeneity. Put another way, a 2007m11-2014m12 sample frame – with twelve months before treatment begins and after treatment ends – ensures that any differential trends between counties is approximately orthogonal to D, our independent variable of interest. This orthogonality implies that differential trends have relatively little effect on our estimates. By contrast, with a larger amount of time before treatment than after treatment, these trends are no longer orthogonal to D, potentially leading to bias.

The variation in estimates is much smaller for the PTT-trimmed estimates: the 2007-2014 estimate is 0.213 and the 2004-2014 estimate is 0.064. We believe that this relative robustness to sample date validates the use of this refined sample (selected based on an absence of pre-treatment trends): even as the sample window becomes more asymmetric around the "treatment" period, the estimates do not change substantially, suggesting that differential trends are much smaller in magnitude in this sample. Additionally, the standard errors for the PTT-trimmed samples are also uniformly lower by between 16% to 42%, consistent with better match quality in the refined BCP sample.

Trimming on pre-treatment trends

The refined BCP strategy trims the pairs with the worst matches—25% of the sample with the biggest absolute differences in pre-treatment employment trends. In **Table 2.5**, we show how our four main estimates (OLS, 2008 IV, 2014 IV, and Pooled IV) vary as our threshold for trimming on PTT varies. We show estimates for different trimming thresholds across 7 rows. The rows are, respectively: no trimming, 10% trimming, 20% trimming, 25% trimming, 30% trimming, 40% trimming, and trimming at the median of the difference in PTT. The 25% trim is our main PTT specification from **Table 2.3**. In all four columns (i.e., for all 4 specifications), the range in the point estimates across trimming thresholds is below 1 standard error in magnitude. The coefficient estimates are fairly robust to changes in the trimming threshold. The standard error is minimized for the full sample at a 25% trim. It is minimized at a 10% trim for the pooled IV sample and the 2014 sample. It is minimized at a 30% trim for the 2008 sample. Thus, our choice of a 25% benchmark trim across all specification is a reasonable one.

Additionally, for all specifications, the primary impact of trimming seems to be a reduction in the standard errors by improving the match between high-treatment and low-treatment counties. It does not seem to systematically change the magnitude of the estimate in a positive or in a negative direction. The reduction in the standard errors is often up to 50% from the baseline sample. The one exception is the 2014 IV estimate where the maximum reduction across trimming thresholds is approximately 20%.

Other robustness checks

In **Table 2.6**, we consider a number of other robustness checks for our OLS estimates on the full 2007-2014 sample and for our pooled IV.[36] We do this both for the baseline BCP sample as well as the PTT-trimmed sample. The first row in the table reproduces the estimates from **Table 2.3**. Each of the remaining rows varies the specification, data, or sample as follows. We show estimates of impacts on private employment only. As an additional strategy to miti-

[36]The corresponding results for the separate 2008 and 2014 IV regressions are shown in **Section 2.8.3**.

gate residual mechanical endogeneity, we drop pairs containing counties that show a high correlation between county EPOP and the EPOP of its state over the 2004m11-2008m10 period ("correlation trimming"). Comparison within these county pairs should be less prone to contamination from state-specific employment shocks that endogenously determine state-level benefit duration. We include an (in sample) county specific linear trend (ISLT) control. We trim based on pre-treatment trends estimated over the 2004m11-2007m10 period (instead of 2004m11-2008m10) to address concerns that PTT trimming could be mechanically removing anticipation effects. Because the lapses (correctly) might not have been seen as changes because they were expected to be reversed in a very short period of time, we recode treatment during temporary lapses at the level of the duration during the last week before the lapses; we do not recode for the IV estimates because none of the lapses occur during the relevant sample periods. We also estimate using quarterly as opposed to monthly data: once using the same QCEW employment data but aggregated to the quarterly level, and once using quarterly employment statistics from a different data set, the Quarterly Workforce Indicators (QWI). We show results using data that have not been seasonally adjusted. To demonstrate that our controls are well matched to our treatments, we show robustness to restricting the sample to a plausibly better-matched group of pairs whose population centroids are less than 100 km apart. We also estimate a specification where we allow for imbalance in our panel by including counties with missing values in the sample. In addition, we show a pooled IV specification where we instrument using the total change in benefits rather than the change in benefits due solely to the expiration of EUC. In this case, the instrument includes the additional decreases below 26 weeks made by state governments in Florida, Georgia, Kansas, and South Carolina, as well as an increase from 26 to 30 weeks in Massachusetts. We also show three different specifications where we alter our baseline treatment of North Carolina, which lost access to EUC benefits earlier than other states.[37] Finally, as a further alternative, we use a log-log specification instead of the level-on-level specification used throughout the paper. We do this using both log employment and log EPOP as outcomes, but also report the EPOP-equivalent estimates in square brackets for comparability.[38]

For the OLS specifications in Columns 1 and 2, the range of the estimates from these changes is not substantial. The coefficients (or EPOP-equivalent coefficients as is the case when using logged outcomes) range between -0.145 and 0.692. In no case do we see any indication of substantial disemployment

[37]Recall that North Carolina lost access to EUC at the end of June 2013. This was a full 6 months before the other states lost access to EUC benefits, which means that North Carolina gets treated half way through the control period in the 2014 IV analysis. In our main specifications analyzing the 2014 EUC expiration, therefore, we drop all county pairs containing a county from North Carolina. We also drop North Carolina from the 2014 part of the sample in the pooled IV regression. As robustness checks, we drop North Carolina from the entire baseline BCP-FE full sample estimation as well as from the entire pooled IV specification. We also include North Carolina in the 2014 portion of the pooled IV specification. Finally, we retain the inclusion of North Carolina in the 2014 portion of the pooled IV sample but redefine the instrument, in North Carolina's case, to reflect the drop in EUC benefits for North Carolina in July 2013.

[38]For instance, the estimate of 0.006 in column 1 for log EPOP would imply that the expansion of UI from 26 to 99 weeks increased EPOP by $((\frac{99}{26})^{.006}-1)\times 42 = 0.35$ percentage points (since the unweighted mean EPOP in this sample is approximately 42), similar to the coefficients that we see in the level-on-level specification (0.430). The level equivalents for the log-log specification are displayed in brackets below the coefficient estimates. The level-on-level equivalents of the log employment estimates are quite close to the original estimates.

effects of the UI extensions. For the IV specifications in Columns 3 and 4, the estimates range between -0.147 and 0.930 for the baseline BCP sample, and between -0.406 and 0.659 for the PTT-trimmed sample. The greater variability for the IV is consistent with the IV estimates being more imprecise, and the standard errors are two to three times as large as the OLS counterparts. However in none of these cases are the estimates statistically distinguishable from zero.

In **Table 2.13**, we show the robustness checks for the 2008 and 2014 IV analyses separately. The results are largely similar to our pooled IV results, though the standard errors are significantly larger for the 2008 IV and often 30-50% smaller for the 2014 IV. The 2008 IV estimates are imprecise because the initial 2008 triggering explains less of the variation in treatment in the surrounding 2 year sample period. In addition, they are imprecise because of the large variation in EPOP during the onset of the Great Recession.

2.5.4 External validity: size and persistence of policy changes

One potential concern with our border county pair design—or any county panel design for that matter—is whether the differences in UI benefit duration between counties across the state border were sizable and persistent, especially as compared to the national level changes in benefit duration that took place during the Great Recession. **Figure 2.10** shows the distribution of differences in maximum benefit duration across county pairs and over time for the full sample. Here each observation is a county pair in a given week between November 23, 2008, and December 22, 2013. As the figure shows, around 40% of pair-week observations in this sample have no difference in UI benefit durations. However, nearly half of the observations have a benefit duration exceeding 10 weeks. To put this in perspective, a 10 week differential is almost 40% of the typical maximum benefit duration of 26 weeks that prevailed in all but two states prior to the Great Recession. Therefore, the gaps across state borders that we are evaluating are economically substantial. In **Figure 2.17,** we show that similar sized duration gaps existed between the two sides of the border just prior to the EUC expiration in 2014.

The gaps in UI benefit duration between neighboring counties across the border were substantial, but were they also persistent? **Figure 2.11** shows the mean benefit duration gap (as a share of the initial gap) by weeks following a particular event.[39] On average, ten weeks after the event, 70% of the original gap in maximum benefit duration between the two sides of the border remained in place. Even 52 weeks after the event, on average, more than 50% of the original gap in duration persisted across the border. Overall,

[39]In this analysis, all changes in relative benefit differences are treated as "events" or "shocks." With the data organized at the pair-by-shock (ps) level, we regress the change in relative duration on a set of $shock_{ps} \times eventdate_{\tau}$ indicator variables, where $shock_{ps}$ is the size of the initial shock and $eventdate_{\tau}$ runs from zero to 51 weeks after the initial shock. For instance, suppose at time t, county A increased duration from 53 to 63 weeks while county B held constant at 47 weeks, then $shock_{ps}$ would be equal to 10. The dependent variable in the regression (for $\tau = 0, 1, ..., 51$) would be equal to $D_{A,t+\tau} - D_{B,,t+\tau} - 6$, since the pre-shock difference was 6 weeks. Therefore, the regression coefficients trace out the share of the original shock that remains after τ weeks.

the evidence suggests that the benefit durations we are using for identification are not transitory policy shocks. The duration series in **Figures 2.8** and **2.9** show similar information for the specific 2008 and 2014 events.

We additionally show that the high average persistence of the policy shocks is not driven by a small number of cases but rather policy persistence was widespread across counties. In panel A of **Figure 2.18**, we show the share of counties where the duration gap continuously remained at least as large as the initial gap by weeks following the the 2008 event. The figure shows that after approximately 20 weeks, the initial gap remained in place or increased in about 60% of the county pairs; by 40 weeks, about 15% of the pairs retained the full gap. Panel B shows evidence for the 2014 expiration. Even 50 weeks before the EUC expiration, over 40% of counties had gaps in duration at least as large as the gap at the time of expiration. Thus, the 2014 event study estimates are based on the expiration of highly persistent differentials across county pairs.

Overall, while the cross sectional differences in size and persistence of the UI benefit durations are not as dramatic as the overall national level changes that occurred during the Great Recession, they are nonetheless quite substantial—especially for the 2014 expiration event. Moreover, the persistence of the events in our samples are quite a bit greater than those used in some of the other papers in the literature. For example, the measurement error based identification used in Chodorow-Reich & Karabarbounis uses treatment events whose half life is roughly 8 weeks (see their Figure 2). In contrast, as shown in our **Figure 2.11**, the half life of the typical event used for our baseline OLS estimate exceeds 52 weeks.

2.6 Rationalizing Macro and Micro Effects of UI Extensions

A higher benefit duration has an unambiguously negative labor supply effect through increasing reservation wages. In the UI literature, the micro-based estimates of extensions on employment reflect only these labor supply considerations. In this section we compute and interpret the gap between our macro estimate and some of the prevalent micro estimates from the literature.

How do the macro effects of UI extensions on employment that we estimate compare with employment change implied by micro-level labor supply elasticities? In order to answer this question, we first express both our macro estimates and the micro literature estimates in numbers of jobs. This entails multiplying our estimates (which are in terms of EPOP) by the 15+ population in 2012 (253 million) and the micro-estimates (which are in terms of unemployment rates) by the 2012 labor force (134 million). The gap between the macro and the micro estimates of the UI extensions on employment can be written as:

$$GAP = \Delta E_{MACRO} - \Delta E_{MICRO} = (\beta_{MACRO} \times P + \beta_{MICRO} \times L)$$

where β_{MICRO} is a micro estimate from the empirical literature of the impact of raising the UI benefit duration from 26 to 99 weeks on the unemployment rate, L is the size of the labor force (in 2012), β_{MACRO} is an estimate from this paper, P is the 15+ population in 2012, ΔE_{MACRO} is the predicted change in national employment from increasing UI benefit duration from 26 to 99 weeks using our estimates, and ΔE_{MICRO} is the predicted change in national employment from increasing UI benefit duration using micro estimates from the literature. In **Table 2.7** we report computations using 6 estimated micro responses to the impact of increasing UI duration from 26 to 99 weeks in the literature. Five of these are from four papers estimated using data from the Great Recession (Daly et al. (2012); Farber & Valletta (2015); Johnston & Mas (2016); Rothstein (2011)). Four of these numbers range between 0.1 to 0.8. Johnston & Mas (2016) is much larger in magnitude at 4.6. We also use one estimate from before the Great Recession which comes from Elsby et al. (201): 2.4.[40] In addition, we use two estimates of β_{MACRO} from Column 2 of **Table 2.3** (-0.069 and 0.213, rounded to -0.1 and 0.2 for simplicity). For each combination of estimates, we calculate the employment gap between the macro and micro employment estimates.

[40] As we noted in the introduction, Johnston & Mas (2016) provide a case study of Missouri where there was a sudden reduction in benefits, and find a much larger micro-level response than most of the literature. Besides providing labor supply based estimates, they also provide synthetic control and difference-in-difference estimates for aggregate employment effects from the benefit reduction. These macro estimates are similarly sized as their micro estimates, and are much larger than the macro effects that we find in this paper. Therefore, the size of the estimates from Johnston & Mas (2016) seem less about the micro versus macro effects than about the Missouri case study. Nonetheless, here we include the implied β_{MICRO} estimates from Johnston et al. study since those are specifically based on the labor supply response to the policy change.

Our macro estimates imply a range of employment change between -0.3 million and 0.5 million. In contrast, the range implied by the micro elasticities is -6.2 million and -0.1 million; excluding the Johnston and Mas estimate, the range is -3.2 million to -0.1 million. For 11 out of the 12 combinations of estimates, the predicted macro employment change is more positive than the predicted micro change, sometimes sizably so.[41]

One explanation for a more positive macro than micro effect is the Keynesian aggregate demand channel. UI puts cash in the hands of unemployed individuals whose earnings in the absence of UI payments are likely to be well below their permanent incomes. These individuals are likely to be liquidity constrained and thus a dollar of UI expenditures is highly likely to be consumed. Empirical work has shown that the marginal propensity to consume out of one-time tax rebates during the Great Recession was 25% (Sahm et al. (2012)). Though lower income individuals responded more, the differences were not large. However, economic theory suggests that liquidity constrained unemployed individuals should have a substantially larger response to cash receipts than other groups. If UI recipients spend most of their money on consumption, this can impact aggregate demand. The total impact will depend upon the fiscal multiplier, over which there is substantial disagreement among macroeconomists. For example, when analyzing the likely impact of the ARRA, the CBO estimated an output multiplier for UI benefits rang-

[41]The exception is when we take the lower bound of the β_{MICRO} estimate (0.1) and the lower bound of the β_{MACRO} estimate (-0.1).

ing between 0.4 and 2.1—with the larger estimate being more relevant when monetary policy is at the zero lower bound.

How large a fiscal multiplier is needed to rationalize the gap between the micro and macro effects of the UI extensions? For this back-of-the-envelope exercise, we assume that the gap between the micro and macro estimates, $\Delta E_{MACRO} - \Delta E_{MICRO}$, arises solely due to aggregate demand effects. Since the multiplier is the ratio of total dollars created to total dollars spent, we first convert the employment effect of increasing UI from 26 to 99 weeks into an impact on overall income, and then divide by UI expenditures. Our estimate of the change in total income is the product of the employment change (rescaling the percentage point change by $\frac{1}{100}$) and the ratio of output to employment $\left(\frac{Y}{E} = \$108,000\right)$.[42] National EB and EUC transfer payments between November 2008 and December 2013 averaged \$49.3 billion annually, and during this time period the average number of weeks of UI available was 74.4. In order to obtain an estimate of UI expenditures corresponding to an increase from 26 to 99 weeks, we scale the actual expenditure by $\frac{99-26}{74.4-26}$ ($\Delta B = \$49.3 \times 10^9 \times \frac{73}{48.4}$).[43]

[42] GDP per worker data from 2012 is from the World Bank: `http://data.worldbank.org/indicator/SL.GDP.PCAP.EM.KD?locations=US`. Note that this implicitly assumes that jobs created from the fiscal stimulus have mean productivity. Chodorow-Reich (2017) provides evidence supporting the validity of this approximation. Assuming that capital is fixed, but hours and employment adjust, he derives the following relationship between change in output and change in (headcount) employment: $\Delta Y \approx \theta \times (1+\chi) \times \frac{Y}{E} \times \Delta E$, where θ is labor's share, while χ is the elasticity of hours with respect to (headcount) employment. Given his estimates of $\chi = 0.5$ and $\theta = 0.7$, the constant-capital and hours adjustment channels cancel each other out, implying $\Delta Y \approx \frac{Y}{E} \times \Delta E$. He also validates the rough approximation using multipliers estimated from ARRA stimulus on state level employment and output. Similarly, Nakamura & Steinsson (2014) report both output and employment multipliers using defense spending shocks, and the magnitudes of both are are consistent with this approximation.

[43] We obtain the data for payments made through the EB and EUC programs from `http://oui.doleta.gov/unemploy/euc.asp`.

Dividing the estimated change in total income by the estimated UI expenditure gives our estimate of the fiscal multiplier, m_f.

$$\begin{aligned} m_f &= \frac{Y}{E} \times \frac{\Delta E_{MACRO} - \Delta E_{MICRO}}{100} \times \frac{1}{\Delta B} \\ &= \$108,000 \times \frac{(\beta_{MACRO} \times 253 + \beta_{MICRO} \times 134) \times 10^6}{100} \times \frac{1}{\$49.3 \times 10^9 \times \frac{73}{48.4}} \\ &= 3.7 \times \beta_{MACRO} + 1.9 \times \beta_{MICRO} \end{aligned}$$

In **Table 2.7**, we find that the implied fiscal multipliers using the first four micro estimates range between -0.2 and 2.3, centered around 1. However, when we use pre-Great Recession micro estimates, our implied multipliers are substantially larger, and range between 4.2 and 5.3. Finally, if we use the micro estimates from Johnston & Mas (2016), our implied fiscal multipliers are extremely large, exceeding 8. Since our macro effects are small, modest micro effects suggest a modest multiplier. However, large negative micro effects require a counterbalancing large fiscal multiplier to rationalize the small macro effect.

A caveat about our estimates is that our employment effects are estimated locally, and may differ from national multipliers for a number of reasons. First, a substantial fraction of the increased spending from UI extension is likely on tradable goods, much of which is produced outside of the local area. We are not capturing these demand leakages in our local analysis. Since a US county is substantially more open than the US as a whole, our local multiplier estimates are, *ceteris paribus*, likely to be smaller (and possibly substantially so) than national multipliers. Second, the multipliers estimated here are "transfer multipliers" as they are financed by transfers to the state from other states as opposed to through taxes or borrowing (Farhi & Werning (2016)). Therefore, the transfer multiplier may reflect a wealth effect which would not be present at the national level when the spending is deficit-financed, making the local multiplier larger than the national multiplier. Nonetheless, Farhi & Werning (2016) and Chodorow-Reich (2017) point out that as an empirical matter, externally funded transfer multipliers may provide a rough lower bound for the national, deficit-financed multipliers during liquidity traps. This is especially true when the transfer is not highly persistent, which was indeed the case for UI extensions. Overall, our estimates imply that a moderately sized, positive multiplier can rationalize the difference between the macro and the micro effects of UI, suggesting that the optimal benefit duration is likely to be countercyclical. This implication is consistent with the arguments in Landais et al. (2015) and Kroft & Notowidigdo (2016).

There are two other potential explanations for the gap between the micro and macro effects that come from recent work in search theory. The standard Pissarides (2000) search and matching model predicts that a higher benefit duration raises the negotiated wage, thereby reducing vacancies and employment through the job-creation effect (HKMM). However, if jobs are rationed, then a decrease in labor supply by some unemployed individuals from a more generous UI policy will tend to increase the job-finding probability of other unemployed workers, which can increase labor market tightness through the rat race effect (e.g., Mhciallat (2012), Landais et al. (2015)). From the search-and-matching perspective, the net effect on employment will be a combination of the direct labor supply effect, the job creation effect, and the rat race effect. The positive macro effect, ΔE_{MACRO}, cannot be explained by the labor supply (which is negative), the job creation (which is negative) and rat race effects alone (which is positive but merely attenuates the negativity of the former two effects). As a result, it is indicative of at least some positive stimulative effect that may offset the negative effects from job creation and labor supply effects. However, the imprecision of the gap between the micro and the macro estimates suggests caution against interpreting this evidence too strongly.[44] Better distinguishing the search and aggregate demand channels remains an important area for future research.

2.7 Conclusion

Despite a large literature that has evaluated the labor supply effects of unemployment insurance, the overall impact of the policy on aggregate employment is a relatively new and understudied area of research. Yet, this is an important question from a public policy perspective. If there are sizable negative effects of UI employment via labor supply, but these are counteracted by positive aggregate demand effects, the overall employment effects can be more positive

[44]Many of our implied employment effects are not statistically distinguishable at a 95% level of confidence from the micro effects. However, for our PTT-Trimmed full sample specification, the 90% confidence interval does not contain the employment impacts implied by Rothstein's upper bound or any of the more negative micro estimates in **Table 2.7**.

than what is implied by the labor supply estimates—making the policy more effective. Conversely, if the labor supply effects are small, but higher reservation wages fuels lower hiring and hence a higher unemployment rate, the policy can be less attractive than may initially appear.

In this paper, we add to the small but growing literature on the impact of UI on overall employment. We utilize variation across counties which straddle state borders where the states differ in their UI duration during the Great Recession. We find that this strategy substantially reduces likely bias from endogeneity that would plague a two-way fixed effects model assuming parallel trends across counties (or states) receiving differential treatment. To account for remaining endogeneity, we utilize a variety of strategies including refining our sample and focusing on variation driven by the national policy changes created by the introduction of differential EUC in 2008 as well as the expiration of the EUC program at the end of 2013.

Whether we use all policy variations, or whether we use variation induced solely by national level policy changes, most of our estimates are quite small in magnitude. Our OLS results using a refined border county pair design suggest the employment to population ratio rose by a statistically insignificant 0.21 due to the 73 week increase in benefits. The IV results that use the national policy variation from 2008 expansion and 2014 expiration of EUC suggests the EPOP ratio changed by -0.07. While the 95% confidence intervals for the OLS estimate rules out change in EPOP more negative than -0.32, the confidence bounds for the IV rule out changes more negative than -1.31. Across a variety of specifications and samples, our preferred point estimates suggest that the extension of unemployment insurance duration from 26 to 99 weeks during the Great Recession led to a change in EPOP between -0.18 and 0.43 percentage points. Finally, our dynamic specifications do not indicate any policy anticipation effects.

Overall, our findings are similar to recent estimates by Chodorow-Reich & Karabarbounis (2016) and Coglianese (2015), who use policy variation that is quite different from what we use in this paper. At the same time, our estimates and conclusions are quite different from those reached by HMM and HKMM, even though they also use a border county pair based strategy. As we show in our **Section 2.8.1**, the differences are in large part due to their use of (model-based) LAUS data, as well as auxiliary parametric assumptions used by authors of the two papers which we do not find to be warranted by the data.

The small macro employment effects of UI found in this paper are consistent with small negative effects on labor supply typically (though not always) found in the existing literature, together with moderately sized, positive effect on aggregate demand in the local economy. Future research should better disaggregate the macro effect into its constituent components: labor supply, demand multiplier, rat race and job creation effects. Nonetheless, our results suggest that the overall employment impact of the sizable UI extensions during the Great Recession was likely modest. At worst they led to a small reduction in aggregate employment, and at best they slightly boosted employment in the local economy.

Table 2.1: Pre-existing employment trends prior to November 2008 UI benefit expansion

	(1) All counties	(2) Border counties	(3) Border counties	(4) PTT-trimmed
Treatment X Date	-0.780***	-0.976***	-0.241	-0.110
	(0.244)	(0.206)	(0.286)	(0.110)
Observations	148896	111456	111456	83520
County fixed effects	X	X	X	X
Pair-period fixed effects			X	X

Notes: In columns 1 and 2, each cell reports the coefficient on $treat_c \times t$ from a regression of the following form: $E_{ct} = \alpha \times treat_c \times t + \lambda_c + \theta_t + \epsilon_{ct}$. In columns 3 and 4, each cell reports the coefficient on $treat_c \times t$ from a regression of the following form: $E_{cpt} = \alpha \times treat_c \times t + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. In all columns, the dependent variable is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression is estimated over the period 2004m11-2008m10 and t is the date divided by 48 (representing the 48 month period between the beginning and the end of this sample). The time-invariant variable $treat_c$ is the average treatment intensity for each county, defined as the average duration over the 2008m11-2013m12 period, minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. In column 1, standard errors are clustered at the state level. In columns 2, 3, and 4, standard errors are clustered two-way at the state and state-pair level. Columns 4 report the estimates from the set of border county pairs in the PTT-trimmed sample. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.2: Summary statistics: High-treatment versus low-treatment counties in border county pair sample

	Baseline					**PTT-Trimmed**				
	High: Mean	Sd	Low: Mean	Sd	p-val	High: Mean	Sd	Low: Mean	Sd	p-val
EPOP (A)	44.227	17.113	44.741	15.269	0.679	42.983	14.641	44.075	13.721	0.285
Private EPOP (A)	34.773	16.168	35.132	14.567	0.762	33.723	13.652	34.804	13.342	0.243
LAUS unemp. rate (A)	5.127	1.674	4.864	1.843	0.135	5.241	1.585	4.916	1.794	0.023
Population age 15+ (A)	79,283	207,157	69,625	148,849	0.214	91,211	231,886	78,193	153,538	0.088
Share white (B)	0.811	0.182	0.811	0.177	0.998	0.815	0.180	0.818	0.176	0.838
Share black (B)	0.085	0.145	0.086	0.147	0.966	0.087	0.146	0.085	0.144	0.862
Share hispanic (B)	0.067	0.111	0.059	0.092	0.491	0.061	0.102	0.054	0.087	0.455
Share H.S. grad (B)	0.569	0.064	0.567	0.065	0.724	0.568	0.063	0.566	0.066	0.726
Share college (B)	0.179	0.078	0.189	0.086	0.010	0.182	0.080	0.193	0.088	0.000
Median h.h. income (B)	42,645	11,459	43,535	12,127	0.198	42,997	11,881	44,145	12,728	0.073
New mortgage debt p.c. (A)	3.456	3.226	3.674	3.039	0.423	3.556	2.961	3.836	3.090	0.251
Share in cities 50k+ (C)	0.190	0.331	0.196	0.331	0.759	0.203	0.338	0.222	0.348	0.267
Min. weeks of UI elig.	24.470	3.495	24.631	3.199	0.718	24.478	3.495	24.720	3.092	0.609
Max. weeks of UI elig.	96.105	6.674	86.996	13.320	0.000	96.452	6.212	87.755	12.787	0.000
Pairs w/ different avg treatment	1131		1131			849		849		
Pairs w/ identical avg treatment	30		30			21		21		

Notes: The first four columns report summary statistics in border counties in the estimation sample, separately for "high" and "low" treatment counties. A county's assignment to the "high" or "low" group is defined by its average treatment intensity relative to its counterpart within each pair. Average treatment intensity ($treat_c$) is a time-invariant, continuous measure defined as the average duration over the 2008m11-2013m12 period, minus average duration over the 2007m11-2008m10 and 2014m1-2014m12 periods. The 30 (baseline) or 20 (PTT-trimmed) border county pairs with identical treatment are dropped in this table. The fifth column reports the p-values from a test that the means for high counties and low counties are equal, robust to clustering two-way at the state and state-pair level. Columns 6-10 report analogous statistics for the subsample of border county pairs in the PTT-trimmed sample. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008. If a border county appears in j county-pairs, then it appears j times for the purpose of creating the estimates in this table. (A) is from 2007 data, (B) is from the 2005-2009 ACS, and (C) is from the 2010 Census. High school graduates are those who have attained a high school degree but not a bachelor's degree. College graduates are those who have attained a bachelor's degree.

Table 2.3: Main Estimates: Effect of UI benefit duration on EPOP using OLS and IV specifications

	(1) BCP-FE	(2) PTT-Trimmed
Full sample		
OLS Estimate	0.430	0.213
	(0.466)	(0.270)
County pairs	1161	870
Observations	199692	149640
Pooled sample (IV)		
IV estimate	0.143	-0.069
	(0.964)	(0.635)
First stage coef.	0.847***	0.842***
	(0.052)	(0.051)
F stat.	[262.2]	[262.3]
County pairs	1161	870
Observations	108000	81120
2008 sample (IV)		
IV estimate	0.549	0.198
	(2.515)	(1.265)
First stage coef.	0.717***	0.726***
	(0.110)	(0.113)
F stat.	[41.3]	[40.3]
County pairs	1161	870
Observations	55728	41760
2014 sample (IV)		
IV estimate	-0.024	-0.182
	(0.562)	(0.521)
First stage coef.	0.915***	0.903***
	(0.046)	(0.043)
F stat.	[392.6]	[423.8]
County pairs	1089	820
Observations	52272	39360

Notes: Each panel reports two coefficients on D_{ct} from a regression of the form $E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt}$. E_{cpt} is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points and D_{ct} is the potential weeks of UI benefits divided by 73. The second column restricts the sample to the PTT sample. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008. Regressions in the first panel use OLS estimated over the 2007m11-2014m12 period. Regressions in the remainder of the table are estimated on subsamples using instrumental variables. The instrument z_{ct} is defined as follows. From 2007m11-2008m10, z_{ct} is equal to the duration available immediately prior to the implementation of UCEA; from 2008m11-2009m10, z_{ct} is equal to the duration available immediately *after* the implementation of UCEA. From 2013m1-2013m12, z_{ct} is equal to the duration available immediately prior to the expiration of EUC; from 2014m1-2014m12, z_{ct} is equal to the duration available immediately after EUC expiration, before any changes in regular benefits took effect. Estimates in the second panel pool the 2007m11-2009m10 and 2013m1-2014m12 samples and replace county fixed effects with county-by-subsample fixed effects. Estimates in the third panel use data from 2007m11-2009m10; estimates in the fourth panel use data from 2013m1-2014m12. In the IV specifications, first stage coefficients and standard errors are also reported. Standard errors are reported in parentheses and first stage F-statistics in square brackets. Standard errors are clustered two-way at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.4: Robustness of the effects of UI benefit duration on EPOP: choice of sample period

	(1) BCP-FE	(2) PTT-Trimmed
2007m11-2014m12	0.430	0.213
	(0.466)	(0.270)
	N = 199692	N = 149640
2006m11-2014m12	0.142	0.175
	(0.451)	(0.322)
	N = 227556	N = 170520
2005m11-2014m12	-0.088	0.138
	(0.440)	(0.356)
	N = 255420	N = 191400
2004m11-2014m12	-0.330	0.064
	(0.452)	(0.378)
	N = 283284	N = 212280
County pairs	1161	870

Notes: Each cell reports the coefficient on D_{ct} from a regression of the form $E_{cpt} = \beta D_{ct} + \lambda_c + \nu_{pt} + \eta_{cpt}$. E_{cpt} is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points and D_{ct} is the potential weeks of UI benefits divided by 73. The second column restricts the sample to the PTT sample. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008. The regression in each row is estimated over the sample-period indicated. The estimates in row 1 correspond to the estimates in the top panel of Table 2.3. Standard errors are clustered two-way at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.5: Robustness of the effects of UI benefit duration on EPOP: choice of cutoffs for trimming on match quality

	(1) Full	(2) Pooled	(3) 2008	(4) 2014
Baseline	0.430	0.143	0.549	-0.024
	(0.466)	(0.964)	(2.515)	(0.562)
	N = 199692	N = 108000	N = 55728	N = 52272
10th percentile	0.161	0.199	0.558	0.049
	(0.304)	(0.599)	(1.274)	(0.456)
	N = 179568	N = 97152	N = 50112	N = 47040
20th percentile	0.170	0.042	0.314	-0.074
	(0.276)	(0.612)	(1.241)	(0.507)
	N = 159616	N = 86544	N = 44544	N = 42000
25th percentile	0.213	-0.069	0.198	-0.182
	(0.270)	(0.635)	(1.265)	(0.521)
	N = 149640	N = 81120	N = 41760	N = 39360
30th percentile	0.221	-0.109	0.184	-0.232
	(0.272)	(0.629)	(1.216)	(0.549)
	N = 139664	N = 75648	N = 38976	N = 36672
40th percentile	0.329	0.085	0.929	-0.257
	(0.286)	(0.660)	(1.305)	(0.558)
	N = 119712	N = 64800	N = 33408	N = 31392
50th percentile	0.340	0.048	1.007	-0.328
	(0.302)	(0.719)	(1.435)	(0.601)
	N = 99760	N = 53952	N = 27840	N = 26112

Notes: Each cell reports the baseline coefficient from the full sample, pooled event sample, and 2008 and 2014 subsamples, estimated over a different subsample of border county pairs. The cells in row 1 correspond to the estimates in column 1 of Table 2.3. In the other rows, the sample of border county pairs (BCPs) is trimmed based on the magnitude of differences in pre-existing trends estimates from 2004m11-2008m10. We rank and then trim all BCPs according to the magnitude of differences in pre-treatment trends (PTT). In the second row, we drop the bottom 10 percent of BCPs with the largest differences in pre-existing trends, in the third row, we drop the bottom 20 percent, and so forth. The fourth row (the 25th percentile) corresponds to the estimates in column 2 of Table 2.3. Standard errors are clustered two-way at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.6: Additional robustness checks on the effects of UI benefit duration on EPOP

	Full sample OLS		**Pooled sample IV**	
	(1) BCP-FE	(2) PTT-Trimmed	(3) BCP-FE	(4) PTT-Trimmed
1. Baseline	-0.709	-0.913	-2.308	-3.093
	(0.645)	(0.780)	(1.813)	(2.256)
2. Private EPOP	-0.549	-0.754	-2.271	-3.117
	(0.601)	(0.769)	(1.951)	(2.481)
3. Correlation-trimmed	-0.931	-0.944	-3.631**	-4.602***
	(0.771)	(0.903)	(1.500)	(1.682)
4. ISLT	0.027	-0.451	-1.857*	-1.537
	(0.603)	(0.727)	(1.088)	(1.281)
5. PTT through 2007m10		-0.512		-2.327
		(0.790)		(2.309)
6. Eliminate lapse	-0.578	-0.872		
	(0.670)	(0.808)		
7. Quarterly data	-0.757	-0.984	-2.367	-3.285
	(0.702)	(0.849)	(1.756)	(2.196)
8. QWI EPOP (quarterly)	-0.771	-1.415*	-2.456	-4.123*
	(0.652)	(0.779)	(1.733)	(2.121)
9. Not seasonally adjusted	-0.582	-0.761		
	(0.638)	(0.775)		
10. Distance trimming	-0.769	-1.043	-2.048	-3.341
	(0.690)	(0.805)	(1.903)	(2.381)
11. Unbalanced panel	-0.709	-0.913	-2.308	-3.093
	(0.645)	(0.780)	(1.813)	(2.256)
12. Hinterland pairs	1.787*	0.701	-0.949	0.475
	(0.943)	(0.660)	(1.726)	(1.344)
13. Exploit Δ reg. benefits			-2.391	-3.217
			(1.781)	(2.205)
14. Drop NC	-0.699	-0.982	-2.308	-3.093
	(0.721)	(0.833)	(1.813)	(2.256)
15. Keep NC			-3.073	-4.445*
			(2.003)	(2.700)
16. NC: Alt. instrument			-1.441	-2.529
			(1.327)	(1.828)
17. $\ln(EPOP)$	-0.036**	-0.035*	-0.034	-0.047
	(0.016)	(0.021)	(0.038)	(0.048)
	[-1.847**]	[-1.832*]	[-1.810]	[-2.434]
18. $\ln(emp)$	-0.052**	-0.057**	-0.026	-0.042
	(0.020)	(0.027)	(0.042)	(0.054)
	[-2.684**]	[-2.921**]	[-1.383]	[-2.202]

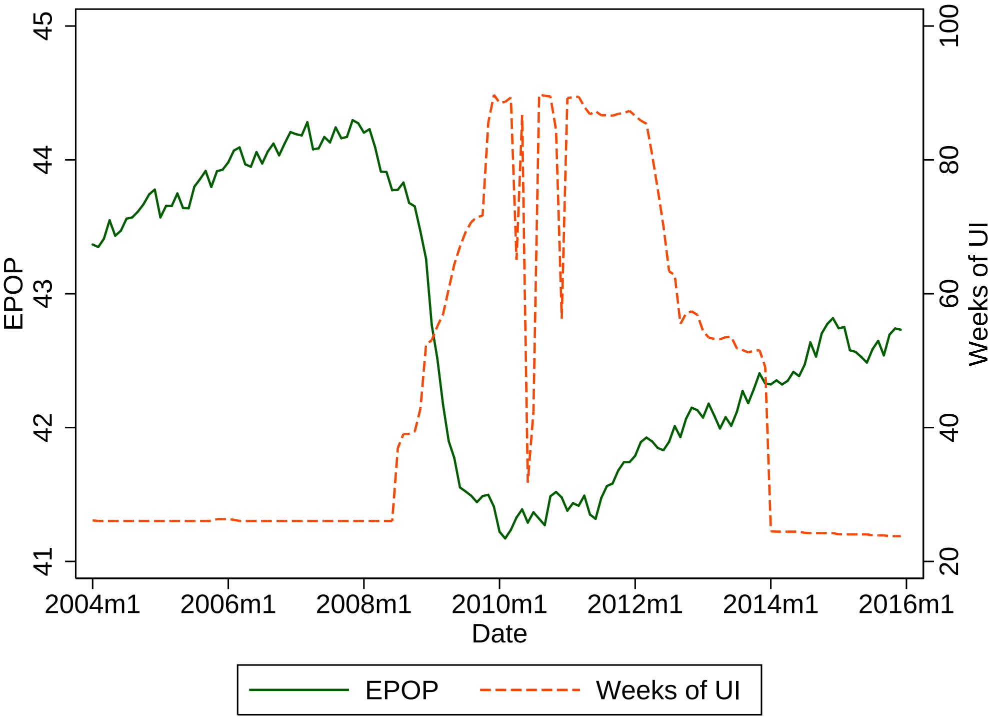
Notes: Each cell reports regressions analogous to those reported in Table 2.3 for the full sample with OLS or the pooled event samples (IV). The estimates in the 1st row correspond to the estimates in the top two panels of Table 2.3. The estimates in the 2nd row replace (total) EPOP with the ratio of private employment to population age 15+. In the 3rd row, we trim the set of border county pairs based on the level of correlation between county EPOP and state EPOP over the period 2004m11-2008m10 (see text for details). The 4th row controls for county-specific linear trends. The 5th row trims based on PTT estimated through 2007m10 instead of 2008m10. The 6th row recodes the periods in 2010 when EUC lapsed by assigning EUC values during these lapses as equal to their prior value. The 7th row uses quarterly data instead of monthly (and estimates over the 2007q4-2014q4 period). The 8th row uses EPOP derived from the QWI (at the quarterly level) instead of the QCEW. The 9th row uses seasonally-unadjusted data. The 10th row drops county-pairs whose population centroids are greater than 100km apart. The 11th row includes counties without full EPOP data for each month, which we drop by default. The 12th row uses a modified version of the instrument z_{ct} which exploits all changes in benefits, including changes in regular benefits, which occur at the end of December 2013. Rows 13-15 report estimates using alternative strategies for dealing with North Carolina (NC); by default, border county pairs (BCPs) with one neighbor in NC are kept in the full sample OLS and the 2008 subsample and dropped in the 2014 subsample. The 13th row completely drops all NC BCPs. The 14th row keeps all North Carolina BCPs. The 15th row keeps NC BCPs but redefines the instrument for NC counties (see text for details). The 16th and 17th row use $\ln(EPOP)$ and $\ln(employment)$, respectively, as dependent variables. The bracketed estimates in these two rows are the level-on-level equivalent, equal to $(\frac{99}{26}^{\hat{\beta}} - 1)\bar{E}$, where $\bar{E}$ is the mean EPOP level in the given sample. Cells which are not applicable in the given sample, or which provide estimates that are mechanically equal to the baseline estimates, are left blank. Standard errors are clustered two-way at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.7: Rationalizing micro and macro employment effects of UI: demand side effects and implied fiscal multipliers

	β_{MICRO}	β_{MACRO}	ΔE_{MICRO}	ΔE_{MACRO}	m_f
Rothstein (2011), lower bound	0.1	-0.1	-0.1	-0.3	-0.2
	0.1	0.2	-0.1	0.5	0.9
Farber and Valletta (2015)	0.4	-0.1	-0.5	-0.3	0.4
	0.4	0.2	-0.5	0.5	1.5
Rothstein (2011), upper bound	0.5	-0.1	-0.7	-0.3	0.6
	0.5	0.2	-0.7	0.5	1.7
Daly et al. (2012)	0.8	-0.1	-1.1	-0.3	1.2
	0.8	0.2	-1.1	0.5	2.3
Elsby et al. (2010), upper bound	2.4	-0.1	-3.2	-0.3	4.2
	2.4	0.2	-3.2	0.5	5.3
Johnston and Mas (2016)	4.6	-0.1	-6.2	-0.3	8.4
	4.6	0.2	-6.2	0.5	9.5

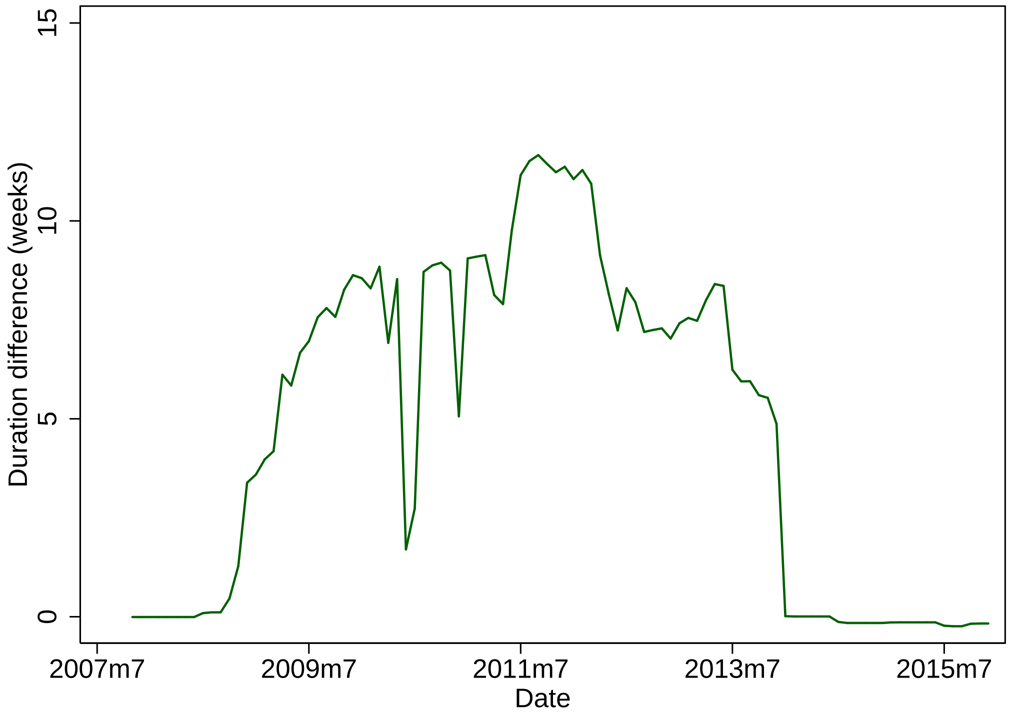
Notes: The table displays estimates of implied fiscal multipliers, using a range of micro estimates from other studies (Column 1) and two macro estimates from this paper (Column 2). β_{MICRO} is an estimate of the change in the unemployment rate resulting from only the micro-level effect of a 73-week increase in maximum UI duration, while β_{MACRO} is a direct estimate of the aggregate change in EPOP. Columns (3) and (4) represent the resulting impact on employment (in millions of workers), and are calculated as $\Delta E_{\text{MICRO}} = \beta_{\text{MICRO}} \times L$ and $\Delta E_{\text{MACRO}} = \beta_{\text{MACRO}} \times P$, where P is the population and L is the labor force, expressed in millions. m_f is the implied fiscal multiplier, computed under the assumption that the entirety of the gap between the macro and micro employment effects is due to aggregate demand. m_f is computed according to the equation $m_f = \frac{Y}{E} \times \frac{\Delta E_{MACRO} - \Delta E_{MICRO}}{100} \times \frac{1}{\Delta B}$, where $\frac{Y}{E}$ is output per worker in 2012, and ΔB is annual EB and EUC expenditure, averaged over the period from November 2008 through December 2013 and scaled to correspond to a 73-week increase in duration.

Figure 2.1: Evolution over time: national QCEW-based EPOP ratio and UI benefit duration



Notes: EPOP is the seasonally-adjusted ratio of employment (from the QCEW) to population age 15+. Weeks of UI represents the maximum number of weeks of UI compensation available. In this figure, both EPOP and weeks of benefits are calculated via an unweighted average of counties.

Figure 2.2: Difference in UI benefit duration between high-treatment and low-treatment counties across state borders



Notes: For each county pair, we compute the difference between maximum duration in the high-duration county and in the low-duration county. We plot the average difference across all county pairs. "High" and "low" status is determined by comparing the difference between average duration from 2008m11-2013m12 and average duration from 2007m11-2008m10 and 2014m1-2014m12. The counties in the 30 pairs where this difference is identical are assigned arbitrarily to the "high" and "low" sets.

Figure 2.3: Reduction in UI benefit duration from the December 2013 expiration of EUC

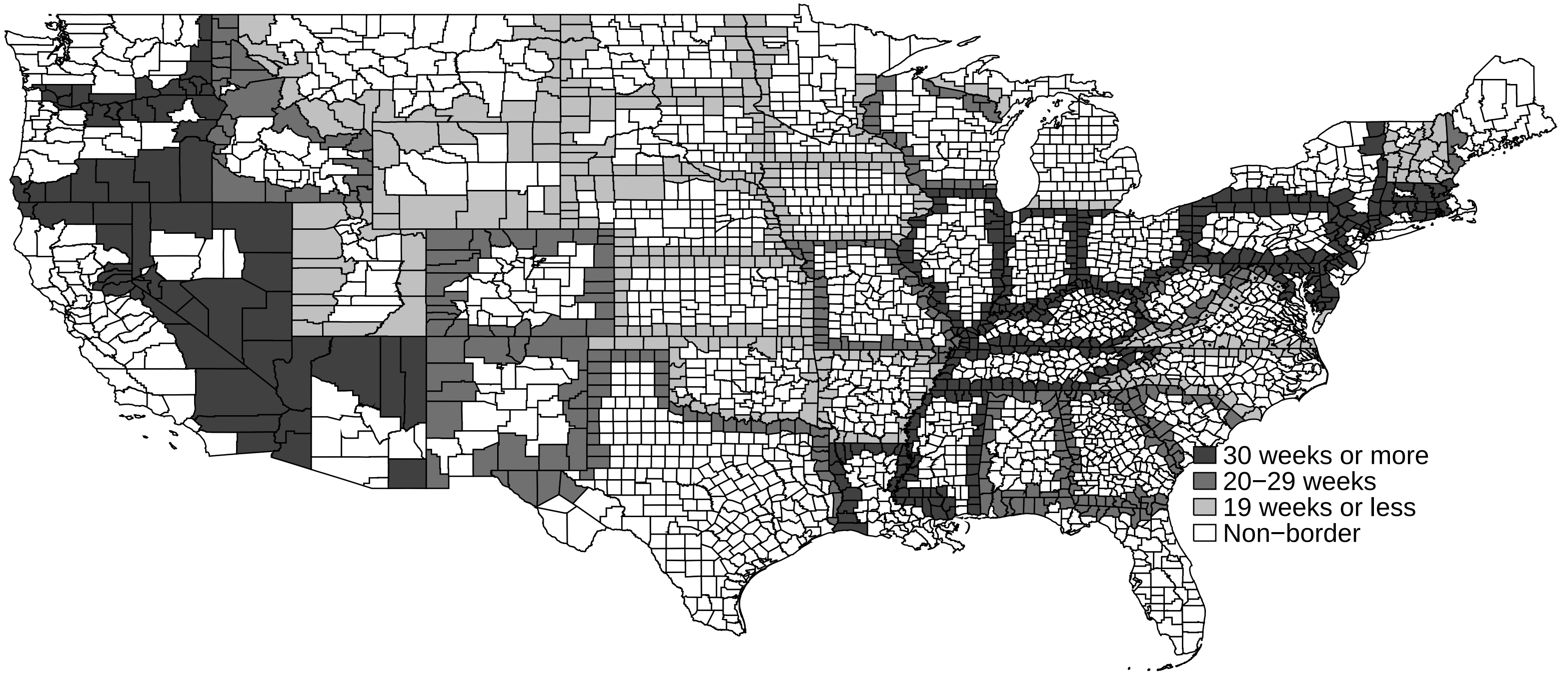
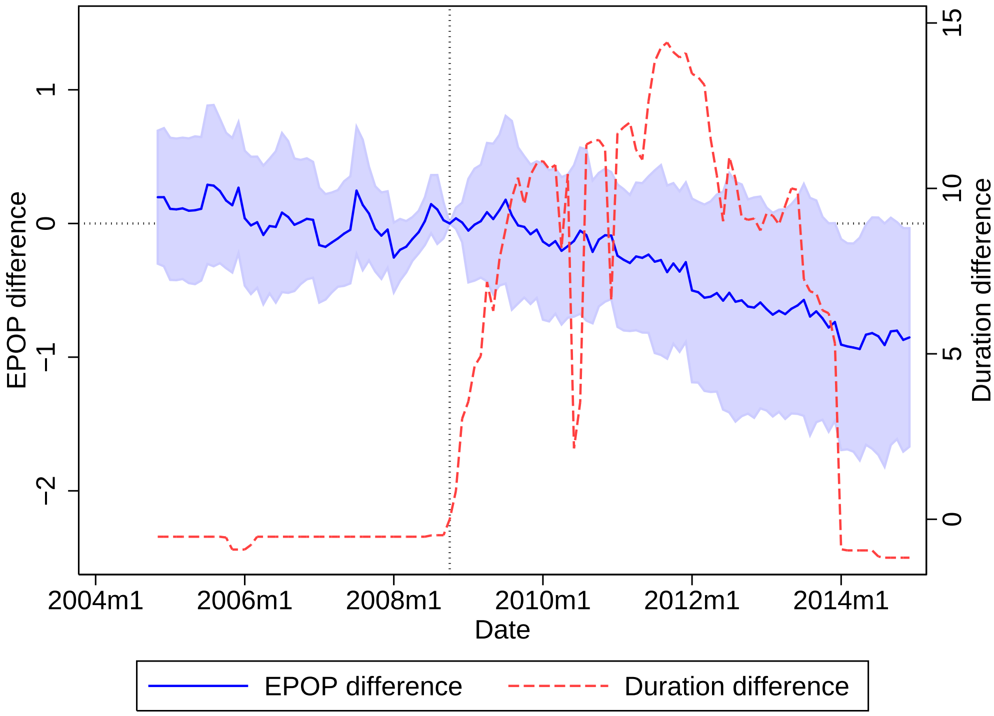
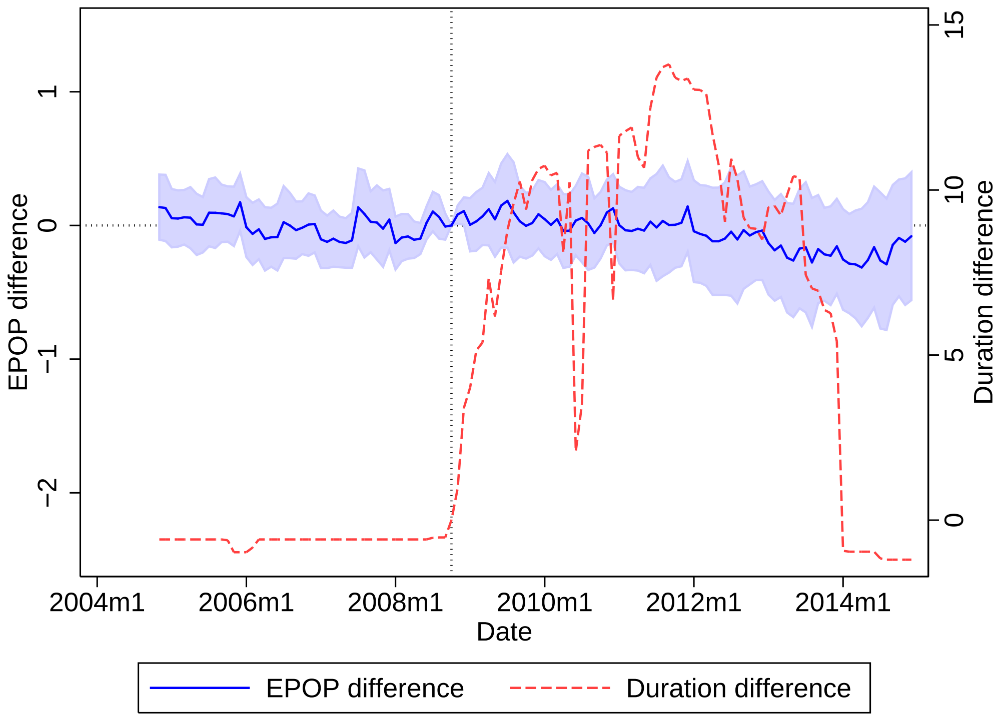


Figure 2.4: Evolution of EPOP and UI benefit duration differentials by average treatment intensity: baseline border county pair sample



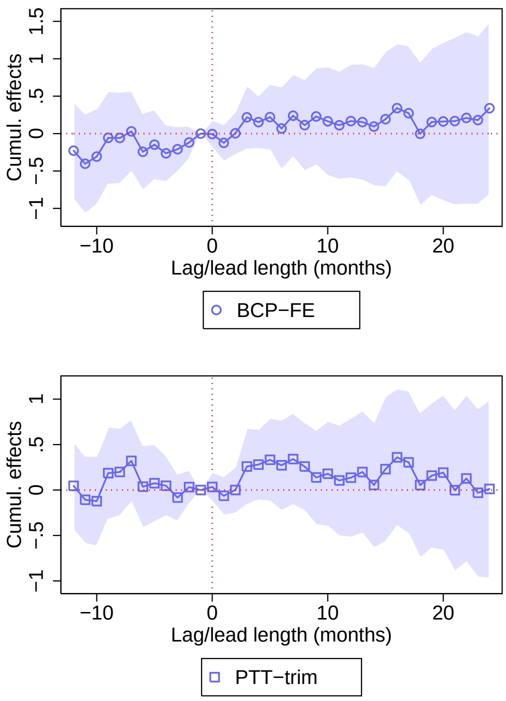
Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. E_{cpt} is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points. The average treatment intensity, $treat_c$, is a time-invariant, continuous measure defined as the average duration during the treatment period (2008m11-2013m12), minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. The shaded region corresponds to the 95% confidence interval, robust to two-way clustering at the state and state-pair level. The dotted line (right axis) reflects the analogous coefficients with D_{ct} as the dependent variable, where D_{ct} is weeks of benefits. The month 2008m10, the last month prior to the first introduction of differential EUC, is marked with a dotted vertical line. The sample includes 1,161 county pairs.

Figure 2.5: Evolution of EPOP and UI benefit duration differentials by average treatment intensity: PTT-trimmed border county pair sample



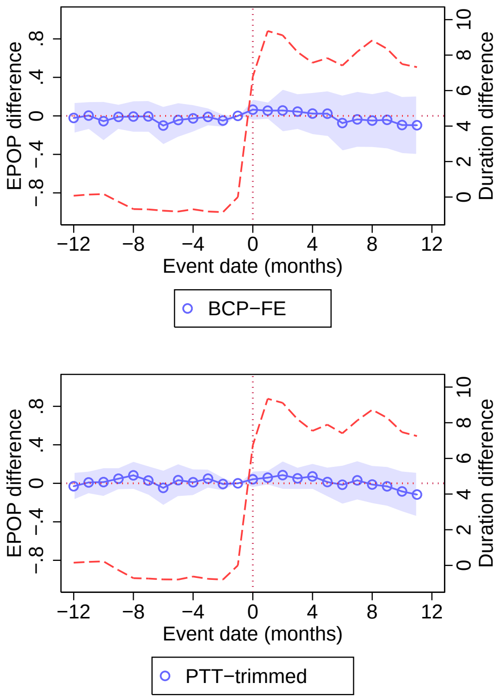
Notes: This figure plots (solid line, left axis) the set of β_s coefficients from the following regression estimated over the set of border county pairs in the PTT-trimmed sample: $E_{cpt} = \sum_{s=\tau_A}^{\tau_B} \beta_s treat_c \mathbb{1}\{t = s\} + \lambda_c + \nu_{pt} + \epsilon_{cpt}$. E_{cpt} is the seasonally-adjusted ratio of total employment to population age 15+, scaled in percentage points. The average treatment intensity, $treat_c$, is a time-invariant, continuous measure defined as the average duration during the treatment period (2008m11-2013m12), minus average duration from the 12 months prior (2007m11-2008m10), divided by 10. The shaded region corresponds to the 95% confidence interval, robust to two-way clustering at the state and state-pair level. The dotted line (right axis) reflects the analogous coefficients with D_{ct} as the dependent variable, where D_{ct} is weeks of benefits. The month 2008m10, the last month prior to the first introduction of differential EUC, is marked with a dotted vertical line. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008. The sample includes 870 county pairs.

Figure 2.6: Cumulative response of EPOP from distributed lags specification: OLS in first-differences



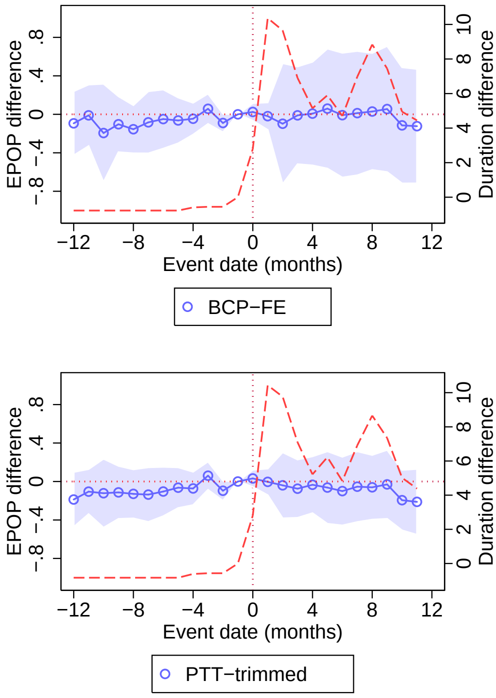
Notes: This figure reports the monthly cumulative response of EPOP from a 73 week increase in maximum UI benefit duration, centered around event date -1 whose cumulative response is defined as zero. The model is estimated on the full sample (2007m11-2014m12), using all border county pairs (BCPs) (hollow circles) and the subset of BCPs in the PTT-trimmed sample (hollow squares), where all independent variables are divided by 73. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 24 lags and 11 leads in first-differenced benefit duration, and is estimated using EPOP data from 2007m11-2014m12 (and thus duration data from 2005m11-2015m11). Lags are to the right of zero; leads are to the left of zero. The zeroth cumulative response is equal to the estimated coefficient on contemporaneous benefit duration. The j^{th} cumulative lag is equal to the estimated coefficient on contemporaneous duration plus the sum of the estimated coefficient on the 1st through jth lag term. The j^{th} cumulative lead is equal to -1 times the sum of the estimated coefficients on the first through the $j-1^{th}$ lead terms. The shaded region corresponds to the 95% confidence interval, robust to two-way clustering at the state and state-pair level.

Figure 2.7: Evolution of EPOP difference and UI benefit duration difference across state borders: Pooled 2008 expansion and 2014 expiration of EUC



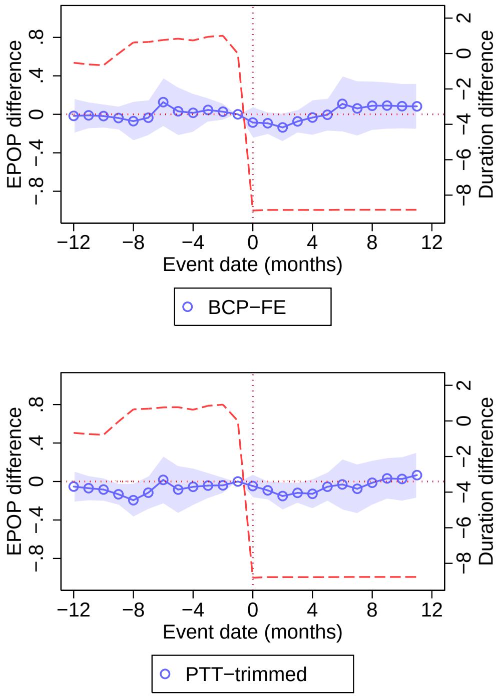
Notes: This figure reports the monthly cumulative response of EPOP (left axis, hollow circles) from the pooled 2008 and 2014 samples, centered around event date -1 whose cumulative response is defined as zero. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 11 lags and 12 leads in first-differenced benefit duration: for the 2008 sample, the duration variable is equal to the increase in weeks of UI duration immediately upon the implementation of UCEA, divided by 10; for the 2014 sample, the duration variable is defined as -1 times the weeks of UI duration lost as a result of EUC expiration, divided by 10. The dashed line (right axis) reports the monthly cumulative response of benefit duration around the event; the regression is identical to the EPOP specification except that the dependent variable is the first-differenced benefit duration in weeks. The upper panel reports the results from the baseline BCP-FE sample consisting of all border county pairs; the lower panel reports results using the PTT-trimmed sample, which drops the quartile of county pairs with the highest differential in pre-treatment linear trends between November 2004 and October 2008. Event date zero is marked with a dotted vertical line; this corresponds to November 2008 for the 2008 sample and January 2014 for the 2014 sample.

Figure 2.8: Evolution of EPOP difference and UI benefit duration difference across state borders: 2008 expansion of EUC



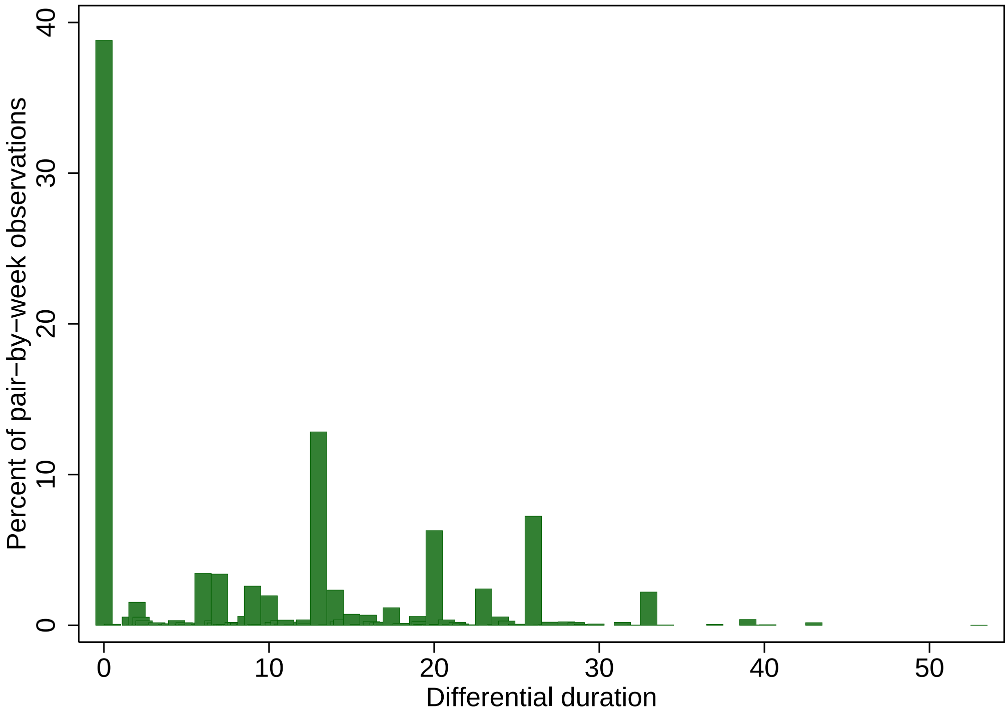
Notes: This figure reports the monthly cumulative response of EPOP (left axis, hollow circles) from the November 2008 EUC expansion, centered around event date -1 whose cumulative response is defined as zero. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 11 lags and 12 leads in first-differenced benefit duration, where this duration variable is equal to the increase in weeks of UI duration immediately upon the implementation of UCEA, divided by 10. The dashed line (right axis) reports the monthly cumulative response of benefit duration around the event; the regression is identical to the EPOP specification except that the dependent variable is the first-differenced benefit duration in weeks. The upper panel reports the results from the baseline BCP-FE sample consisting of all border county pairs; the lower panel reports results using the PTT-trimmed sample, which drops the quartile of county pairs with the highest differential in pre-treatment linear trends between November 2004 and October 2008. Event date zero is marked with a dotted vertical line, and corresponds to November 2008.

Figure 2.9: Evolution of EPOP difference and UI benefit duration difference across state borders: 2014 expiration of EUC



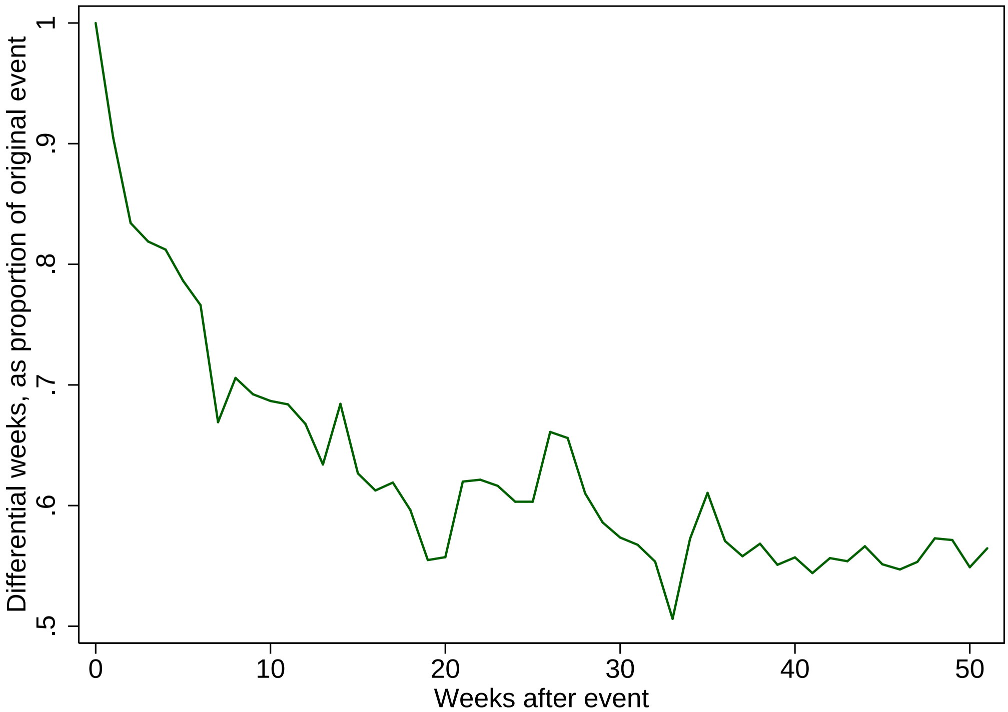
Notes: This figure reports the monthly cumulative response of EPOP (left axis, hollow circles) from the EUC expiration at the end of 2013, centered around event date -1 whose cumulative response is defined as zero. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 11 lags and 12 leads in first-differenced benefit duration, where this duration variable is defined as change in weeks available as an immediate result of EUC expiration, divided by 10. The dashed line (right axis) reports the monthly cumulative response of benefit duration around the event; the regression is identical to the EPOP specification except that the dependent variable is the first-differenced benefit duration in weeks. The upper panel reports the results from the baseline BCP-FE sample consisting of all border county pairs; the lower panel reports results using the PTT-trimmed sample, which drops the quartile of county pairs with the highest differential in pre-treatment linear trends between November 2004 and October 2008. Event date zero is marked with a dotted vertical line, and corresponds to January 2014.

Figure 2.10: Distribution of differences in UI benefit duration across border county pairs



Notes: This figure plots the distribution of duration differences across border county pairs, with each observation at the pair-by-(calendar)-week level. The sample is restricted to weeks between November 23, 2008, and December 22, 2013.

Figure 2.11: Persistence of differential change in UI benefit duration across border county pairs



Notes: This figure plots the persistence of all changes in relative duration in the full sample. In particular, the data is organized at the pair (p), event (s), event-week (τ) level, where an event is any change in the duration difference across a county pair. The dependent variable $y_{ps\tau}$ is the difference in duration across the county pair, minus that same difference immediately prior to the event. This dependent variable is regressed on the size of the initial event interacted with 52 dummies for the 52 event-weeks τ immediately following the event. This figure plots those coefficients. See text for details.

2.8 Appendix

2.8.1 Comparison with HKMM and HMM

The results in this paper are quite different than the results in Hagedorn, Karahan, Manovskii, & Mitman (2016) (which studies the effect of UI from 2005 to 2012) and the results in Hagedorn, Manovskii, & Mitman (2016) (which studies the effect of EUC expiration at the end of 2013). Similar to this paper, both HKMM and HMM use border county pairs for their estimation. However, there are differences in data, in econometric specification, and in sample definitions between our paper and these two studies. Some differences are minor, while others are quite important. In this section, we offer a reconciliation of these sets of results, and assess which of the differences in sample definition, data and specifications drive the differences in the estimates and conclusions reached by our respective papers.

In this section, we compare our OLS estimates from the baseline BCP sample to the baseline estimates of HKMM. The HKMM estimation equation is as follows, where data for a given pair p at time t has already been spatially differenced (after taking logs):

$$\ln(u_{pt}) - \beta(1 - s_t)\ln(u_{pt+1}) = \alpha * \ln(D_{pt}) + \lambda_p' F_t + \epsilon_{pt} \tag{2.10}$$

Here, u_{pt} is the unemployment rate from LAUS,[45] β is the discount factor (equal to 0.99), s_t is the separation rate, D_{pt} is the same measure of maximum benefit lengths that we use, and $\lambda_p' F_t$ are interactive effects. Thus, the dependent variable is a quasi-forward difference (QFD) of the log of the un-

[45]The LAUS data used by HKMM has been substantially revised since they accessed it. We have estimated the models using both the pre-revision version of the LAUS data used by HKMM and the more recent, revised version of the data. We have found both versions of the data give similar results in the HKMM specifications. We use pre-revision data throughout the discussion of HKMM.

employment rate. They then calculate the total effect of UI on unemployment by considering the steady state ($u_{pt} = u_{pt+1}$) impact of a persistent increase in D_{pt}. In the steady state, $\ln(u_p) = \frac{\alpha}{1-\beta(1-s)} \ln(D_p)$. Therefore, HKMM's headline claim comes from multiplying their main estimate by a factor $\frac{1}{1-\beta(1-s)}$, which is approximately equal to 10. They perform their estimation over the period 2005q1-2012q4.

Our full-sample BCP-FE estimation strategy is different from HKMM in five distinct ways. These differences are: (1.) we do not transform our dependent variable using quasi-forward-differencing, (2.) we use employment data from the QCEW rather than unemployment data from LAUS, (3.) we estimate the results using monthly data from 2007m11-2014m12, instead of quarterly data from 2005q1-2012q4, (4.) we control for differences across county pairs using a fixed effects model rather than the Bai (2009) interactive fixed effects model, and (5.) we use levels instead of logs.

Table 2.8 describes the impact of each of these five steps. Because different specifications have different dependent variables, and because the implied effect is not equal to the coefficient in some specifications, we standardize each specification into an implied effect of the 26-to-99 week expansion on EPOP.[46] We "translate" between implied effects on the unemployment rate and implied effects on EPOP by using the total peak-to-trough impact of the Great Recession. We measure this peak-to-trough impact using the unweighted average of counties in our border-pair sample. In particular, in this sample, EPOP fell from 44.3% to 41.2% and the unemployment rate increased from 4.8% to 9.7%. So, if one estimation suggests that the impact of the 26-to-99 week expansion was 3 percentage points of unemployment, we would convert that specification's estimate into an EPOP effect of $3 \times (\frac{41.2-44.3}{9.7-4.8}) \approx -1.9$ percentage

[46]Importantly, we scale up the estimates in QFD specifications by $\frac{1}{[1-\beta(1-s)]}$, as HKMM do.

points.

Table 2.8 analyzes one-off changes either starting from the HKMM specification (column 1), or moving to our specification (column 2). The first row begins with reporting the estimates: our replication of the HKMM estimates suggest that the UI benefit expansion from 26 to 99 weeks has an implied EPOP effect of -2.72, which is nearly 90% of the decrease in EPOP during the Great Recession within our sample. This corresponds to a coefficient estimate of 0.052, while HKMM report a very similar estimate of 0.049. We find that this estimate is statistically significant, as HKMM do. In contrast, the point estimates for the full sample BCP-FE estimates in this paper suggest that the decline in EPOP would have been about 10% *greater* without the UI expansions, though this is not distinguishable from zero.

The next five rows report the marginal impact of each of the five steps. In column 1, we show what happens when the step reported in the row is added starting with the HKMM specification. In column 2, we show what happens when this step is added to our specification. Finally, in column 3, we consider *all* possible transition paths between HKMM's estimates and our estimates, and report the average marginal contribution of each of the steps, across all of these transition paths.[47]

The key findings are as follows. Quasi-forward differencing, the use of the LAUS unemployment data as opposed to the QCEW employment data,

[47]We do not consider the step of switching from logs to levels in column 1, because the quasi-forward-differencing is motivated by theory which requires the data to be in logs. With quasi-forward-differenced data in levels, it is neither clear what we are measuring, nor what the total effect of UI on employment would be. For the same reason, we do not consider adding quasi-forward-differencing to our specification in column 2 (which is in levels). In addition, when calculating the averages in column 3, we discard transition paths that involve using quasi-forward-differenced data in levels. In the end, we estimate 24 models with all allowable combinations of the five sources of differences; we then take 60 paths (equal to 5! paths with 1/2 thrown out because eliminating quasi-forward differencing happens after the logs to levels conversion) between the HKMM and BDGK estimates, and calculate the contribution of each of these five factors averaged across these 60 paths.

and sample alignment are all consequential choices. In contrast, the use of interactive fixed effects as opposed to linear fixed effects and the use of logs versus levels are not consequential choices.

Column 1 shows that, starting from the HKMM estimate, switching from the LAUS unemployment rate, or getting rid of quasi-forward differencing, dramatically reduces the HKMM estimates in magnitude towards zero. In particular, just switching from the LAUS unemployment rate to the QCEW EPOP (as shown in Row 4) changes the estimates to $-2.724 + 1.356 = -1.368$, suggesting the UI benefit expansion explained around 40% of the fall in EPOP rather than 90% as implied by HKMM's estimates. Similarly, removing quasi-forward differencing (Row 2) changes the estimates to $-2.724 + 2.688 = -0.036$ percentage points of EPOP. Column 2 shows that use of the LAUS unemployment rate also leads to a (mistaken) suggestion of job loss when we start from our specification, although the impact of this is more modest. Starting from our BCP-FE specification, when we use the LAUS unemployment rate as the outcome, the translated result suggests the UI benefit expansion led to a change in EPOP equal to $.430 - 1.133 = -0.703$, just under a quarter of the overall change during the Great Recession. When we average the incremental contribution of these two steps across all permissible paths going between the HKMM specification and ours (in column 3), we find that dropping quasi-forward differencing increases the estimates by around 1.32 percentage points of EPOP (about 40% of the change in unemployment rate during the Great Recession), while switching the outcome from LAUS unemployment rate to QCEW based EPOP increases the estimate by about 0.74 percentage points of EPOP.

Aligning our samples also has a meaningful impact. The HKMM sample of 2005q1-2012q4 starts and ends earlier than our sample of 2007m11-2014m12.

As we showed in **Table 2.4**, while the baseline BCP-FE approach greatly reduces the pre-existing trend, it does not completely remove it. Use of an earlier start date, as well as an end date prior to the phase-out of differential UI benefits across state borders, can produce a more negative estimate in the presence of such trends. We find that use of this altered sample period leads to somewhat smaller magnitudes of estimates, reducing the impact of the policy by around 0.846, 1.461, and 0.863 percentage points of EPOP in columns 1, 2, and 3, respectively.

In contrast, the use of Bai (2009) interactive effects versus fixed effects, and use of logs versus levels, make fairly small contributions in explaining the difference between our two sets of estimates.

This analysis shows that (1) changing the sample period (and frequency) from HKMM's specification to ours, (2) eliminating quasi-forward-differencing, and (3) changing the dependent variable from the LAUS unemployment rate to QCEW EPOP all reduce the implied negative impact of UI on employment, by 0.74 to 1.25 percentage points of EPOP when averaged over all possible paths. We next discuss our justification for making the specification choices that we do.

Quasi-Forward Differencing

HKMM derive **Equation (2.10)** by considering a search-and-matching framework where the rate of vacancy posting or firm job creation depends on a firm's expectation about future wages. Since unemployment insurance puts upward pressure on wages, an increase in benefits would reduce the expected profits of the firm and lead to a reduction in job creation. Because expectations about *future* benefit changes can affect employment *today*, HKMM make the point that an empirical approach that only relates current employment to current or past policy changes would be misspecified. In order to capture these anticipation effects, HKMM use a quasi-forward-differencing procedure. Their argument is as follows: the value of an employee to an employer is equal to the current-period flow profits, plus $\beta(1-s)$ times the expected value of the employee tomorrow (since the value of a vacant job is driven to zero by free entry). Therefore, HKMM argue, we can isolate the impact of UI on current-period flow profits by considering the quasi-forward difference of the unemployment rate (which they consider to be proportional to current period flow profits, in logs). The theory predicts that, in the case of an increase in generosity that was a surprise and immediately known to be persistent, firms would move from a low-unemployment steady state to a high-unemployment steady state, according to the equation $\Delta \ln(u_p) = \frac{\alpha}{1-\beta(1-s)} \Delta \ln(D_p)$.[48] As we noted above, this choice is quite important—removing forward differencing essentially erases the entirety of their effect even in their sample.

We are generally less favorable toward the use of quasi-forward differencing for several reasons. This model-driven approach relies on strong parametric assumptions—most notably that labor demand is well-characterized by the vacancy-posting problem captured in the model. Unfortunately this results in an empirical approach that is very sensitive to misspecification. For example, if an increase in UI generosity (D_{pt}) tends to be associated with a decrease in future unemployment (u_{pt+1}) in the data, then the estimated coefficient α will be positive. However, such a pattern could also be consistent with a Keynesian aggregate demand effect that operates with a small delay. That is, if an increase in benefits in one period leads to increased aggregate demand and lower unemployment in the next period, the HKMM strategy would find that UI *increased* the unemployment rate, when in fact the opposite occurred. Second,

[48] Here α is the regression coefficient, β is the discount factor, s is the probability that the job ends, u is the unemployment rate, and D is the number of weeks of UI benefits.

as a practical matter, the size of the final estimate is sensitive to assumptions in the model required for translating a flow result to a steady state effect, and in the exact magnitudes of separation and discount rates. Both the heavy dependence on a specific model and the inability to distinguish between alternative explanations make quasi-forward differencing an unattractive strategy from our perspective.

Instead, our preferred strategy is to capture the dynamics in a less model-driven and a more transparent manner using distributed lags. That specification directly estimates employment changes around benefit duration innovations, allowing us to assess possible pre-existing trends, anticipatory effects, and delayed or slow moving response within the window. As we discussed in Section 2.5.2, we find no evidence of significant anticipation effects in the 12 months prior to benefit changes. The lack of any anticipation effect raises questions about the value of quasi-forward differencing the outcome, especially given the drawbacks discussed above.

LAUS versus QCEW

HKMM predominantly use the LAUS employment data rather than the QCEW employment data to compute county level measures of employment.[49] However, the LAUS data is partly model-based. In particular, while the LAUS data uses actual movement to unemployment based upon UI claims, they do not observe those entering (or re-entering) the labor force. Therefore, the county level estimates for unemployment are based on state-level data on labor force entry and re-entry—something BLS states explicitly in their online

[49]They do report results using the log employment from the QCEW and QWI as a robustness check, in columns 3 and 4 of Table 5. The log employment result, -0.03, would imply that the 26-99 week expansion of UI caused a reduction of employment by 3.9%, which would translate to about 1.6 percentage points of EPOP. This is about 40% less than implied EPOP effect of HKMM's main result, consistent with the average marginal effects reported in **Table 2.8**. The log employment results from the QWI are modestly larger.

manual (`http://www.bls.gov/lau/laumthd.htm`):

> “The second category, "new entrants and reentrants into the labor force," cannot be estimated directly from UI statistics, because unemployment for these persons is not immediately preceded by the period of employment required to receive UI benefits. In addition, there is no uniform source of new entrants and reentrants data for States available at the LMA [labor market area] level; the only existing source available is from the CPS at the State level. Separate estimates for new entrants and for reentrants are derived from econometric models based on current and historical state entrants data from the CPS. These model estimates are then allocated to all Labor Market Areas (LMAs) based on the age population distribution of each LMA. For new entrants, the area’s proportion of 16-19 years population group to the State total of 16-19 years old population is used, and for reentrants, the handbook area’s proportion of 20 years and older population to the State total of 20 years and older population is used.”

The use of state-level information in estimating county-level unemployment rates is problematic for a border discontinuity design. The border county design attempts to purge reverse causation present at the state level by using more local comparisons. Use of state-level information raises the possibility of finding a (spurious) discontinuity in the measured unemployment rate across the state borders even when there is no such discontinuity in reality.

The QCEW data are based on administrative payroll records provided to the BLS by states, which protects against finding spurious discontinuities. Moreover, the QCEW data includes around 98% of all formal sector workers, making them very close to the true total employment counts in these counties.

For these reasons, we consider the QCEW to be the preferred data source for county-level employment. When the results using the QCEW and LAUS data differ non-trivially—which they do in this case—the QCEW findings are much more likely to be accurate.

Sample Alignment

HKMM's sample goes from 2005 through 2012 and uses quarterly data. By contrast, our main specification uses monthly data, starts in 2007m11, and goes through 2014m12. Using quarterly versus monthly data has virtually no impact. For our preferred specification, for example, changing to quarterly data increases the standard errors by a little more than 0.04 and increases the the mean estimate by 0.02 (see **Table 2.6**). Though that represents a 7% increase, since the baseline estimates are small to start with, the impact is quite small. However, switching the time period of estimation from 2005-2012 to 2007m11-2014m12 does make a difference. First of all, as we discussed in **Section 2.5.3**, the 2007m11-2014m12 sample exhibits a fairly symmetric rise and then fall in treatment intensity, orthogonalizing possible trends. Moving to the 2005-2012 sample makes this less so. As can be seen in **Figure 2.4**, the 2005-2012 period is largely a period of (1) increasing benefit duration and and (2) decreasing relative employment on the high-treatment side of the border. However, after 2012, the high-treatment side of the border starts to experience a relative decline in duration, while continuing its relative decline in employment. This is in part due to federal policy changes and in part due to differential changes in unemployment levels. Thus, it is not surprising that adding 2013 and 2014, and removing 2005 to 2007m10, has a noticeable positive impact on the UI duration impact upon employment.

Furthermore, we note that the choice of sample date matters little for the PTT-trimmed sample. **Table 2.4** shows that the OLS estimates in the BCP-FE specification fall from 0.41 to -0.35 when the sample is changed from 2007m11-2014m12 to 2004m11-2014m12. However, the OLS estimates in the PTT-trimmed sample fall only from 0.18 to 0.03. The IV estimates show a similar pattern, although the range is larger in both samples. This leads us to be confident that the large negative effects seen in full-sample specifications with earlier start dates (and/or end dates) reflect endogeneity from pre-existing trends. Furthermore, since the 2007m11-2014m12 sample window effectively orthogonalizes these trends with treatment, we believe that our sample window provides for more reliable estimates than other sample windows, including HKMM's 2005q1-2012q4.

2.8.2 HMM comparison

HMM find that the expiration of EUC at the end of 2013 increased employment, though the implied effect of UI generosity is smaller than that of HKMM. Whereas the latter suggests that approximately 80% of the increase in unemployment during the Great Recession can be explained by the increase in benefit generosity, applying the coefficient estimates of HMM to the 26-to-99 week expansion would imply that UI policy can explain about one third. Scaled another way, HMM finds that the employment effect of the *expiration* is on the same order as total employment gains during 2014. HMM estimate a variety of different empirical models, all of which are motivated by a desire to exploit variation in UI benefits solely coming from the EUC expiration, while at the same time incorporating information over a longer period to formulate a counterfactual for the county-level employment which would have occurred had EUC not expired. Broadly, these specifications can be broken into two groups, which we call the "interaction term" models and the "event study" models.[50] We discuss each of them in turn.

The following is equivalent to HMM's "benchmark" interaction term model, where e_{ct} is log employment, measured either in the QCEW or LAUS:[51]

$$e_{ct} = \kappa[\ln(D_{ct})\mathbb{1}(t \leq 2013q3)] + \alpha[\ln(D_{ct})\mathbb{1}(t \geq 2013q4)] + \mu_c + \nu_{pt} + \gamma_c t + u_{cpt} \tag{2.11}$$

That is, the model includes pair-period fixed effects, county fixed effects, as well as a county-specific time trend. The coefficient of interest is α, which measures the effect of duration on employment solely using variation from 2013q4 onward (i.e., from no earlier than the quarter immediately prior to expiration). The other independent variable, the log of benefit duration in periods prior to 2013q4, soaks up the effect of duration up to 2013q3; this ensures that, after taking out county fixed effects and county-specific linear trends, the model is comparing employment differences in 2013q4 to employment differences in all quarters in 2014.

The first column of the top panel of **Table 2.10** shows HMM's estimate of this specification over the 2010q1-2014q4 period, as well as our replication. They estimate a coefficient of -0.0190, with a p-value of zero (to three decimal places) from a block bootstrap procedure. To place this in the context of our other estimates, this would translate into a -1.05 percentage point reduction in EPOP from a 26-to-99 week expansion of duration. While this is smaller than the corresponding estimate in HKMM, it is still substantial, representing about one third of the EPOP drop of the Great Recession; it would also imply that the expiration of EUC was responsible for increasing employment

[50] The former correspond to models discussed in Sections 3 through 5 of HMM and the latter correspond to models discussed in Section 6 of HMM.

[51] We understand that HMM takes the spatial difference across pairs manually; as discussed above, this is equivalent to including a full set of pair-period fixed effects.

in 2014 by about 2 million jobs. When we estimate this equation using the LAUS data that they use on the county pairs in our sample, we estimate a very similar coefficient of -0.0200, with an analytical standard error (clustered at the state-pair level) of 0.0082,[52] which implies a p-value of about 0.015.[53] However, since HMM accessed their data, the entire LAUS series has been redesigned by the BLS, largely to incorporate information from the American Community Survey rather than the Decennial Census.[54] The second column of the first panel shows our estimate from the same specification but with employment derived from the revised data. The coefficient falls in magnitude by three quarters to -0.0048 and becomes statistically indistinguishable from zero. Thus, when using the most recent version of the LAUS employment series, this specification no longer finds that the 2014 EUC expiration caused an employment boom.

HMM also estimate this model using log employment derived from the QCEW and find a modestly negative estimate of -0.0100. In our scale, this would translate to an EPOP effect of -0.558 percentage points from a 26-to-99 week expansion. When we estimate their model we obtain a similar coefficient of -0.0078, corresponding to an EPOP effect of -0.435.[55] While -0.558 is more negative than our 2014 IV specification (-0.024 in the full BCP-

[52] In our baseline specifications, we cluster two-way at the state and state-pair level in order to account for any common state-level shocks (including mechanical correlation of errors for those counties that border multiple states). For the sake of this reconciliation exercise, we cluster at the state-pair level. Clustering at the two-way level in this specification increases the standard error to 0.0097.

[53] Our baseline sample includes 1,161 county pairs, and we drop an additional two pairs due to missing data in this specification. While our baseline specification studying the 2014 EUC expiration drops pairs in which either county is in North Carolina, we do not drop such pairs in this reconciliation exercise. HMM report using 1,175 pairs with full data. Such a discrepancy could arise due to reasonable differences in interpretation regarding, e.g., whether counties that touch only on a corner should be included as a "county pair."

[54] See `http://www.bls.gov/lau/lauschanges2015.htm` for details. We downloaded the current LAUS data on November 10, 2016.

[55] In our baseline specifications in this paper, we seasonally adjust the QCEW data as described in the text. For the sake of this reconciliation exercise, we use not-seasonally-adjusted data.

FE sample, or -0.094 in the PTT-trimmed sample), the difference is at the bottom end of the range of estimates that can be generated using QCEW data from robustness checks on our main specifications. In results available upon request, we re-estimate our baseline 2014 BCP-FE IV specification using all combinations of the following specification choices: (1) using EPOP, log EPOP, or log employment as the dependent variable,[56] (2) using duration in logs or in levels as the independent variable of interest, (3) keeping county pairs involving North Carolina or dropping them, (4) defining the instrument based on changes in duration immediately upon the EUC expiration, or defining it based on the change between average duration in 2013q4 and the average duration in 2014, (5) starting the sample in 2013q1 or 2013q4, and (6) using seasonally-adjusted or not-seasonally-adjusted data. After translating each estimate to its implied effect on EPOP in levels, we find that these 96 estimates range between -0.637 and 0.473. The EPOP-equivalent estimate from HMM specification using QCEW data (either -0.558 using their estimate or -0.435 using our replication) is within that range, though at the negative end. Furthermore, as with the LAUS specification, we find a lower level of statistical precision than HMM: our standard error of 0.0068 would mean that HMM's point estimate of -0.0100 would not be statistically distinguishable from zero at conventional levels.

HMM repeat their analysis with two variants of their benchmark model. First, they replace the county fixed effects and linear trends with interactive effects (Bai (2009)) and estimate the model over the 2005q1-2014q4 period. Second, they add to the benchmark model county-specific coefficients on three aggregate time series: the price of oil, aggregate construction employment, and reserve balances with the Fed system. We show these estimates in Panels 2 and 3, respectively, of **Table 2.10**. The first column shows HMM's estimate

[56] We do not estimate a specification using employment in levels.

and our replication using the pre-redesign LAUS data.[57] These estimates are qualitatively similar to the estimates from the benchmark model. And, like the benchmark model, the coefficient estimates come much closer to zero when post-redesign LAUS employment data is used, consistent with the null effect of benefit expansions that we find in our baseline specifications. We have not been able to replicate their results with the QCEW.

Additionally, HMM estimate "event study" specifications, as described in their Section 6. These specifications are designed to compare employment in 2014 to what is predicted to have occurred in the absence of the EUC expiration based on pre-expiration data. These predictions are formed by estimating a model using data solely from 2005q1 to 2013q4, and by using the resulting parameter estimates to project the future path of employment in a given county. To estimate the pre-event model, HMM regress county-level log employment on county fixed effects, date fixed effects, a county-specific cubic in the quarterly date, and four lags of log employment. They then define their dependent variable e^*_{ct} as the difference between actual log employment and predicted log employment based on the model parameters. Finally, they recover the effect of the EUC expiration by estimating the following model using observations only from 2014:

$$e^*_{ct} = \alpha\Big(\ln(D_{c,2014}) - \ln(D_{c,2013q4})\Big) + \nu_{pt} + \epsilon_{cpt} \tag{2.12}$$

They estimate a coefficient of approximately -0.02, both using employment from LAUS and from the QCEW, meaning that counties which saw larger declines in benefits than their neighbors (i.e., whose independent variable is more negative) experienced higher growth of log employment in 2014, relative

[57] We calculate standard errors in Panel 2 via a block bootstrap at the state-pair level. We use four factors, as HMM report using for LAUS employment, throughout Panel 2.

to their neighbors, relative to the prediction of their model. As with the estimates found in the "interaction term" models using pre-revision LAUS, this estimate would imply that the 26-to-99 week expansion would explain about one third of the EPOP drop during the Great Recession.

While we have not been able to replicate their results exactly, we do obtain qualitatively similar results. The main result from the event study strategy can be seen immediately in **Figure 2.12**, which plots the time series of the average value of log employment, as well as the series of predicted log employment, for high-benefit counties relative to low-benefit counties (where "high" and "low" status is defined by the size of the drop in log duration between 2013q4 and 2014, relative to the county pair partner). The model predicts that employment in high-benefit counties will continue to fall in 2014 relative to their lower-benefit neighbors, when in fact, a modest reversal occurs. The event study approach attributes this to the effect of the EUC expiration. As in the "interaction term" models discussed above, the redesign of the LAUS series affects the results substantially. When we repeat the analysis using the revised data, we find that the coefficient estimate becomes slightly (and insignificantly) positive, as shown in **Figure 2.13**. HMM also estimate the event study with QCEW data, and find an estimate of -0.0236, which is larger (in magnitude). When we estimate this model using employment from the QCEW, we find a coefficient of -0.0126 (with a standard error of 0.0113), which is in between our estimates for the specifications with revised and vintage LAUS log employment, respectively.[58] This is shown graphically in **Figure 2.14**.

When translated to a change in EPOP, our replication of HMM's event study estimate using the QCEW (-0.703) is substantially more negative than

[58]This standard error takes the parameters of the model estimated in the pre-change period as non-random, likely causing us to understate this standard error. HMM use a bootstrapping procedure to construct these standard errors.

our estimates using EUC expiration, which ranged between -0.024 (full BCP-FE sample) and -0.182 (PTT-trimmed sample). HMM's event study strategy estimates a negative effect of EUC expiration using QCEW data because it constructs a counterfactual where the employment differential between the high and low treatment counties is expected to become more negative in 2014. However, this HMM counterfactual is largely driven by a county-specific polynomial time trend, whose identification is heavily reliant on employment changes that occur up to nine years before the treatment event.[59] As an indication of the type of problem with such a parametric strategy, the employment reversal (both in the QCEW data and, in fact, in the pre-revision LAUS data as well) appears to begin a few quarters prior to the expiration of EUC—a "pre-reversal" which casts doubt on the plausibility of a continuing downward trend as the appropriate counterfactual. In contrast, we take a much more flexible approach by showing whether the employment rates were following parallel trends prior to 2014 by treatment status on the two sides of the border in our 2014 expiration IV. We find that they were, indeed, following parallel trends—as shown clearly in **Figure 2.9** for the full set of border county pairs. And that this employment gap between the two sides of the border remained largely unchanged following the 2014 expiration. We think the more transparent evidence from the 2014 event that we provide in **Figure 2.9** raises questions about the causal import of the parametric model used by HMM to construct the counterfactual employment path.

[59]The use of a cubic trend, rather than some other degree of polynomial, does not affect these results substantially.

Table 2.8: Decomposition of difference between estimates from HKMM and BDGK into contributing factors

Step	From HKMM	To BDGK	Average Marginal Effect
Base Estimate	-2.7238***	0.4299	
	(0.6636)	(0.4946)	
No QFD	2.6883***		1.3156***
	(0.6311)		(0.4192)
Align sample	0.8460	1.4613*	0.8629**
	(0.6930)	(0.8803)	(0.3409)
Urate to EPOP	1.3562	1.1334**	0.7421*
	(1.1691)	(0.4869)	(0.4023)
Bai to FE	0.8300	0.6469	0.2186
	(0.7012)	(0.5636)	(0.2968)
Logs to levels		0.0777	0.0146
		(0.3348)	(0.1392)

Notes: The first row reports the total effect of the expansion of UI from 26 to 99 weeks, in percentage points of EPOP, implied by the coefficient estimates of HKMM (column 1) and the full sample BCP-FE estimates of this paper (BDGK) (column 2). The remaining estimates in the first column represent the increased total implied effect of UI when one specification change is made from the original HKMM estimate. The remaining estimates in the second column represent the effect of taking each final step to arrive at the BDGK estimate. Because the total implied effect is not well motivated by theory when using quasi-differenced data in levels, we leave two cells blank in these first two columns. The third column represents the average incremental effect of taking each step along all possible transition paths between HKMM and BDGK estimates, except that we discard transition paths that involve estimating models with quasi-differenced data in levels. See text for details regarding each step and the conversion of each coefficient estimate into an effect on EPOP. Standard errors are calculated via a block bootstrap at the state-pair level with 300 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.9: Transitioning from HKMM to BDGK estimates: Contribution of factors along three particular paths

	Path 1			Path 2			Path 3	
	Coefficient	EPOP effect		Coefficient	EPOP effect		Coefficient	EPOP effect
HKMM reported result	0.0490***	-2.5885***						
HKMM replication	0.0519***	-2.7238***	HKMM replication	0.0519***	-2.7238***	HKMM replication	0.0519***	-2.7238***
	(0.0093)	(0.6564)		(0.0093)	(0.6564)		(0.0093)	(0.6564)
Eliminate QD	0.0086	-0.0355	Urate to EPOP	-0.0025	-1.3676	Align sample	0.0153***	-1.8778***
	(0.0321)	(0.1327)		(0.0020)	(1.0434)		(0.0030)	(0.4642)
Bai to FE	0.1304***	-0.5825***	Elimate QD	-0.0021	-0.1220	Elimate QD	0.0061	-0.0251
	(0.0415)	(0.2021)		(0.0054)	(0.3180)		(0.0224)	(0.0925)
Urate to EPOP	-0.0275**	-1.6064**	Logs to levels	-0.0298	-0.0298	Logs to levels	0.3197*	-0.2046*
	(0.0123)	(0.7034)		(0.2440)	(0.2440)		(0.1692)	(0.1083)
Align sample	0.0059	0.3523	Align sample	-0.2170	-0.2170	Bai to FE	1.0995***	-0.7035***
	(0.0081)	(0.4870)		(0.1405)	(0.1405)		(0.2473)	(0.1582)
Logs to levels (BDGK)	0.4299	0.4299	Bai to FE (BDGK)	0.4299	0.4299	Urate to EPOP (BDGK)	0.4299	0.4299
	(0.4662)	(0.4662)		(0.4662)	(0.4662)		(0.4662)	(0.4662)

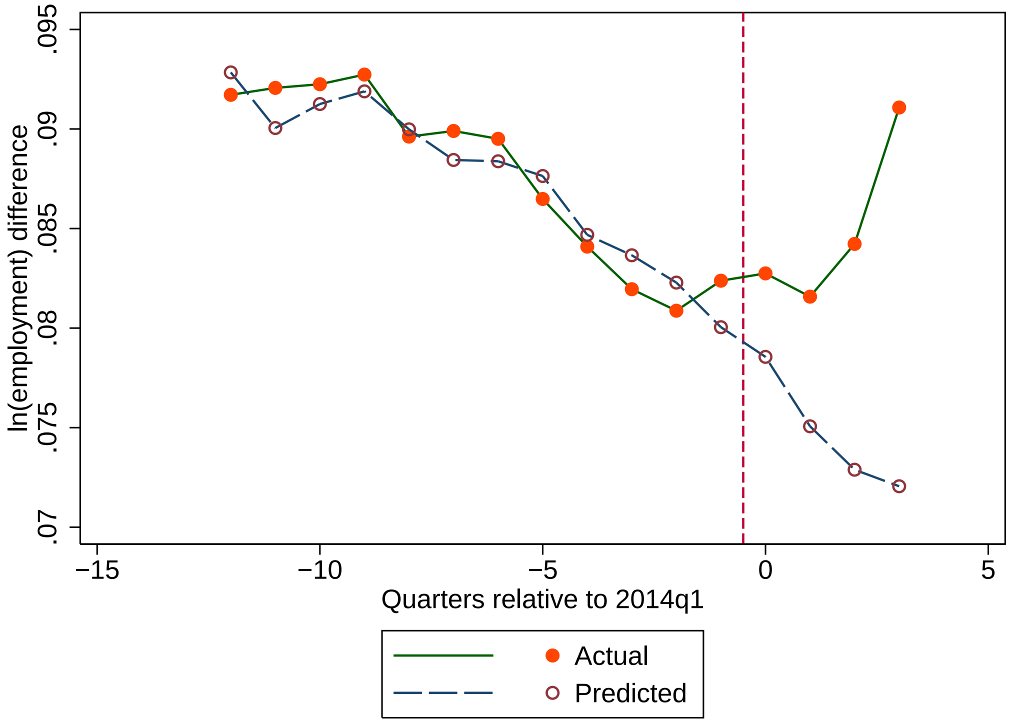
Notes: This table presents three transition paths from HKMM's estimates to the full sample BCP-FE estimates of this paper (BDGK). Each cell presents the coefficient estimate, as well as the implied total effect of the 26-99 week expansion of UI expressed as an implied impact of EPOP, in percentage points. Once a step is made in a given path, it is retained in subsequent specifications in the same path. See text for details regarding each step. Standard errors for specifications involving the Bai (2009) interactive effects estimator are calculated via a block bootstrap at the state-pair level with 300 replications. Standard errors for other specifications are clustered twoway at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.10: Estimates using the HMM interaction-term model: Alternative data sets and specifications

	(1) LAUS (orig.)	(2) LAUS (rev.)	(3) QCEW
Benchmark			
HMM's estimate	-0.0190***		-0.0100***
	[0.000]		[0.050]
Our estimate	-0.0200**	-0.0048	-0.0078
	(0.0082)	(0.0060)	(0.0069)
Observations	46440	46440	46440
Interactive Effects			
HMM's estimate	-0.0233***		-0.0121***
	[0.000]		[0.030]
Our estimate	-0.0231**	-0.0050	-0.0031
	(0.0099)	(0.0073)	(0.0086)
Observations	92720	92720	92880
Natural Factors			
HMM's estimate	-0.0144***		-0.0141***
	[0.000]		[0.020]
Our estimate	-0.0138	-0.0013	-0.0065
	(0.0104)	(0.0070)	(0.0067)
Observations	46440	46440	46440

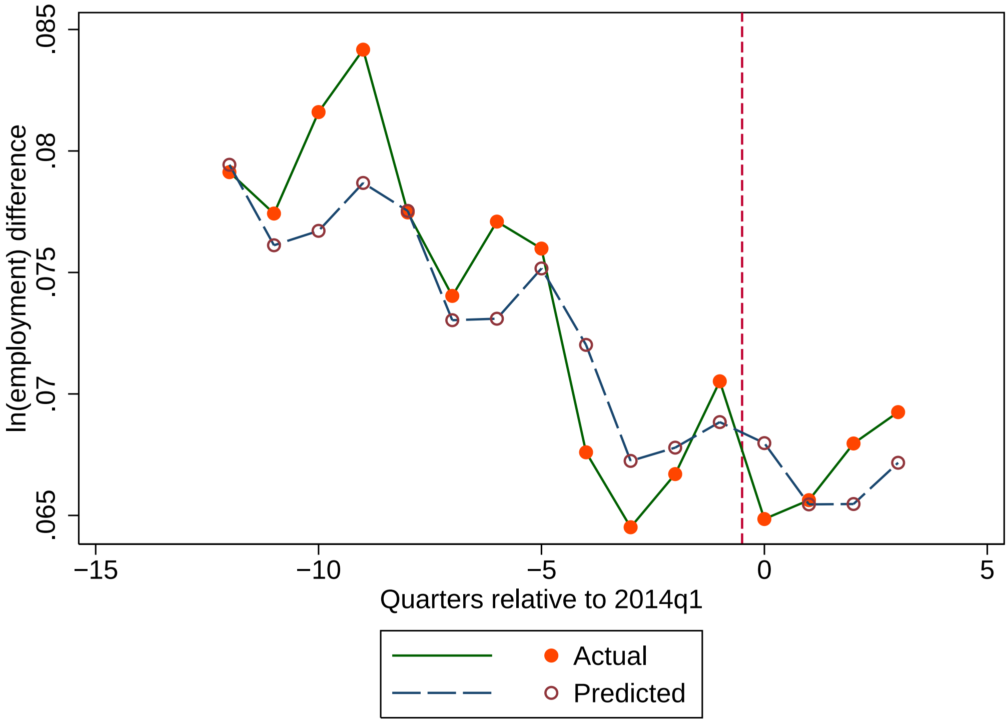
Notes: This table reports estimates of α from HMM's "interaction-term" model: $e_{ct} = \kappa[\ln(D_{ct})\mathbb{1}(t \leq 2013q3)] + \alpha[\ln(D_{ct})\mathbb{1}(t \geq 2013q4)] + \nu_{pt} + \epsilon_{cpt}$, under different characterizations of the error term ϵ_{cpt}. In each panel, the top row reports the estimates reported by HMM, with p-values (from a block bootstrap at the state-pair level) in brackets. The second row reports our replication, with standard errors in parentheses. The first column uses log employment from LAUS, prior to the 2015 redesign. The second column uses post-redesign LAUS data, downloaded on September 9, 2016. The third column uses (not-seasonally-adjusted) log employment from the QCEW. The first panel represents the "benchmark" specification, in which $\epsilon_{cpt} = \mu_c + \gamma_c t + u_{cpt}$. The second panel replaces the fixed effects and county-specific trends with interactive effects (Bai (2009)): $\epsilon_{cpt} = \lambda_c' F_t + u_{cpt}$. The third panel adds to the benchmark specification county-specific coefficients on three national time series: the price of oil, employment in the construction industry, and reserve balances with the Fed system. Standard errors in the first and third panel are analytical, clustered at the state-pair level. Standard errors in the second panel are derived from a block bootstrap at the state-pair level. * p < 0.10, ** p < 0.05, *** p < 0.01.

Figure 2.12: Replication of HMM event study: Pre-revision LAUS employment



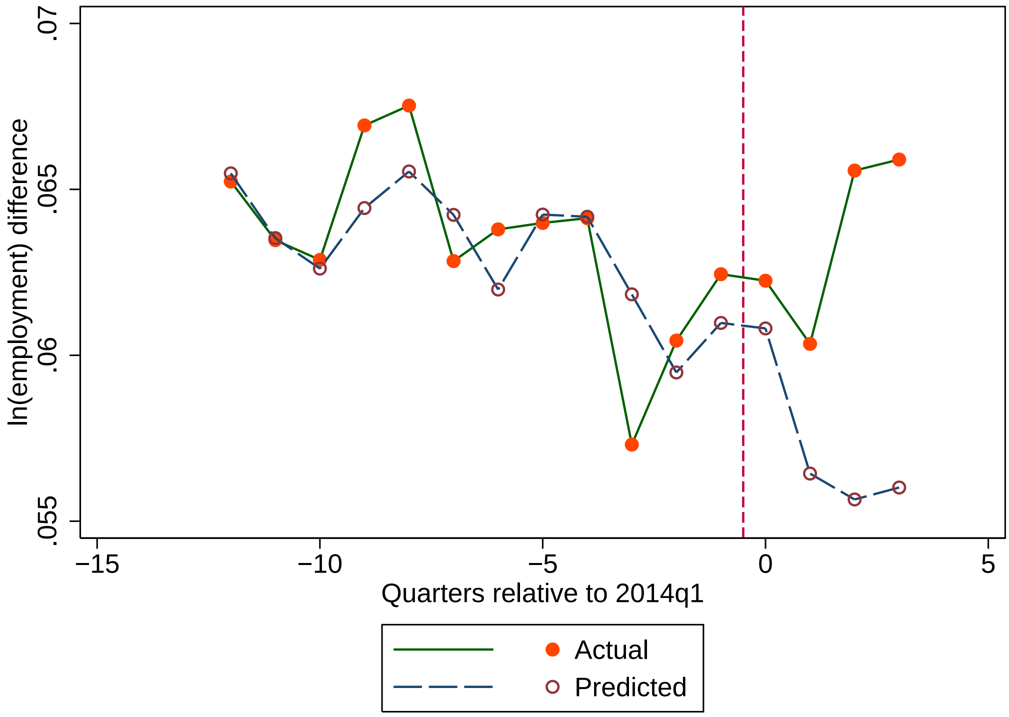
Notes: This figure plots (solid line, solid points) the average difference in log employment between "high" and "low" counties, where a "high" county is defined to have experienced a larger drop in log duration between 2013q4 and 2014 than its neighbor; pairs which experienced identical drops in log duration are not included. The figure also plots (dashed line, hollow points) the average difference in predicted log employment between high and low counties, where the prediction is computed by regressing (on quarterly data from 2005q1 through 2013q4) county log employment on four lags of log employment, time fixed effects, and a county-specific cubic function of the date. Predictions in 2014q1 through 2014q4 are computed recursively. This figure uses employment data from LAUS, prior to the March 2015 redesign.

Figure 2.13: Replication of HMM event study: Post-revision LAUS employment



Notes: This figure plots (solid line, solid points) the average difference in log employment between "high" and "low" counties, where a "high" county is defined to have experienced a larger drop in log duration between 2013q4 and 2014 than its neighbor; pairs which experienced identical drops in log duration are not included. The figure also plots (dashed line, hollow points) the average difference in predicted log employment between high and low counties, where the prediction is computed by regressing (on quarterly data from 2015q1 through 2013q4) county log employment on four lags of log employment, time fixed effects, and a county-specific cubic function of the date. Predictions in 2014q1 through 2014q4 are computed recursively. This figure uses current LAUS data.

Figure 2.14: Replication of HMM event study: QCEW employment



Notes: This figure plots (solid line, solid points) the average difference in log employment between "high" and "low" counties, where a "high" county is defined to have experienced a larger drop in log duration between 2013q4 and 2014 than its neighbor; pairs which experienced identical drops in log duration are not included. The figure also plots (dashed line, hollow points) the average difference in predicted log employment between high and low counties, where the prediction is computed by regressing (on quarterly data from 2015q1 through 2013q4) county log(employment) on four lags of log employment, time fixed effects, and a county-specific cubic function of the date. Predictions in 2014q1 through 2014q4 are computed recursively. This figure uses employment data from QCEW.

2.8.3 Additional Tables and Figures

Table 2.11: Summary statistics for all counties, all county border pairs, and PTT-trimmed sample of county border pairs

	All counties		**Border counties**		**PTT-trimmed**	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
EPOP (2007)	44.19	18.33	44.51	16.20	42.73	15.17
Private EPOP (2007)	34.58	17.45	34.88	15.47	33.24	14.78
LAUS unemployment rate (2007)	4.857	1.686	4.948	1.777	5.046	1.795
Population age 15+ (2007)	76818.0	243398.5	72692.4	178383.3	55884.2	120677.9
Share white (2005-2009 ACS)	0.796	0.190	0.812	0.181	0.817	0.187
Share black (2005-2009 ACS)	0.0885	0.144	0.0834	0.145	0.0884	0.154
Share hispanic (2005-2009 ACS)	0.0755	0.128	0.0620	0.101	0.0540	0.0961
Share high school grad, less than Bachelor's (2005-2009 ACS)	0.564	0.0665	0.568	0.0640	0.570	0.0610
Share Bachelor's degree or higher (2005-2009 ACS)	0.187	0.0852	0.184	0.0818	0.178	0.0785
Median household income (2005-2009 ACS), 2009 dollars	43299.6	11419.7	42949.1	11725.8	41847.9	11682.7
Newly acquired mortage debt per capita (2007)	3.535	3.216	3.508	3.120	3.216	2.829
Share in cities 50k+ (2010 census)	0.186	0.333	0.188	0.328	0.160	0.304
Minimum weeks of UI eligibility over sample period	23.78	4.365	24.17	4.040	24.20	4.025
Maximum weeks of UI eligibility over sample period	91.37	12.15	90.74	12.38	91.00	12.13

Notes: If a border county appears in j county-pairs in the sample in question, then it appears j times for the purpose of creating estimates in this table. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008.

Table 2.12: Estimated effect of UI benefit duration on EPOP in specifications without pair-period fixed effects

	(1) All counties	(2) Border counties
No fixed effects	-3.037***	-3.244***
	(0.556)	(0.756)
County fixed effects	-1.826***	-1.768***
	(0.107)	(0.135)
Time fixed effects	-9.550***	-10.670**
	(3.551)	(4.241)
Twoway fixed effects	-0.385	-0.382
	(0.355)	(0.361)
Observations	266944	199692
No. of counties	3104	1129

Notes: This table reports estimates of the form $E_{ct} = \beta D_{ct} + \text{FE} + \epsilon_{ct}$, where E_{ct} is the ratio of employment to population aged 15+, scaled in percentage points. Row 1 considers models without fixed effect. Rows 2-4 consider models with different sets of fixed effects (FE). Standard errors are clustered at the state level in column 1, and two-way at the state and state-pair level in column 2. If a border county appears in j county-pairs, then it appears j times when creating the estimates in column 2.* p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01.

Table 2.13: Additional robustness checks on the effects of UI benefit duration on EPOP: 2008 and 2014 event samples

	2008 sample IV		**2014 sample IV**	
	(1) BCP-FE	(2) PTT-Trimmed	(3) BCP-FE	(4) PTT-Trimmed
1. Baseline	-0.075	1.550	-2.920**	-4.336***
	(6.180)	(7.298)	(1.235)	(1.507)
2. Private EPOP	0.116	1.186	-2.935**	-4.290***
	(6.188)	(7.476)	(1.291)	(1.646)
3. Correlation-trimmed	-1.510	-1.510	-4.132**	-5.474***
	(3.893)	(3.893)	(1.656)	(1.952)
4. ISLT	-1.241	-1.202	-2.060	-1.657
	(2.357)	(2.626)	(1.369)	(1.661)
5. PTT through 2007m10		-1.335		-2.563*
		(9.437)		(1.403)
6. Quarterly data	-0.259	0.770		
	(6.092)	(7.370)		
7. QWI EPOP (quarterly)	-1.709	-3.287	-2.651**	-4.340***
	(6.113)	(7.899)	(1.338)	(1.451)
8. Distance trimming	2.448	1.550	-3.298**	-4.748***
	(5.801)	(7.298)	(1.339)	(1.628)
9. Unbalanced panel	-0.075	1.550	-2.920**	-4.336***
	(6.180)	(7.298)	(1.235)	(1.507)
10. Hinterland pairs	1.886	3.496	-2.380	-1.189
	(2.968)	(3.310)	(2.125)	(0.774)
11. Exploit Δ reg. benefits			-3.038**	-4.525***
			(1.230)	(1.479)
12. Drop NC	-0.075	1.550		
	(6.180)	(7.298)		
13. Keep NC			-3.857**	-6.065**
			(1.737)	(2.463)
14. NC: Alt. instrument			-1.683*	-3.341**
			(1.001)	(1.330)
15. ln($EPOP$)	0.065	0.114	-0.056**	-0.083**
	(0.131)	(0.145)	(0.027)	(0.033)
	[3.691]	[6.682]	[-2.889**]	[-4.209**]
16. ln(emp)	0.077	0.101	-0.049	-0.075*
	(0.125)	(0.148)	(0.032)	(0.040)
	[4.394]	[5.862]	[-2.516]	[-3.795*]

Notes: Each cell reports regressions analogous to those reported in Table 2.3 for the 2008 and 2014 subsamples (each estimated via IV). The estimates in the 1st row correspond to the estimates in panels 2 and 3 of Table 2.3. The estimates in the 2nd row replace (total) EPOP with the ratio of private employment to population age 15+. In the 3rd row, we trim the set of border county pairs based on the level of correlation between county EPOP and state EPOP over the period 2004m11-2008m10 (see text for details). The 4th row controls for county-specific linear trends. The 5th row trims based on PTT estimated through 2007m10 instead of 2008m10. The 6th row uses quarterly data instead of monthly. The 7th row uses EPOP derived from the QWI (at the quarterly level) instead of the QCEW. The 8th row drops county-pairs whose population centroids are greater than 100km apart. The 9th row includes counties without full EPOP data for each month, which we drop by default. The 10th row uses a modified version of the instrument z_{ct} which exploits all changes in benefits, including changes in regular benefits, which occur at the end of December 2013. Rows 11-13 report estimates using alternative strategies for dealing with North Carolina (NC); by default, border county pairs (BCPs) with one neighbor in NC are kept in the 2008 subsample and dropped in the 2014 subsample. The 11th row completely drops all NC BCPs. The 12th row keeps all North Carolina BCPs. The 13th row keeps NC BCPs but redefines the instrument for NC counties (see text for details). The 14th and 15th row use ln($EPOP$) and ln($employment$), respectively, as dependent variables. The bracketed estimates in these two rows are the level-on-level equivalent, equal to $(\frac{99}{26}^{\hat{\beta}} - 1)\bar{E}$, where $\bar{E}$ is the mean EPOP level in the given sample. Cells which are not applicable in the given sample are left blank. Standard errors are clustered two-way at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.14: Cumulative response of EPOP from distributed lags specification: OLS in first-differences

	(1) BCP-FE		(2) PTT-Trimmed	
	Leads	Lags	Leads	Lags
Contemp.		-0.006		0.034
		(0.096)		(0.084)
Lead/lag 1	0	-0.123	0	-0.063
	(0)	(0.127)	(0)	(0.110)
Lead/lag 2	0.118	0.004	-0.032	0.001
	(0.113)	(0.149)	(0.097)	(0.132)
Lead/lag 3	0.208	0.218	0.083	0.259
	(0.156)	(0.219)	(0.133)	(0.216)
Lead/lag 4	0.263	0.154	-0.047	0.280
	(0.196)	(0.183)	(0.171)	(0.201)
Lead/lag 5	0.148	0.220	-0.075	0.333
	(0.241)	(0.226)	(0.218)	(0.233)
Lead/lag 6	0.243	0.069	-0.037	0.273
	(0.265)	(0.286)	(0.233)	(0.256)
Lead/lag 7	-0.030	0.239	-0.321	0.341
	(0.278)	(0.284)	(0.234)	(0.259)
Lead/lag 8	0.058	0.113	-0.199	0.259
	(0.313)	(0.314)	(0.248)	(0.250)
Lead/lag 9	0.056	0.229	-0.186	0.139
	(0.319)	(0.333)	(0.260)	(0.267)
Lead/lag 10	0.307	0.165	0.122	0.179
	(0.329)	(0.372)	(0.253)	(0.297)
Lead/lag 11	0.403	0.112	0.107	0.106
	(0.341)	(0.372)	(0.247)	(0.312)
Lead/lag 12	0.228	0.168	-0.046	0.136
	(0.334)	(0.390)	(0.249)	(0.336)
Lead/lag 13		0.154		0.198
		(0.399)		(0.346)
Lead/lag 14		0.094		0.056
		(0.406)		(0.355)
Lead/lag 15		0.193		0.230
		(0.465)		(0.409)
Lead/lag 16		0.341		0.360
		(0.440)		(0.386)
Lead/lag 17		0.273		0.304
		(0.463)		(0.402)
Lead/lag 18		-0.003		0.054
		(0.494)		(0.410)
Lead/lag 19		0.155		0.159
		(0.505)		(0.411)
Lead/lag 20		0.162		0.192
		(0.542)		(0.437)
Lead/lag 21		0.168		-0.002
		(0.574)		(0.456)
Lead/lag 22		0.210		0.127
		(0.591)		(0.472)
Lead/lag 23		0.181		-0.030
		(0.577)		(0.476)
Lead/lag 24		0.340		0.012
		(0.590)		(0.502)

Notes: This table reports cumulative monthly lags and leads estimated on the full sample (2007m11-2014m12), using all border county pairs (BCPs) (column 1) and the subset of BCPs in the PTT-trimmed sample (column 2), where all independent variables are divided by 73. The dependent variable is the first-differenced seasonally adjusted ratio of total employment to population age 15+, scaled in percentage points. The regression includes 24 lags and 11 leads and is estimated using EPOP data from 2007m11-2014m12 (and thus duration data from 2005m11-2015m11) in first differences. The zeroth cumulative lag is equal to the estimated coefficient on contemporaneous duration. The jth cumulative lag is equal to the estimated coefficient on contemporaneous duration plus the sum of the estimated coefficient on the 1st through jth lag term. The jth cumulative lead is equal to the sum of the estimated coefficient on the 1st through the $j-1^{th}$ lead term. The 1st cumulative lead is normalized to zero. PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2008. Standard errors are clustered two-way at the state and state-pair level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 2.15: Increase in UI benefit duration from the November 2008 expansion of EUC

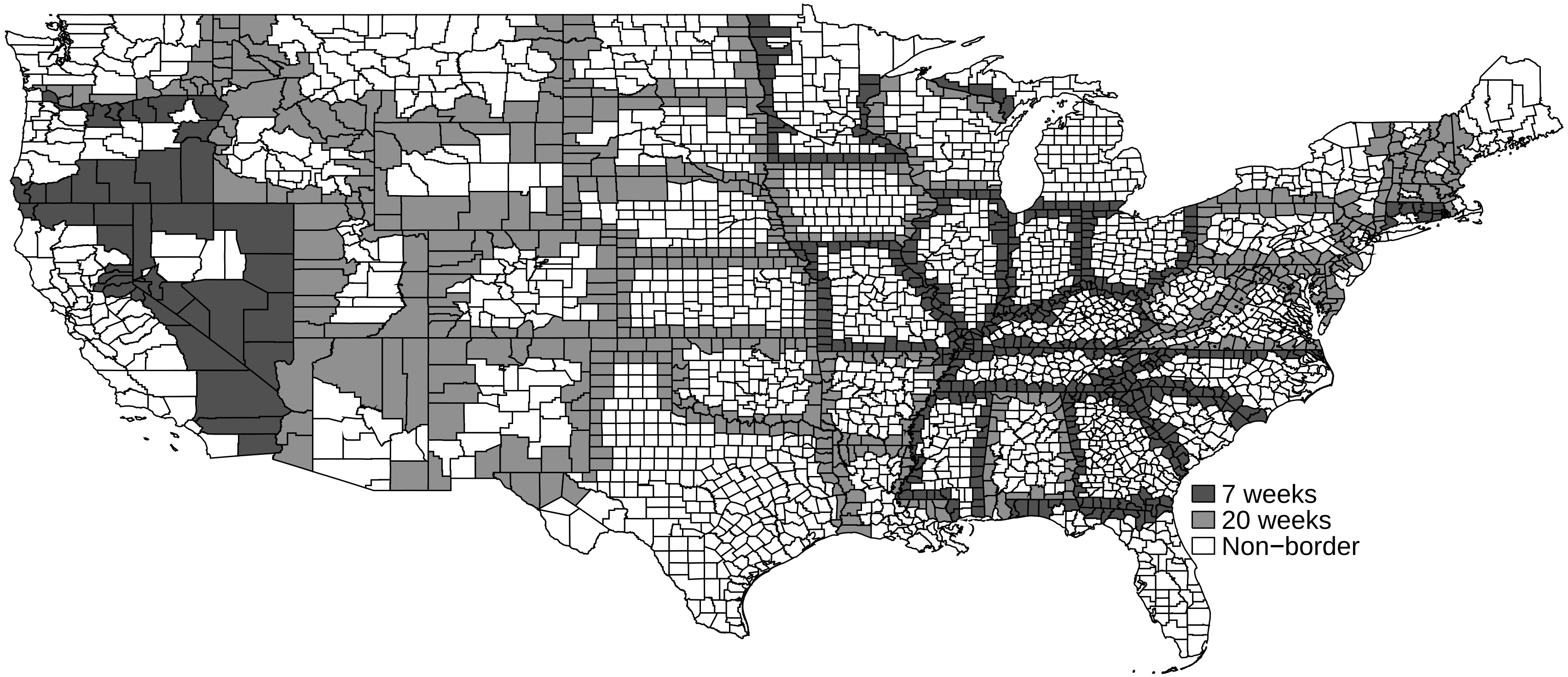
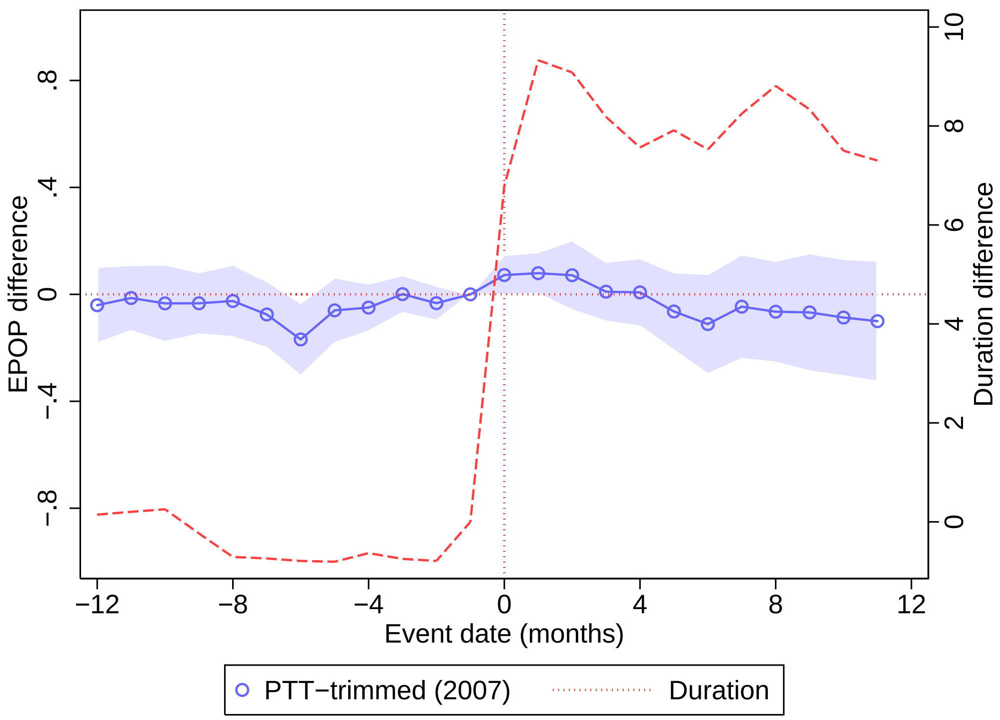


Figure 2.16: Evolution of average EPOP difference and UI benefit duration difference across state borders: Pooled 2008 expansion and 2014 expiration of EUC, sample trimmed based on PTT estimated through 2007m10



Notes: This figure reports the monthly cumulative response of EPOP (left axis, hollow circles) from the pooled 2008 and 2014 samples, using an alternative trimmed sample. In this figure, PTT-trimming removes the quartile of county pairs with the highest differential in linear trends between November 2004 and October 2007 (not October 2008, as before). See notes to Figure 2.7 for additional information.

Figure 2.17: Distribution of EUC differences across border county pairs immediately prior to EUC expiration

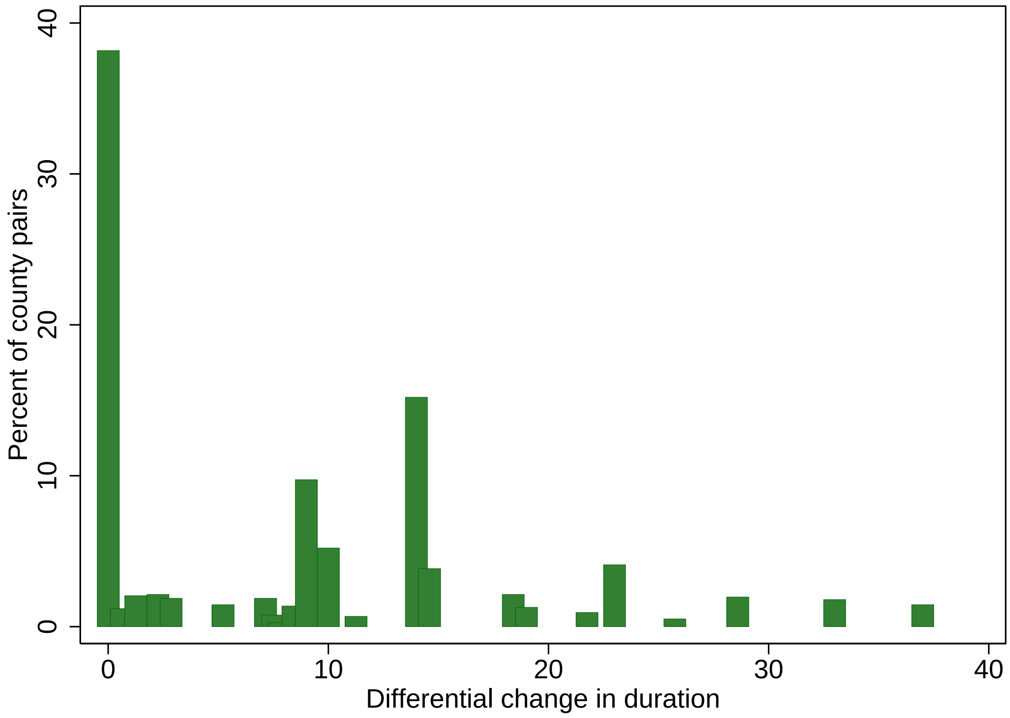
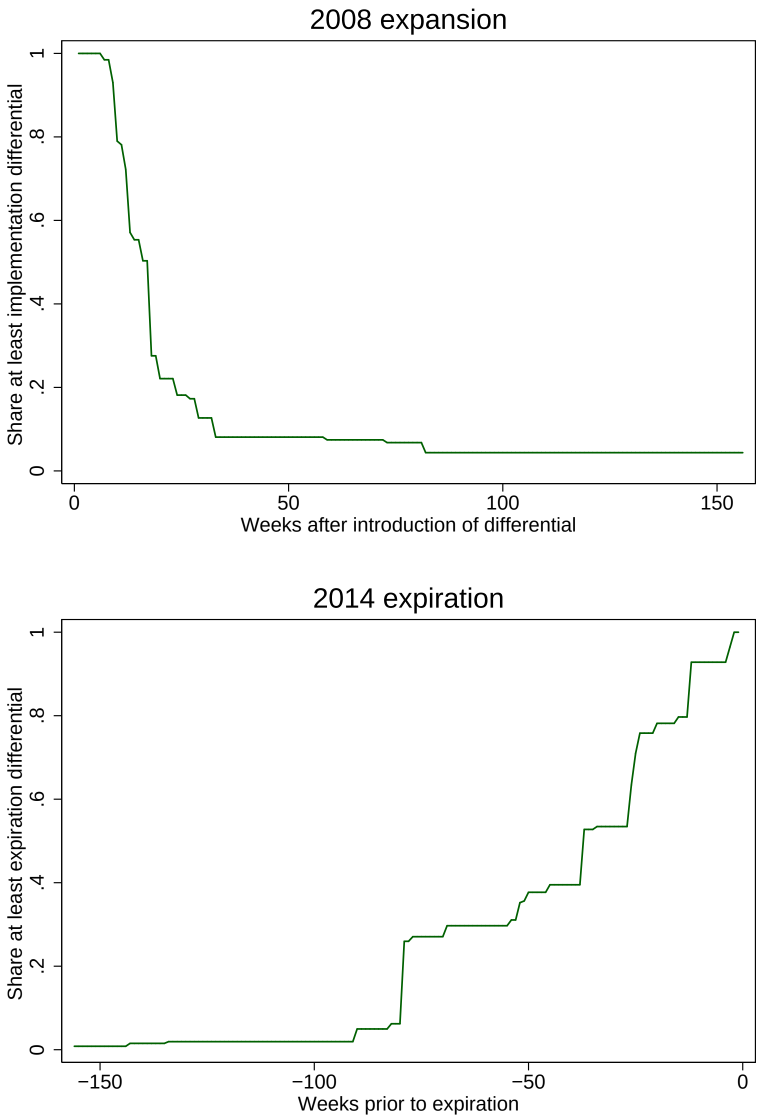


Figure 2.18: Persistence of duration differences in 2008 and 2014 events



Notes: The top graph plots the share of county pairs that continuously have a duration difference at least as large as immediately after the implementation of UCEA in November 2008. The bottom graph plots the share of county pairs that continuously have a duration difference (moving backward in time) at least as large as immediately prior to the 2014 expiration of EUC. The sample of pairs is restricted to those with differential duration at the time of the event in question.

Chapter 3

Shifting by S Corporation Shareholders: Evidence from the 2013 Kansas Tax Reform

3.1 Introduction

In 2013, 4.3 million businesses taxed under subchapter S (of Chapter 1) of the Internal Revenue Code recorded an estimated $6.9 trillion in gross receipts (Internal Revenue Service (2017)). The number of such "S corporations," as they are known, has increased over three-fold since the mid-1980s, compared to a modest decline in the popularity of their cousin, the C corporation.[1] For this reason, there is substantial interest in understanding the behavior of S corporations and their shareholders (Auten, Splinter, & Nelson (2016), Bull & Burnham (2008), Cooper, et al. (2015), Gordon & Slemrod (2000), Nelson (2016), Smith, et al. (2017), among others).

[1]Most S corporations are modest in size; in the tax data that I use in this paper, about 60% of firms have a single shareholder and 99% have six shareholders or fewer. C corporations can be much larger than S corporations, as S corporations are allowed no more than 100 shareholders. But many smaller firms are C corporations as well. The Internal Revenue Service (2017) estimates that there were 1.6 million active C corporations in 2013, with total gross receipts of $20.3 trillion.

In this paper, I focus on a unique decision that owners of S corporations must make: how much of their operating income (roughly \$55,000 for the median profitable firm in 2012) to pay themselves in the form of wages, leaving the remainder to be labeled as "profits." For certain S corporations, such as those with a single owner, this distinction between wages and profits would be meaningless aside from tax consequences. Yet, wages face an additional tax (FICA) of up to 15.3 percent, meaning that the tax-minimizing strategy for most S corporation owners is to pay themselves a wage of zero.[2] In order to protect the tax base, however, the Internal Revenue Service (IRS) requires owners of S corporations to pay themselves "reasonable compensation" under threat of audit. Nevertheless, given the lack of objective standard for what defines "reasonable compensation," there is considerable leeway for S shareholders to reclassify wage income as business income.

The reclassification of wages as profits has a non-trivial effect on tax revenue. The Government Accountability Office (2009) estimates that \$24 billion in wages was underreported by S corporations in 2003 and 2004, in the sense of being recoverable by audit. Additionally, there is a substantial amount of underreporting does not meet the strong standard of being recoverable by audit. In particular, Bull and Burnham (2008) estimate that S corporation shareholders understate their wages by a factor of one-third, which would correspond to approximately \$140 billion in underreporting over this same two-year period. Furthermore, subject to certain restrictions, the Tax Cuts and Jobs Act will allow for a partial deduction for pass-through income beginning in 2018 – effectively reducing the tax rate applicable to that income. Because this partial

[2]This assumes that the increase in the present value of future Social Security benefits brought about by a one dollar increase in wages is less than the FICA tax payment, which will generally be true for all but very low-income individuals under plausible assumptions. Furthermore, 50 percent of the FICA tax is deductible under the income tax, so the true wedge is slightly smaller than 15.3 percent.

deduction will generally apply to profits earned by S corporations but not the wages paid to shareholders, this has a side effect of increasing the incentive for some owners of S corporations to reclassify wages as profits. Thus, it is of considerable interest to determine how elastically owners of S corporation respond to an increase in this tax wedge.

This paper analyzes such an increase in the tax wedge between wages and profits. It exploits a natural experiment that took place in Kansas beginning in 2013 which exempted (among other things) the profits earned by S corporations in Kansas from state income tax, but did not exempt wage income. As a result, this reform added up to 4.9 percentage points to the tax wedge between wage and subchapter-S business income, more than doubling the tax wedge for high-earnings shareholders and increasing the tax wedge for lower-earnings shareholders by about one-third. I use administrative tax data to compare the evolution of wages paid to shareholders in Kansas, relative to the evolution in other states. The baseline difference-in-differences estimates imply only a small reduction in wages paid to shareholders in Kansas of about three percent. Using a randomization inference technique, this estimate is insignificantly different than zero, and I can rule out effects more negative than 16 percent of baseline shareholder-employee wages. I additionally control for the change in operating income (i.e., income calculated before deducting wages to shareholders) to ensure that the estimates correspond to a shifting effect, rather than a combination of a shifting effect and an effect on general business activity of Kansas firms. This control variable has little effect, as the change in operating income in Kansas was similar to that of other states.

I then use two methods to refine the construction of the control group. First, I reweight the sample by inverse propensity scores, based on the predicted probability of being a Kansas firm based on observed covariates. Second, I use the synthetic controls method. Neither of these strategies affect the coefficient estimate substantially. However, the improved matching between treatment and control causes the confidence intervals to shrink; in the most precision specifications, I can rule out effects more negative than 6 to 8 percent.

This muted response stands in contrast to the substantial shifting found in the literature along this margin. In particular, Auten, Splinter, and Nelson (2016) show that high-income owners of S corporations reduced the wage share of their income by up to 40% after the 1994 "uncapping" of the Medicare portion of the payroll tax introduced an incentive to do so.[3] Bull and Burnham (2008) make a similar finding: the authors estimate that wages to shareholders are 33% lower due to the tax wedge. Smith, et al. (2017) estimate that firms that switch from C corporations to S corporations reduce their total wage deductions by approximately two percent of gross sales.[4]

Furthermore, there is considerable evidence that taxpayers respond sharply to shifting incentives in other contexts.[5] For instance, the growth of S corporations and other pass-through entities itself is plausibly an example of shifting: it is a stylized fact that income migrated away from the corporate base (i.e., the base applying to C corporations) and toward pass-through entities such as S corporations after the Tax Reform Act of 1986 reversed the order of the top corporate and individual tax rates (Slemrod (1995, 1996), Carroll and Joulfaian (1997), Auerbach and Slemrod (1997), Saez (2004), among others). Additionally, Auten, Splinter, and Nelson (2016) showed that the amount of

[3]Auten, Splinter, and Nelson (2016) defined "high-income" to mean having combined wages and S corporation income in excess of $1 million.

[4]Of note, during the period studied by Smith, et al. (2017), C corporations generally faced an incentive to over-report wages relative to the arm's-length wage. Thus, this estimate is an upper bound for the under-reporting of wages by S corporations.

[5]Generically, I define "shifting" to mean changes in the "location" of income recognition, where "location" can mean different time periods, different tax bases, or different jurisdictions.

"passive" S corporation income decreased substantially after the Net Investment Income Tax was introduced in 2013, which applied to "passive" but not "active" income from S corporations. Furthermore, there is a deep literature showing that taxpayers are responsive to incentives regarding the timing of transactions. For instance, Burman and Randolph (1994), Auerbach and Poterba (1988), Burman, Clausing, and O'Hare (1994), and others have shown that taxpayers increase their realization of capital gains right before scheduled increases in capital gains rates take effect. Goolsbee (2000) showed that executives of large firms accelerated the exercise of stock options before tax increases took effect in 1993; Sammartino and Weiner (1997) confirm the results of Goolsbee (2000) using taxable income more generally. Gorry, et al. (2017) similarly find that deferred executive compensation increases when tax rates are high. Relatedly, Dowd and McLelland (2017) find that taxpayers tend to delay the realization of capital gains until the preferential rates associated with holding capital assets for 365 days come into effect.

To rationalize the relatively small response of S corporation owners to HB 2117, I adapt standard models of tax evasion (e.g., Yitzhaki (1987)) to this present setting, and make a minor modification. I show that these models predict that the size of the shifting effect is a decreasing function of the baseline wedge; put differently, the amount of shifting in response to the introduction of a wedge can be much larger than the response to the strengthening of an existing wedge. There are two stylized forces that drive this result. First, one component of the marginal cost of shifting is that it increases the risk of detection, which leads to reversal of the entire stock of shifting; when the tax wedge increases and the amount of existing shifting is greater than zero, the existing shifting becomes more valuable and taxpayers become more cautious. Second, the shape of the perceived probability of audit function may directly cause the marginal cost curve to be convex. Intuitively, taxpayers may have substantial leeway to shift income at low cost, as there may be a wide range of wages that the taxpayer believes will pass IRS scrutiny with high probability. But, as the amount of shifting increases, the IRS becomes more suspicious at a faster rate. The resulting convexity of the marginal cost curve causes the magnitude of the shifting effect to shrink when the baseline wedge is larger.

Consistent with either of these two forces inherent in these models, I find evidence that S corporations that faced a lower tax wedge prior to the reform showed a stronger shifting response. In particular, owners of S corporations with initial wages or self-employment income (from all sources) in excess of the Social Security cap face an initial tax wedge of 2.9% or 3.8%, rather than 15.3%. I find a somewhat larger shifting effect for these firms. However, these results are consistent with other explanations as well, such as the tax change being more salient to larger, more sophisticated firms. Furthermore, the precision of these estimates drops noticeably.

In addition to the literature studying the effect of tax incentives on shifting or reclassifying income, this paper also fits into a nascent literature studying the effect of HB 2117 in Kansas. The work most similar to mine is DeBacker, Heim, Ramnath, and Ross (2017). Like me, the authors use tax data to study the effect of HB 2117 on pass-through businesses. They find a modest increase in the number of taxpayers reporting Schedule C (sole proprietor) income, along with suggestive evidence that some of this increase is driven by recharacterizing labor income as business income through independent contracting arrangements. Their findings regarding Schedule E (partnerships, S corporations, rent, and royalties) are more mixed. They do not directly examine wages paid by S corporations to their shareholders. By contrast, I link tax returns and information returns to explicitly analyze wages paid to shareholders, thereby providing direct evidence regarding one of the most plausible channels for income shifting in response to HB 2117.

Additionally, McCloskey (2017), Turner and Blagg (2017), Mazerov (2016) study the effect of HB 2117 on employment. They generally find that Kansas employment growth fell modestly after the implementation of HB 2117 relative to employment growth in control states and counties, though the statistical significance of these findings is not clear. The present paper focuses primarily on a different outcome (reported wages to shareholders), which need not be linked closely to the effect on economic outcomes such as employment.

The remainder of this paper is organized as follows. Section 3.2 presents the policy background. Section 3.3 describes the administrative tax data used in this paper. Section 3.4 presents the main empirical strategy and results. Section 3.5 presents the results from two refinements of the baseline empirical strategy. Section 3.6 provides a discussion of the results in the context of standard theories of tax evasion. Section 3.7 performs a suggestive empirical test of this model. Section 3.8 provides for sensitivity analyses. Section 3.9 concludes.

3.2 Policy Background

Broadly speaking, business income is taxed under the federal income tax in one of two ways.[6] Most large businesses are taxed under Subchapter C (of Chapter 1) of the Internal Revenue Code; such firms are known as "C corporations." C corporations pay an entity-level tax on income. Additionally, taxable shareholders of C corporations generally pay individual income tax (sometimes at preferential rates) on dividends and capital gains derived from

[6]This section presents a very brief overview of the law as it existed prior to Tax Year 2018. See JCX-42-17 (Joint Committee on Taxation (2017)) for a more full treatment.

their holdings in these corporations.

Most other businesses fall into the category of "pass-through" businesses. The income from these firms is allocated to each owner, whether distributed or retained in the business. During the period under study in this paper (i.e., prior to 2018), each owner then paid tax on this income under the usual income tax rules that apply to him or her. The three main subcategories of pass-through businesses are partnerships, sole proprietorships, and S corporations. The present study concerns S corporations, which are so named because their tax rules are provided in Subchapter S.[7] **Figure 3.5** uses data from Joint Committee on Taxation (2017) to show the evolution of the number of firms organized as C corporations, S corporations, partnerships, and sole proprietorships, normalized to one in 1985, the year before the Tax Reform Act (TRA) of 1986. The figure shows the enormous growth in the number of firms organized as S corporations in the aftermath of TRA 1986, which made pass-through entities more tax-favored relative to C corporations.

An entity can elect to be taxed as an S corporation if it meets certain requirements. In particular, it must have no more than 100 shareholders, generally all of these shareholders must be U.S. natural persons (e.g., not partnerships or other corporations), and all shareholders must be permanent residents and/or citizens of the United States.[8] Furthermore, S corporations may not have multiple classes of stock.

The present study concerns how S corporations allocate their operating income, which I define as revenues minus costs other than shareholder com-

[7]Limited liability companies (LLCs) organized under state law can generally elect treatment as one of several subcategories if they meet certain requirements. Single-owner LLCs can choose to be taxed as a sole proprietorship (which is the default) or as an S corporation. Multi-member LLCs can choose to be taxed as a partnership (which is the default) or as an S corporation. For the purpose of this study, the definition of "S corporation" includes LLCs that elect to be taxed as an S corporation.

[8]Under certain circumstances, trusts and employee stock option plans may also be shareholders of S corporations.

pensation. A given firm must allocate some portion of this income as "wages" that it pays to its shareholder-employees, and the remainder as "profits" earned by the firm. For a single-owner firm, this distinction is meaningless but for tax: wages paid to the shareholder reduce the profits of the firm – over which the shareholder has exclusive ownership – dollar for dollar. Furthermore, prior to 2018, both the wages and profits of the firm were subject to federal income tax in the same way.[9] However, there exists a tax wedge which makes the wages relatively costlier than profits. This wedge arises because FICA tax of up to 15.3 percent is imposed on wages to shareholders (just as it is generally imposed on wages generally) but not on the profits of the firm.[10] Formally, half of this 15.3 percent is statutorily leveed on the employer, while the other half is levied on the employee; the "employer share" is deductible against income tax, which effectively reduces the wedge to $15.3 \times (\frac{1}{2} + \frac{1}{2}(1-\tau))$, where τ is the marginal income tax rate.[11] Importantly, the business income earned by the firm is subject neither to FICA nor the essentially equivalent self-employment tax under SECA. This creates an incentive for owners of S corporations to reduce the wages they pay themselves and thus increase the business income earned by the firm.[12]

[9]There are exceptions to this general rule. For instance, in the unusual case when the S corporation passes through losses to the taxpayer in excess of the taxpayer's basis in the corporation, an additional dollar in wages would be taxable while the additional dollar of losses would not be deductible. As basis cannot easily be observed in the data, this study abstracts from this complication.

[10]As of 2013, S corporation profits allocated to a high-income passive shareholder are generally subject to the 3.8% Net Investment Income Tax (NIIT). I abstract from this complication.

[11]For total wages/self-employment income ("earnings") under the Social Security cap ($117,000 in 2014), the tax rate is generally 15.3%. In 2011 and 2012, the employee share was temporarily reduced by 2 percentage points. For earnings between this cap and $200,000 ($250,000 for joint filers), the rate reduces to 2.9%. As of 2013, for earnings above this threshold, the rate increases to 3.8%, with the additional 0.9% paid by the employee.

[12]On the margin, wages (and self-employment income) earned below the Social Security cap can increase the Social Security benefits that an individual will be entitled to upon retirement. This tax-benefit linkage – if understood by taxpayers – can reduce the true wedge between wages and other types of income below the statutory wedge. However, this tax-benefit linkage effects only the initial tax wedge; the policy variation under study in this

To prevent a loss of tax revenue, the IRS requires shareholders of an S corporation to take "reasonable compensation." In IRS Fact Sheet 2008-25, the IRS explicitly says that there is no explicit test for what makes compensation reasonable: "There are no specific guidelines for reasonable compensation in the Code or the Regulations." Instead, the IRS lists nine factors that courts have used on a case-by-case basis.[13] In several court cases, the IRS has successfully increased the amount of wages treated as having been paid to certain taxpayers. Most notably, in *Watson v. U.S.*, the United States Court of Appeals for the 8th Circuit upheld the IRS's determination that accountant David Watson's $24,000 salary (in years in which he received profit distributions in excess of $175,000) was unreasonably low. In this case, the IRS increased the wage payment by over $65,000 in two years, and Watson was required to pay FICA tax on the difference. Nevertheless, the definition of "reasonable compensation" remains highly subjective, leaving taxpayers with considerable leeway to "push the envelope."

Furthermore, practical details suggest that adjustment frictions are not very large. In particular, the tax withholding associated with wages to shareholders (just as with other wages) must be made at least quarterly, with monthly or more frequent payments being required in many circumstances. Thus, S corporations must make an active decision regarding wages to shareholders fairly early in the year, though this decision can be partially changed later in the year. The data provide further evidence of a lack of adjustment

paper increased the tax wedge between wages and profits, and did not affect relationship between wages and future Social Security benefits. Furthermore, the progressivity of Social Security benefits means that the Social Security-inclusive tax wedge (holding current wages fixed) is increasing in lifetime wages.

[13]These factors are "[t]raining and experience," "[d]uties and responsibility", "[t]ime and effort devoted to the business," "[d]ividend history," "[p]ayments to non-shareholder employees," "[t]iming and manner of paying bonuses to key people," "[w]hat comparable businesses pay for similar services," "[c]ompensation agreements," and "[t]he use of a formula to determine compensation."

frictions: in my data, only 16% of those firms with positive wages to shareholders in year t pay their shareholders the same nominal amount in year $t+1$.[14]

3.2.1 Policy variation: HB 2117

This study exploits variation created by a tax reform in Kansas, known as HB 2117, which took effect January 1, 2013. Among other reforms, the law exempted pass-through income — specifically, amounts recorded on Schedules C (sole proprietorships), E (partnerships, S corporations, rents, and royalties), and F (farms) on federal Form 1040 — from state income tax. The exemption did not apply to wages, including wages earned by owners of S corporations. Thus, the wedge between wages and business income increased by 4.9 percentage points for most affected taxpayers, the rate that applied to taxable income greater than $30,000.[15] This tax change was likely quite salient for affected individuals. The bill received considerable local press at its signing in May of 2012.[16] Furthermore, it continued to receive substantial press over the next several years, as it became a controversial political issue within Kansas. In 2017, the Kansas legislature repealed (among other things) the exemption of pass-through income, overriding the governor's veto.

Although HB 2117 affected sole proprietorships and partnerships, I do not study such entities in this paper. Unlike S corporations, owner of sole proprietors generally do not pay themselves wages for tax purposes; instead,

[14]Anecdotally, one common piece of advice for S corporation shareholders is to pay wages to shareholders that are a fixed share (e.g., 50%) of operating income. In unreported results, I find no bunching in the distribution of the ratio of wages to operating income, or the ratio of wages to lagged operating income.

[15]This top rate of tax was reduced from 6.45% to 4.9% in 2013, 4.8% in 2014, and 4.6% in 2015.

[16]`http://cjonline.com/news/2012-05-22/brownback-signs-large-state-tax-cut`, `http://www.kansas.com/news/politics-government/article1092505.html`, `http://www.kansas.com/news/business/article1092681.html`

all income earned by sole proprietors is generally subject to self-employment tax and all such income was exempted under HB 2117. Furthermore, some partners (e.g., limited partners) face an incentive to reduce wage-equivalent payments known as "guaranteed payments" in order to avoid self-employment tax. However, HB 2117 initially applied the exemption from state income tax to all income earned by partners, including guaranteed payments, meaning that HB 2117 did not initially increase the relevant tax wedge for partners. For these reasons, I restrict attention to S corporations in this study.

3.2.2 Section 199A deduction for qualified business income

As part of Public Law No: 115-97 (originally referred to as the "Tax Cuts and Jobs Act"), Congress introduced Section 199A to the Internal Revenue Code. Under Section 199A, taxpayers may generally claim a deduction equal to 20 percent of income earned through pass-through businesses, including S corporations – reducing the effective marginal tax rate on that income. This deduction is subject to various restrictions. First, the deduction is not available to a "specified service trade or business", which includes firms in such industries as law, accounting, financial services, and consulting.[17] Second, the deduction is limited to $\max(0.5W, 0.25W + 0.025A)$, where W is the total W2 wages paid by the business (including but not limited to wages to shareholders) and A is the original purchase price of certain depreciable assets held by the business. However, neither of these two limitations (the denial to service firms and the wage/asset limitation) apply if the taxpayer has taxable income (computed without regard to this deduction) less than \$157,500 (or \$315,000 for married couples filing jointly).

[17] To be clear, this is a non-exhaustive list.

Generally, Section 199A will have the effect of increasing the wedge between wages to shareholders and profits earned by S corporations, since the profits will effectively be subject to a lower federal income tax rate (by $0.2 \times \tau$, where τ is the marginal rate) than wages. However, unlike HB 2117, this will not be true for all firms: some service-firms owned by high-income shareholders will not experience this increase in wedge. Furthermore, for a small number of S corporations, the wage limitation (limiting the deduction to $\max(0.5W, 0.25W + 0.025A)$) could actually cause the wedge to *fall*.[18] Thus, while studying HB 2117 can provide insights on the effect of wages to shareholders generally, one must be careful in applying this result to predict the effects of Section 199A.

3.3 Data and Sample Construction

This paper uses data drawn from the database of administrative tax records. The underlying database contains the near-universe of certain tax forms over certain years. I use data primarily from three forms. First, I use data from Form 1120S, which is filed by all S corporations. Form 1120S identifies the corporation (typically through the Employer Identification Number (EIN)) and reports various income and deduction items as well as the number of shareholders. Second, I use Schedule K-1 of Form 1120S, which is issued to each shareholder of an S corporation. In addition to identifying the identity of each shareholder (generally their Social Security Number (SSN)), Schedule K-1 allocates income and/or loss items to each shareholder. Third, I use Form W2, which reports wage amounts paid by an firm to an individual. I use Form W2

[18]Consider a firm that earns operating income z before paying a wage w to its shareholders. It has no other wages, and owns no assets. The owner has taxable income high enough to be subject to the wage limitation. For this firm, the deduction will equal $\min(0.2(z - w), \max(0.5w, 0)) = \min(0.2(z - w), 0.5w)$. A higher wage will lead to a larger deduction (reducing or reversing the wedge) whenever $0.5w < 0.2(z - w)$, or $\frac{w}{z} < \frac{2}{7}$.

(in combination with the shareholder-firm links established by Schedule K-1) to calculate the total wages paid to all shareholders in a given year by a given firm.[19] The unit of observation is therefore at the firm-by-year level. Because HB 2117 applies to Kansas residents earning income from activities performed in Kansas, I restrict the sample to firms with at least one shareholder living in the same state (according to Schedule K-1) as the firm's home state (as indicated by Form 1120S).[20]

I use the universe of S corporations from 2003 to 2015 with personal identifiers (such as SSNs and EINs) masked to protect taxpayer confidentiality.[21] I make several restrictions to the data. First, the CDW data is raw and unedited, and thus have several imperfections that require careful attention. I drop firm-year observations with any of these imperfections. For instance, I drop firm-year observations for which I cannot find a Schedule K-1. I also drop firm-year observations in which the income reported on Form 1120S does not match the sum of the income reported on the Schedules K-1. Further details are discussed in **Section 3.10.1**.

Second, I keep only those firms for which wages and profits are one-for-one substitutes (at least over some range). This is most apparent for single-shareholder firms, as observed by Bull and Burnham (2008), but more generally this property applies to all firms in which all shareholders are employees.

[19] I define wages as Box 1 wages from Form W2. However, owners of S corporations generally include in Box 1 wages health insurance premiums paid or reimbursed by the corporation. Thus, ideally, one would prefer a measure of wages that strips out these health insurance premiums. However, strategies for doing so introduce their own problems, as discussed in **Section 3.10.1**. The inclusion of health insurance premiums in the definition of wages will have the effect of adding noise and, as also discussed in **Section 3.10.1**, could conceivably have the effect of biasing my results downward, suggesting that the true shifting effect is less negative than my estimates.

[20] Additionally, I use Form 1041, Schedule K-1 to identify the beneficial owners of the small number of trusts that are themselves the shareholders of S corporations. See **Section 3.10.1** for details.

[21] Prior to 2003, the Form 1120S data do not include the number of shareholders, which I use as a data quality check (see below).

Such firms can construct a wage change that holds pre-tax allocations fixed, and thus face the shifting incentive most strongly.[22] By contrast, if at least one shareholder has zero wages, then it is impossible to reduce wages without changing pre-tax allocations: intuitively, the share of profits owned by the set of shareholder-employees increases by less than one dollar for every dollar of wage reduction. Ideally, I would operationalize this sample restriction by dropping firms with at least one shareholder that is truly not an employee (i.e., a shareholder that would not receive a wage under a counterfactual of no tax wedge). This is not observable. Thus, I approximate this ideal restriction by dropping firms that, as of time $t-1$, have (1) multiple shareholders and (2) at least one of these shareholders does not receive a wage from the firm.[23] I do not drop any single-shareholder firms in this step, as the vast majority of shareholders of single-shareholder firms perform services for the corporation and thus are "potential" employees. For the sake of brevity, I will refer to the survivors of this restriction as "Type A" firms (and the remainder as "Type B" firms).

This second restriction (and the definition of the dependent variable, discussed below) leads mechanically to the third restriction: I require a firm to be in the raw data, and free of the data imperfections described above, at time t and $t-1$ in order for the firm to be included in the estimation sample at time t.

These various selection criteria are non-trivial and introduce some non-randomness into the sample. **Figure 3.6** shows that the share of firms that are dropped starts just over 0.6 in 2004 – meaning that only 0.4 of firms are kept – and decreases to 0.45 by the end of the sample period. **Figure 3.7**

[22]In particular, suppose shareholders $1, \ldots, N$ (indexed by i) have ownership shares $\lambda_1, \ldots, \lambda_N$ that sum to one. If each shareholder's wage is reduced by $\theta\lambda_i$ for some positive amount θ, pre-tax allocations are unchanged.

[23]I will examine such firms separately in the Robustness section.

shows the effect of each of the three steps sequentially. The upper-left panel shows the share of observations dropped when requiring presence in the raw data in $t-1$ and t. In Kansas, this share drops from about 0.11 to 0.08 over the sample period, with a share consistently slightly higher in states other than Kansas. The upper-right panel enforces sample-selection due to data quality (and re-forces sample balance over $t-1$ and t). This restriction – to be free of data problems for two consecutive years – causes me to drop about 40% of the remaining sample in the beginning period, and about 20% of the remaining sample at the end of the sample period. Finally, the bottom panel restricts the sample to "Type A" firms, which causes me to drop about 30% of the remaining sample in Kansas and 25% in other states. Reassuringly, the trends in the share dropped in all three steps are quite parallel in Kansas relative to other states.

Table 3.1 presents summary statistics for the main estimation sample in 2010 through 2012. There are approximately 51,400 such firm-year observations in Kansas and 6,962,300 such firm-year observations in the sample in other states.[24] I report the share of firms that are single-shareholder firms, the average number of shareholders, and medians of total income (gross receipts less cost of goods sold), ordinary income, wages paid to shareholders, and total reported assets.[25] The first column reports these statistics in Kansas, the second column reports these statistics in the four bordering states of Colorado, Missouri, Nebraska, and Oklahoma, and the third column reports these statistics for all states other than Kansas (including D.C.). The number in brackets indicates Kansas's ranking among all 51 states for the statistic in question. Of note, Kansas firms tend to be larger and earn higher profits than firms in other states. Total income and total assets are each about 50% larger in

[24]To protect taxpayer confidentiality, these counts are rounded to the nearest 100.

[25]The median is rounded to the nearest $100 in order to protect taxpayer confidentiality.

Kansas than in other states, and wages are about 70% larger in Kansas than in other states, and ordinary income is nearly double other states. However, these statistics seem to match other similarly rural states. For instance, Kansas has the 8th highest ordinary income median of all states (as indicated by the "8" in brackets); the seven states with a higher ordinary income pseudomedian are (in order) North Dakota, Alaska, South Dakota, Nebraska, Wyoming, Iowa, and Oklahoma.

Of note, roughly 27% of firms in Kansas (and 37% in other states) pay zero wages to shareholders, consistent with results reported in Bull and Burnham (2008). At first glance, this appears surprising, as we might intuitively expect these firms to be at high risk for audit. However, as shown in **Table 3.14**, the data shows that these firms are qualitatively different than firms that pay positive wages to shareholders. As a first sanity check, only 8% of firms that do not pay a wage to shareholders report a positive amount of officer compensation on Form 1120S, compared to 93% for firms paying positive wages to shareholders. This suggests that the firms that I identify as paying zero wages to shareholders are, in fact, paying zero wages to shareholders, and that this is not a data anomaly. Furthermore, it appears that zero-wage firms are operating on a somewhat smaller scale. The median amount of assets is approximately $12,600 for firms paying zero wages to shareholders and $77,700 for firms paying positive wages to shareholders. Furthermore, only 34% of zero-wage firms have any employees, according to Form 941 (which records payroll tax payments), while 65% of firms that pay positive wages to shareholders have a non-shareholder employee. Similarly, firms that pay positive wages to their shareholders are much more likely to have positive operating income and total income. In sum, based on these statistics, it is perhaps less surprising that a large number of firms pay zero wages to shareholders. As some number of these firms paying zero wages to shareholders may be on the margin between paying and not paying, I include them in the sample.

3.4 Empirical Strategy and Main Results

3.4.1 Regression equation

To motivate the regression equation, I assume that the data generating process for wages paid to shareholders, w_{ist}, is given by the following equation, where i indexes firms, s indexes state, t indexes time, $treat_s$ is a Kansas dummy, and z_{ist} is operating income (i.e., ordinary income plus wages to shareholders).

$$\begin{aligned}\ln(w_{ist}) = \lambda_t + \mu_i + \alpha_s \times t + \delta \ln(z_{ist}) \\ + \beta_1\Big(treat_s \times \mathbb{1}(t \geq 2013)\Big) + \beta_2\Big(treat_s \times \mathbb{1}(t \geq 2014)\Big) \\ + \beta_3\Big(treat_s \times \mathbb{1}(t \geq 2015)\Big) + \epsilon_{ist} \quad (3.1)\end{aligned}$$

That is, log wages are given by firm (μ_i) and year (λ_t) fixed effects, a state-specific linear trend, plus some multiple of log operating income, plus three treatment effects. The first treatment effect, β_1 represents the immediate effect in 2013 and the second two effects, β_2 and β_3 represent the additional effects in 2014 and 2015 respectively, which could reflect a delayed response due to information or other adjustment frictions. The total (persistent) effect is equal to $\beta_1 + \beta_2 + \beta_3$, which is the primary object of interest.

However, this regression equation needs to be adjusted to account for observations where $w_{ist} = 0$ or $z_{ist} \leq 0$. This adjustment becomes more tractable after taking a first difference, as follows, where $\tilde{\lambda}_t = \lambda_t - \lambda_{t-1}$ and

$\tilde{\epsilon}_{ist} = \epsilon_{ist} - \epsilon_{is,t-1}$:

$$\ln(w_{ist}) - \ln(w_{is,t-1}) = \tilde{\lambda}_t + \alpha_s + \delta\Big(\ln(z_{ist}) - \ln(z_{is,t-1})\Big)$$
$$+ \beta_1\Big(treat_s \times \mathbb{1}(t = 2013)\Big) + \beta_2\Big(treat_s \times \mathbb{1}(t = 2014)\Big)$$
$$\beta_3 + \Big(treat_s \times \mathbb{1}(t = 2015)\Big) + \tilde{\epsilon}_{ist} \quad (3.2)$$

I approximate the left-hand-side of Equation (3.2) by its second-order Taylor approximation: the percent change at the midpoint, also known in the firm dynamics literature as the "Davis, Haltiwanger, and Schuh" (DHS) difference (Davis, Haltiwanger, and Schuh (1996), Tornqvist, Vartia, and Vartia (1985)). This functional form has the advantage of being defined when w_{ist} or $w_{is,t-1}$ is equal to zero and is defined as follows:

$$y_{ist} = \begin{cases} 0 & w_{ist} = w_{is,t-1} = 0 \\ \frac{w_{ist} - w_{is,t-1}}{0.5(w_{it} + w_{is,t-1})} & \text{else} \end{cases}$$

Unfortunately, one cannot make the same approximation to $(\ln(z_{ist}) - \ln(z_{is,t-1}))$, since operating income can take on negative values. Thus, I replace $\ln(z)$ with a modified version of the inverse hyperbolic sine, $\sinh^{-1}(z)$. The inverse hyperbolic sine of z is equal to $g(z; a) = \ln(z + \sqrt{z^2 + \exp(2a)})$, with $a = 0$. When z is positive and not close to zero, this expression is approximately $\ln(2) + \ln(z)$. This means that the difference $g(z_{ist}; a) - g(z_{is,t-1}; a)$ will approximate the log difference whenever z_{ist} and $z_{is,t-1}$ are both positive and not close to zero. The parameter a controls the size of this difference when z_{ist} and $z_{is,t-1}$ are of different "sign" (where "sign" is defined for this purpose as "positive and not near zero", "near zero", and "negative and not near zero"). A larger value of a reduces this difference.[26] I use the value $a = 9$;

[26] This is equivalent (aside from an additive constant) to the functional form discussed in

this value preserves the approximation between $g(z_{ist};a)-g(z_{is,t-1};a)$ and the log difference when z is greater than \$15,000 or so, while minimizing the noise created by changes in sign. This choice of a has little effect on the results.[27]

Additionally, some preliminary data exploration suggests that this specification can be further improved. **Figure 3.8** shows a binned scatter plot (using all observations, in all states, from 2010-2015) of y_{ist} (the DHS difference in wages to shareholders) against $g(z_{ist};9)-g(z_{is,t-1};9)$ (the change in operating income). The relationship is more appropriately described as logistic rather than linear. This is unsurprising, given that y_{ist} is bounded above and below, while $g(z_{ist};9)-g(z_{is,t-1};9)$ is not. For this reason, I will replace $\Delta z_{ist} \equiv \Big(g(z_{ist};9)-g(z_{is,t-1};9)\Big)$ with its logistic transformation, $\frac{1}{1+\exp(-\Delta z_{ist})}$.

In sum, the first empirical specification in this paper is the following regression equation, where y_{ist} is the DHS difference in wages to shareholders, and Δz_{ist} is defined as above as the difference in the modified inverse hyperbolic sine of operating income, with $a=9$:

$$\begin{aligned} y_{ist} = \tilde{\lambda}_t + \alpha_s + \delta \times \frac{1}{1+\exp(-\Delta z_{ist})} \\ + \beta_1\Big(treat_s \times \mathbb{1}(t=2013)\Big) + \beta_2\Big(treat_s \times \mathbb{1}(t=2014)\Big) \\ + \beta_3\Big(treat_s \times \mathbb{1}(t=2015)\Big) + \tilde{\epsilon}_{ist} \end{aligned} \quad (3.3)$$

I consider two sets of control firms: all non-Kansas firms, and the set of firms

Pence (2006): $\ln\left(\theta w_{it} + \sqrt{\theta^2 w_{it}^2 + 1}\right)$. This can be seen by considering $\theta = \exp(-a)$.

[27] Of course, I could use this same functional form to approximate $\ln(w_{ist}) - \ln(w_{is,t-1})$. I show these specifications in the Robustness section. While results are qualitatively similar, the choice of a does affect the precision of the results. Additionally, **Table 3.13** shows the nature of these approximations in general. It shows the value of $f(x_t, x_{t-1})$ for three types of differences (log differences, DHS differences, and IHS differences), holding x_{t-1} fixed at \$30,000, and varying x_t down the columns. We see that both the DHS difference and IHS difference are within 10 percent of the log difference when x_t is between \$15,000 (i.e., a 50 percent decrease from \$30,000) and \$90,000 (i.e., a 200 percent increase). The functions diverge when x_t approaches zero.

in the four states that border Kansas (Colorado, Missouri, Nebraska, and Oklahoma).[28]

While this specification is most general, it is natural to consider a specialization in which we impose that $\delta = 0$ by dropping the control for change in operating income. This specialization leads to a regression equation that is more transparent – a simple twoway fixed effects models (albeit where the dependent variable is already an approximate first difference). Furthermore, this specialization is computationally easier as it can be estimated on appropriately-weighted collapsed data; it will become clear in the following section that computational simplicity is an important consideration. Thus, in the main results below, I will show results from these specialized specifications as well.[29]

One additional decision needs to be made regarding when to start the sample. To illustrate the role of this decision, and to provide a visual summary of the results, I plot the average percent change in wages (i.e., average y_{ist}) over time from 2004 through 2015 in **Figure 3.1**, separately for Kansas firms and control firms. In the top panel, the set of control firms is all non-Kansas firms. In the bottom panel, the set of control firms is all firms in the four border states. In both panels, the Kansas and non-Kansas series are fairly parallel between 2010 and 2012.[30] However, the Great Recession appeared to affect Kansas firms to a smaller degree than non-Kansas firms, leading to some

[28] I estimate this model using the Stata command "reghdfe" (Correia (2016)).

[29] One could also specialize by redefining the dependent variable as the change in the wage-share of operating income and by again dropping the control for change in operating income. This effectively imposes a linear, unit elasticity of wages to shareholders with respect to operating income. As we have seen, this elasticity appears non-linear. And, as we will see, the elasticity is substantially less than one, even in the range where it is largest. Thus, this specialization would be more severe than the specialization of imposing a coefficient of zero.

[30] Note, however, that the Kansas series is persistently lower than the non-Kansas series, suggesting that a specification without state fixed effects would lead to misleading results.

non-parallel trends prior to 2010.[31] For that reason, I start the difference-in-differences regressions in 2010. Furthermore, there does appear to be a modest drop in the dependent variable in Kansas in the post-period (relative to non-Kansas firms), previewing the modestly negative regression results.

3.4.2 Inference

As discussed by Abadie, et al. (2014), the standard errors of the regression estimates in this paper are precisely zero if we think of them as purely descriptive: abstracting from data imperfections, there is zero uncertainty regarding what happened to the population-level parameters that are being estimated in these regressions – among the population of "Type A" firms under consideration – because my sample includes the entire population.[32] However, under a causal interpretation of the estimates, we must account for uncertainty regarding the counterfactual population aggregate wages to shareholders. Put differently, we will know the path of wages to shareholders in Kansas relative to control states in the aftermath of HB 2117. But we do not know what aggregate Kansas wages to shareholders *would have been* in the absence of HB 2117. This missing information creates uncertainty.

The natural approach in this setting is to use a randomization inference technique in the spirit of Fisher (1935) (see MacKinnon and Webb (2016)

[31] In results available upon request, an event study regression (omitting 2012) does not reject (using standard errors clustered by firm) that the 2011 and 2010 differences are the same as the 2012 difference. The regression *does* reject that 2009, 2010, and 2011 differences are equal to the 2012 difference. **Figure 3.9** plots the analogous series after controlling for the change in operating income. These plots do show a slight break in trend in 2012, suggesting a possible threat to the parallel trend assumption (or, perhaps, indicating a small anticipation effect). However, this trend break is not statistically significant, using standard errors clustered by firm: the event study regression does not reject that the 2011 and 2010 differences are the same as the 2012 difference.

[32] If one instead considers my sample of firms to be drawn from a hypothetical infinite population, then some sampling uncertainty is created which is well-measured by traditional standard erorrs.

and Conley and Taber (2011) for more recent treatments).[33] In broad terms, this procedure calculates the distribution of the pre-post change in counterfactual aggregate wages in each other state.[34] This distribution then serves as the estimated distribution of the uncertain change in counterfactual wages in Kansas. If the actual change in Kansas is large relative to this distribution (in particular, if the Kansas change is most positive or most negative), then we reject the null hypothesis of no effect, with size $\frac{2}{51} \approx 0.04$.[35] The key assumption of randomization inference is that (under the null) the estimated effect for Kansas has the same distribution as the estimated effects for all other states. One simple threat to this assumption is heterogeneity in sampling error: e.g., the estimated effects for states with fewer observations might be estimated more noisily. In results available upon request, I find that correcting for such heterogeneity does not change the estimated precision of the coefficient estimates.

This method can also be used to construct confidence intervals. Consider the bottom of the confidence interval, over which we pay special attention. As discussed above, the predominant source of uncertainty is over the counterfactual – i.e., the population-level average wages to shareholders that would have occurred in Kansas in the absence of HB 2117. The true effect would be more *negative* than the estimated effect only if the Kansas counterfactual were more *positive* than that of the rest of the country. Under the randomization inference method, we use the distribution of effects in other (untreated)

[33]If more states were treated, one could instead use the cluster-robust variance estimator, or CRVE (Cameron and Miller (2015)). However, the CRVE becomes degenerate in a setting such as the present setting when exactly one cluster is treated (MacKinnon and Webb (2016)).

[34]Specifically, I estimate Equation (3.3) a total of 51 times, with each estimate corresponding to a difference state (including D.C.) being coded as treated.

[35]This procedure gives essentially the same results as the procedure described in Conley and Taber (2011); the only difference is that Conley and Taber (2011) imposes the null hypothesis when estimating the distribution of placebo effects, which has an (imperceptible) effect on the estimates of the *treat* and time fixed effects.

states to place a bound (with $\frac{49}{51}$ $\approx$96% confidence) on how good the Kansas counterfactual might have been. So long as the Kansas counterfactual is no better than the most positive untreated state – an outcome to which we assign a confidence $\frac{50}{51}$ $\approx$98% – then the true treatment effect is no more negative than the estimated effect minus this "good counterfactual."[36] Thus, we calculate the bottom of the confidence interval as the Kansas estimate minus the largest non-Kansas estimate.[37] The top of the confidence interval is computed analogously.

The randomization inference procedure needs to be adjusted when the set of control states is restricted to the four states bordering Kansas. Left unchanged, the test would be close to uninformative, as the probability of the Kansas coefficient being the most positive or most negative under the null would be 0.4, far higher than conventional confidence levels. So, for such specifications, I modify the procedure such that when state j is coded as treated, the sample is restricted to state j and the states bordering that state. In this manner, I can again construct a large number of estimates (49, in this case, as Alaska and Hawaii must be dropped) and reject the null only when the Kansas estimate is the largest or smallest. Confidence intervals can be computed analogously.

[36]The confidence interval is a $\frac{49}{51}$ $\approx$96% confidence interval because there is a $\frac{1}{51}$ probability that the bounds are too low and $\frac{1}{51}$ probability that the bounds are too high.

[37]More formally, the bottom of the confidence interval is the largest value of c such that one can reject $H_0 : \hat{\beta} - c = 0$ from above. We reject H_0 from above when $\hat{\beta} - c$ is the largest coefficient estimate. Let $\hat{\beta}^{max}$ denote the largest placebo estimate. This implies that the bottom of the confidence interval is $\hat{\beta} - \hat{\beta}^{max}$. The top of the confidence interval is analogous. One undesirable feature of this approach is that the confidence interval bounds, which are simple transformations of a first and last order statistic, are themselves estimated with substantial noise. One could instead form confidence intervals parametrically, using the sample variance of the estimated placebo effect as the variance estimator for the coefficient. This reduces the noise in the estimation of the bounds of the confidence intervals at the cost of imposing an assumption of normality. Reassuringly, in unreported results, the confidence intervals estimated in this manner are broadly similar.

3.4.3 Baseline Results

Table 3.2 presents the baseline difference-in-difference estimates. In column 1 and 2, all states are included as control states. In columns 3 and 4, only border states serve as control states. In the first column, I impose the restriction that $\delta = 0$ by not including the control for the change in operating income. The dependent variable, measured at the firm-year level, is the percent change at the midpoint (also known as the DHS difference) in wages paid to shareholders. The regression is estimated from 2010 (that is, using the first difference from 2009 to 2010) to 2015 on the set of "Type A" firms – all single shareholder firms as well as those firms in which all shareholders are employees (as of $t-1$).

In particular, the top panel presents the individual coefficient estimates on $treat \times \mathbb{1}(t = 2013)$, $treat \times \mathbb{1}(t = 2014)$, and $treat \times \mathbb{1}(t = 2015)$ from Equation (3.3). I estimate that wages to shareholders fell by 1.4 percent in Kansas in 2013, an additional 0.7 percent in 2014, and an additional 1.1 percent in 2015. Using standard errors clustered by firm (in braces), one can reject the null of no effect for $treat \times \mathbb{1}(t = 2013)$ and $treat \times \mathbb{1}(t = 2015)$ (but not $treat \times \mathbb{1}(t = 2014)$). This standard error represents the uncertainty associated with the fact that the population is finite rather than infinite. However, as discussed above, the primary source of uncertainty is not this sampling error, but rather uncertainty over counterfactual population mean outcomes. Consistent with this hypothesis, the randomization inference confidence intervals, in brackets, are substantially wider and include zero. Thus, using the randomization inference method, we do not reject the null of no effect.

The bottom panel reports the sum of the coefficients on $treat \times \mathbb{1}(t = 2013)$, $treat \times \mathbb{1}(t = 2014)$, and $treat \times \mathbb{1}(t = 2015)$. This sum represents the estimated persistent effect of HB 2117 on wages paid to shareholders. The estimate of -0.032 implies that the total effect on wages paid to shareholders was -3.2 percent. As above, this total effect is significantly different than zero using standard errors clustered by firm, but is insignificant using the more conservative randomization inference procedure. The remainder of the entries in each column describe the operation of this randomization inference procedure with respect to the total effect. In particular, the Kansas total effect (the sum of the coefficients) was the 12th most negative estimate out of the 51 total estimates. The most positive placebo estimate occurred when Idaho was coded as treated; the corresponding estimate is 0.129. Similarly, the most negative placebo estimate occurred when North Dakota was coded as treated, with an estimate of -0.147. This implies that the bottom of the confidence interval is $-0.032 - 0.129 = -0.161$, as indicated in confidence interval in brackets.

The left panel of **Figure 3.2** illustrates the operation of the randomization inference method with respect to this specification. The dark line corresponds to the mean dependent variable (percent change in wages to shareholders) in Kansas in each year minus the mean dependent variable in control states in that year, normalized such that the average difference prior to 2013 is zero. The remaining lines correspond to the analogous series when each of the other 50 states are coded as treated. The dark dashed lines in particular correspond to series that represent the most negative and most positive placebo estimates for the total effect. This figure shows that Kansas was roughly following a parallel trend with its control states prior to 2013, after which point there was a small drop. Yet, this figure also confirms visually that the Kansas treatment effect is well within the distribution of placebo treatment effects, and thus is insignificantly different than zero.

The second column of **Table 3.2** shows the same estimates when the sample is restricted to Kansas and the four states that border Kansas. The point estimates in this table are quite similar to those in the first two columns. The total effect is estimated to be -0.029 in column 2 relative to -0.032.[38] While the choice of control states has little effect on the point estimates, the confidence intervals do narrow somewhat when only border states are used as controls. Consider the right panel of **Figure 3.2**. This plot is analogous to the left panel in that it plots the Kansas series in dark black, and the analogous series for other states in dashed lines. However, each series is relative to the that state's border states. For instance, the Alabama series is relative to Florida, Mississippi, Tennessee, and Georgia.[39] Comparing the two graphs shows that more of the mass in the right (border) graph is concentrated near zero; furthermore, the series corresponding to the most positive estimate – which plays a large role in determining the top of the confidence interval – has also shrunk noticeably toward zero. For this reason, the confidence intervals shrink: in column 2, one can rule out effects more negative than 10.5 percent.

In the third and fourth columns, I include the change in operating income – using the functional form described in the previous section – as a control variable. This specification is strictly more general than the specification in the previous column. This yields an estimated total effect of about -0.025 using all states as controls and -0.026 using border states as controls. Furthermore, the coefficient on the control variable is approximately one. However, recall that this control variable enters after a logistic transformation ($\frac{1}{1+\exp(-x)}$). This function has a slope of $\frac{1}{4}$ through the origin; thus, the coefficient of one means that the elasticity of wages to shareholders with respect to operating income is about $\frac{1}{4}$ for small changes in operating income.

The results in this table show that controlling for the change in operating income does not change the coefficient estimates substantially: in either case,

[38] Additionally, the results in columns 3 and 4, discussed below, are also similar.
[39] There is no Alaska series or Hawaii series in this graph.

the effect of HB 2117 appears to be small. This control variable is largely irrelevant not because it has no effect on the dependent variable (indeed, one can easily reject $\delta = 0$), but rather because it is approximately orthogonal to treatment. The first column of the top panel of **Table 3.15** shows this explicitly: operating income fell by a modest 4 percent in Kansas relative to all other states, and one percent relative to the states that border Kansas. In related specifications using a propensity-score reweighted sample and using the synthetic control method (to be discussed later), this estimate is closer to zero and sometimes switches sign.[40]

Even though the control for operating income does not affect the point estimate, it could still affect the confidence intervals by absorbing residual variation. E.g., it could be the case that the highly positive and negative placebo estimates reflect a boom or bust for the fortunes of S corporations in those placebo states more generally – which has an effect on wages to shareholders as a byproduct. This does appear to be the case, to a modest extent: the randomization confidence intervals (especially the upper bound) in columns 2 and 4 shrink somewhat toward zero relative to those in columns 1 and 3.

3.5 Refinements

3.5.1 Propensity score reweighting

The difference-in-differences specification presented in **Table 3.2** relies on a strong parallel trend assumption. In particular, we must assume that the difference in the dependent variable between Kansas and control states in the pre-period would remain the same in the post-period, but for the causal

[40]The bottom panel of **Table 3.15** shows that this orthogonality continues to hold after the logistic transformation.

effect of HB 2117 (perhaps conditional on operating income). The visual evidence in **Figure 3.1** indicates that the trends look fairly parallel between 2010 and 2012 suggesting that, perhaps, the trends could be expected to stay parallel in in 2013 and beyond under the counterfactual. However, the figure also shows that trends were markedly non-parallel prior to 2010, with Kansas firms responding differently to the Great Recession than non-Kansas firms. Additionally, summary statistics in **Table 3.1** show substantial differences in the size and profitability of firms in Kansas relative to other states, which could affect counterfactual outcomes. Thus, it would be desirable to relax the parallel trend assumption.

To make things concrete, suppose that the true data generating process under the counterfactual is given by the following interactive effects model, where κ_t and θ_s are each vectors of arbitrary length K and η_{st} and ϵ_{ist} are each mean-zero idiosyncratic error terms.[41]

$$\ln(w_{ist}) = \mu_i + \kappa_t'\theta_s + \eta_{st} + \epsilon_{ist} \tag{3.4}$$

This model is a generalization of the error structure in Equation (3.1) in that it allows a more general form of state-time specific shocks. After taking a first difference and aggregating to the state-year level, this can be written as follows, where $\tilde{\kappa}_t = \kappa_t - \kappa_{t-1}$ and u_{st} is $\eta_{st} - \eta_{s,t-1}$ plus the average within s, t of $\epsilon_{ist} - \epsilon_{is,t-1}$:

$$y_{st} = \tilde{\kappa}_t'\theta_s + u_{st} \tag{3.5}$$

Thus, each time period has a set of factors common across all states, while each state has a set of factor loadings constant over time. Under this model,

[41]More generally, one can heuristically interpret $\ln(w_{ist})$ in this equation as having been already residualized on operating income.

it is possible that a national shock arrives after HB 2117 ($\tilde{\kappa}_t$) to which Kansas reacts differently than other states due to heterogeneity in θ_s. Two-way fixed effects at the state and year level cannot absorb such time-varying effects, even after taking a first difference.

One strategy for addressing this possibility is to “guess” what the key factors θ_s are – call them θ_s^g – and make the sample look as similar as possible according to these θ_s^g. I implement this by using propensity score reweighting. In particular, I estimate one probit model per year, where the dependent variable is one when the firm is a Kansas firm, and the independent variables include year $t-1$ values of the following variables: industry dummies (at the two-digit NAICS level), dummies for the year in which the firm was first observed in the data (as a proxy for age of the firm), dummies for number of employees in six bins, a dummy for wages to shareholders being positive, a dummy for being a single-shareholder firm, total income, total assets, ordinary income, and wages to shareholders.[42] Using the results of this probit regression, I assign a weight to each firm in the year in question. The weights equal $\frac{1}{\hat{p}_{ist}}$ for those firms in Kansas and $\frac{1}{1-\hat{p}_{ist}}$ for other firms. That is, firms in Kansas receive a higher weight if their covariates make them look like non-Kansas firms (i.e., if $\hat{p}_{ist}$ is small), and vice versa.

Table 3.16 shows that this procedure is quite successful at creating a sample that matches along the targeted covariates. The first two columns show the unweighted means of the various covariates among observations in the regression sample (i.e., $t = 2010, \ldots, 2015$). Kansas firms are somewhat more likely to be in the agriculture or finance industries, to have more employees, and to be older. Furthermore, Kansas firms are noticeably larger in terms of total income, ordinary income, assets, and wages to shareholders. Propensity

[42]Total income, total assets, ordinary income, and wages paid to shareholders enter using a modified inverse hyperbolic sine function $g(x; a) = \ln(x + \sqrt{x^2 + \exp(2a)})$, with $a = 9$.

score reweighting reduces these disparities dramatically. For instance, in the unweighted sample, average log wages to shareholders in Kansas are about 0.20 (that is, 20 log points) larger than in other states. In the reweighted sample, this gap shrinks to about 0.004.[43]

The regression results are presented in **Table 3.3**.[44] They are slightly less negative than those presented in **Table 3.2** for the unweighted sample. The estimated total effects are between -0.019 and -0.020. All estimates remain insignificantly different than zero under the randomization inference method. Furthermore, the confidence interval is narrower; the lower bound of the confidence interval shrinks from -0.161 to -0.084 in column 1 and from -0.105 to -0.061 in column 3, suggesting that the inverse propensity score weighting is able to better match (placebo) treated and control states.[45] Furthermore, as in the case in **Table 3.2**, adding the control for the change in operating income does not affect the coefficient estimates, but does improve precision: the bottom of the confidence interval shrinks from -0.084 in column 1 to -0.065 in column 3, and from -0.061 in column 2 to -0.055 in column 4.

In sum, using inverse propensity score weights has very little effect on the estimates, though it does improve the precision of the estimates substantially. In the framework of the interactive effects model, the lack of an effect on the point estimates would be consistent with two explanations. First, $\tilde{\kappa}_t$ could be very similar in the post-period (i.e., 2013 through 2015) and the pre-period

[43]**Table 3.17** presents similar results when the control states are limited to the four border states.

[44]**Figure 3.10** plots the mean dependent variable over time in the propensity-score-reweighted sample, separately in Kansas and in all other states. The figure is qualitatively similar to **Figure 3.1**. Furthermore, **Figure 3.11** plots the Kansas series relative to placebo series, in a manner analogous to **Figure 3.2**. This figure shows that the dispersion of placebo estimates has reduced.

[45]To be clear, these randomization inference procedures involve repeating the entire estimation – including generating the weights – with each state s being coded as treated. E.g., when Alabama is coded as treated, the probit regressions are re-estimated with the dependent variable equal to one when the firm is an Alabama firm.

(2010 through 2012) – i.e., even if Kansas would react differently to shocks than other states, there were no new shocks to react to. Second, it could be the case that the covariates used in the propensity score estimation (θ_s^g) are poor proxies for θ_s. In any case, if the counterfactual outcome in Kansas were not described well by the parallel trend assumption, it cannot be explained by initial differences in these observed covariates.

3.5.2 Synthetic Control

The synthetic control method (Abadie, Diamond, and Hainmueller (2010)) is another strategy for estimating treatment effects when only one unit is treated. One version of the method constructs a "synthetic" control unit to match solely the evolution of the dependent variable prior to treatment. Specifically, suppose there are $N+1$ units (states) indexed by s, with the first unit treated at time T_0 and the remainder untreated. With the data aggregated to the state-year level, a set of weights v_s is chosen for the control states (with $\sum v_s = 1$ and $0 \leq v_s \leq 1 \forall s$) in order to minimize the following expression:[46]

$$\sum_{t=1}^{T_0-1} \left(y_{1t} - \sum_{s=2}^{N+1} y_{st} v_s \right)^2 \tag{3.6}$$

Another version of the synthetic control method effectively adds a constraint to the selection of the weights v_s. In particular, let z_s denote a vector of covariates for state s. In addition to the "adding up" ($\sum v_s = 1$) and non-negativity constraints ($0 \leq v_s \leq 1 \forall s$), the weights are chosen in order to minimize (3.6), subject to the constraint that the weights also minimize the following expression for some diagonal, positive definite matrix Ω, where the diagonal elements of Ω reflect the "penalty weight" given to each covariate in

[46] I.e., y_{st} is the average of y_{ist} within state s in year t, where y_{ist} is the D.H.S. difference in wages to shareholders. Because of the need to aggregate data to the state-year level, I do not attempt to control for the change in operating income in these specifications.

z:

$$(z_1 - \sum_{s=2}^{N+1} z_s v_s)' \Omega (z_1 - \sum_{s=2}^{N+1} z_s v_s) \tag{3.7}$$

Thus, the additional constraint requires that the weights not only match the pre-treatment evolution of the dependent variable, but also some predetermined set of covariates chosen by the econometrician. In practice, I include in z a full set of two-digit industry dummies, as well as the 2012 (i.e., $T_0 - 1$) values of total income, ordinary income, assets, and wages to shareholders.[47]

Under either version of the synthetic control method, the estimated treatment effect in year t is equal to the difference between y_{1t} and $\sum_{s=2}^{N+1} y_{st} v_s$ in the post-period. In practice, however, I estimate the treatment effect as a difference-in-differences (i.e., using Equation (3.3) with δ imposed to be zero, at the state-year level). As the number of pre-treatment periods approaches infinity, the difference in the pre-period will converge to zero, so these estimators are asymptotically equivalent. In finite samples, the difference-in-differences will control for any persistent level differences in the dependent variable that have not been perfectly eliminated by the synthetic control method. To perform inference, I continue to use the RI-β procedure, which is standard in the synthetic controls literature (Abadie, Diamond, and Hainmueller (2010), Abadie and Gardeazabal (2003)).

When the true data generating process is described by an interactive effects model, Abadie, Diamond, and Hainmueller (2010) show that, under standard assumptions, this method consistently estimates the treatment effect as the length of the pre-treatment period goes to infinity. Intuitively, in order for a

[47] Total income, ordinary income, assets, and wages to shareholders enter as transformed by a modified inverse hyperbolic sine function, $g(x; a) = \ln(x + \sqrt{x^2 + \exp(a)})$, with $a = 9$.

synthetic unit to match the treated unit on the pre-treatment evolution of y, it must also match on the factor loadings θ, as the idiosyncratic error terms u average out. A key feature of this method is that one need not guess the identity of covariates that affect the future path of the dependent variable. On the other hand, the modest length of the pre-treatment period in the present setting – nine years – may make the appeals to asymptotic theory fairly heroic. In other words, the weights may be non-trivially affected by idiosyncratic shocks u rather than the factor loadings θ.

Thus, the choice between propensity score reweighting and synthetic controls is a classic mean/variance tradeoff. Propensity score reweighting makes the implausible assumption that the true covariates are known, but the matching is highly precise. The synthetic control method does not require the econometrician to know the identity of the covariates – and further allows for these covariates to be unobservable – but the matching might be somewhat noisier.

Table 3.4 presents the regression results using the synthetic controls method. Column 1 does not impose the constraint in Equation (3.7) and Column 2 does so. The coefficient estimates are quite similar to the estimates using the propensity score reweighted sample (which are themselves less negative than the baseline difference-in-differences estimates). The total estimated effects are -0.014 in column 1 and -0.027, indicating only a very slight reduction in wages paid to shareholders in Kansas after HB 2117. The confidence intervals are broadly similar to those in the propensity score reweighted sample. The most positive placebo estimate in column 1 comes from California, with an estimate of 0.052. The bottom of the confidence interval is thus $-0.014-0.052 = -0.066$ in this column. Column 2 is similar: the most positive placebo estimate comes from Massachusetts, with an estimate of 0.051; the bottom of the confidence interval is -0.078.[48]

In sum, the synthetic control results echo the difference-in-differences results. They show a small negative effect after the implementation of HB 2117, but this effect is modest when compared to the distribution of placebo treatment effects in other states.

3.6 Discussion

The empirical methods so far have not uncovered a large negative effect of HB 2117 on wages paid to shareholders: using the randomization inference method, I can reject effects more negative than 6 to 8 percent in the most precise specifications. This small response stands in contrast to the findings in the literature that a substantial amount of income is shifted along this margin. One explanation for these divergent results is that the set of taxpayers affected by HB 2117 is less sophisticated than taxpayers studied in the literature. In particular, evidence put forward by Auten, Splinter, & Nelson (2016) focused on high-income S corporation owners, defined as having combined wages and S corporation income in excess of $1 million (in early-1990s dollars). Similarly, Smith, et al. (2017) examine firms that switch from C corporations to S corporations, whose owners may be more sophisticated than S corporation owners as a whole.[49]

However, I argue that there is a more fundamental potential explanation to the modest results found in this paper. This explanation relates to the sign of the counterfactual (i.e., pre-reform) tax wedge. In the literature on

[48]**Figure 3.12** considers the visual evidence regarding the match quality produced by the synthetic control method in this context. It finds that the match quality in the pre-period is "reasonable" in Kansas; some states match much better and other states match much worse. Furthermore, the figure shows that the Kansas estimate in the post-period is small relative to the distribution of placebo estimates.

[49]On the other hand, the evidence in Bull and Burnham (2008) suggests that smaller firms also engaged in such shifting.

shifting by S corporation owners, the counterfactual is no tax incentive, or a tax disincentive. For instance, the uncapping of the Medicare component of FICA (Auten, Splinter, and Nelson (2016)) introduced the shifting incentive for the first time for affected taxpayers. Similarly, in the case of firms switching from C corporation to S corporation (Smith, et al. (2017)), the tax incentive switches from being favorable to wages to disfavorable to wages.[50] Additionally, in most other shifting contexts, a reform introduces an incentive (e.g., in the case of the NIIT coming into effect, applying to passive S corporation income but not active) or changes the sign of the incentive (e.g., in the case of the reduction of the individual rate below the corporate rate in the Tax Reform Act of 1986). By contrast, in the present setting, the shifting incentive already existed prior to HB 2117 and was made stronger by that legislation. In this section, I show that standard models of tax evasion (e.g., Yitzhaki (1987), Slemrod & Yitzhaki (2002)), as adapted to the present context, can easily explain why the shifting effect may be decreasing in the size of the baseline wedge.

3.6.1 Model

I assume that firms earn some exogenous positive operating income z (before paying wages to shareholders). This amount can be allocated as wages to shareholders $w \geq 0$ and profits π, subject to $w + \pi = z$. Wages face an additional tax t that profits do not. Let w^* denote the wage that the shareholders would receive in the absence of tax incentives. Define shifting s to be the amount of wage underreporting relative to w^*; i.e., $s \equiv w^* - w$.

As in the standard models of evasion, I model this decision as a gamble.

[50] In the context of Bull and Burnham (2008), who compare multi-shareholder firms to single-shareholder firms, the relevant counterfactual is somewhat less obvious. Multi-shareholder firms (in particular, those in which at least one shareholder is not an employee) still face a shifting incentive, but this incentive is qualitatively different as it must overcome a desire to hold pre-tax allocations fixed.

In a "good" state of the world, the firm enjoys the monetary value of shifting st. However, there are two potential "bad" states of the world. I assume that the firm is audited with probability $p_A(s)$; in this state, the firm pays a fixed cost of intrusion or embarrassment, equal to F. The shifting is reversed by the IRS with unconditional probability $p_R(s)$ (or, equivalently, the probability of reversal given audit is $\frac{p_R(s)}{p_A(s)}$, assuming that audit is a pre-requisite for reversal); in this state, the firm additionally repays the amount of shifting, plus a fine equal to θ times the amount shifted. I assume that $p_R(s)$ and $p_A(s)$ are each weakly increasing and weakly convex for $s \geq 0$.[51] Thus, the firm's problem is to choose s in order to solve the following optimization problem.[52]

$$(1 - p_R(s))st + p_R(s)(-\theta st) - p_A(s)F \tag{3.8}$$

The first order expression can be written as follows (suppressing the arguments of p_R and p_A for conciseness):

$$\underbrace{t(1 - p_R(1 + \theta))}_{MB(s)} = \underbrace{st(1 + \theta)p'_R + p'_A F}_{MC(s)} \tag{3.9}$$

The expression to the left of the equals sign corresponds to the marginal benefit (MB) of shifting an extra dollar. It is equal to the expected increase in monetary payoff resulting from the marginal dollar of shifting, holding probabilities fixed. The right-hand-side corresponds to the marginal cost (MC), which comes from the increase in the probability of the "bad" states of the world.

There are two forces potentially pushing the shifting effect (that is, $\frac{ds}{dt}$) to be decreasing in t. To see this, let us consider two extreme cases. First,

[51]I further assume that the taxpayer cannot reduce $p_R(s)$ or $p_A(s)$ by reporting $w > w^*$; i.e., I assume that $p'_R(s) \leq 0$ for $s < 0$ and likewise for $p'_R(s)$.

[52]For simplicity, I assume risk neutrality, which is common in evasion models with endogenous probabilities of detection.

suppose we take the limit as $F \to 0$. In this case, when $t = 0$, the amount of shifting is zero.[53] But so long as $t > 0$, the size of the tax wedge has no effect on the amount shifted – any positive wedge will lead to the same amount of shifting.

Intuitively, the MB curve is downward sloping in s, since the expected monetary gain to shifting is decreasing in s (because the probability of reversal is increasing in s). The MC curve is upward sloping: the increase in reversal probability brought about by the marginal dollar of shifting is costlier when the stock of shifting is larger. An increase in t increases the MB curve – which makes sense, since each dollar of shifting has a higher monetary value so long as $p_R < \frac{1}{1+\theta}$. But, perhaps less obviously, an increase in t also increases the marginal cost of shifting (pivoting the curve counter-clockwise around zero). This occurs because any positive stock of shifting becomes more valuable when t increases. This causes the increased probability of reversal (brought about by the marginal dollar of shifting) to be costlier. In this extreme case, these two effects precisely offset each other. This is illustrated in the linear case in **Figure 3.3**; however, this result is more general, which can be seen by observing that t cancels out from the first order condition when $F = 0$. When $F > 0$, by contrast, shifting will generally increase as t increases ($\frac{ds}{dt} > 0$), since the associated component of the marginal cost ($p'_A(s)F$) curve does not increase when t increases. But even when $F > 0$, the pivoting of the marginal cost curve still has the effect of making the shifting effect diminish in t (i.e., $\frac{d^2s}{dt^2} < 0$).

In the other extreme case, we set $F > 0$ but $p^R = 0$. That is, the only cost of audit is a fixed cost – i.e., in this extreme case, shifting attracts extra scrutiny from the IRS, but the IRS is never successful at actually reversing the shifting.

[53] The amount of shifting at $t = 0$ and $F = 0$ is indeterminate. An amount $\epsilon > 0$ of fixed cost pins this down at zero.

In this setup, the mechanism causing a diminishing shifting effect is perhaps more intuitive. In particular, the first order condition reduces to $t = Fp'_A(s)$. Thus, the solution in this case is heavily driven by the shape of $p'_A(s)$. In **Figure 3.4**, I plot what the marginal cost curve ($Fp'_A(s)$) would need to look like in order to generate a diminishing shifting effect when $p_R(s) = 0$. In particular, the marginal cost curve must be convex – i.e., slowly increasing for shifting near zero, and more steeply increasing for larger amounts of shifting. This may be reasonable approximation of the truth. It reflects the case when there is a wide range of w that taxpayers are very confident that the IRS will deem "reasonable." As wages decrease further (i.e., as shifting increases further), taxpayers fear that the IRS will get suspicious at an increasingly fast rate. As a concrete example, suppose that taxpayers believe that the IRS will audit returns with $s > \underline{s}$ with a fixed positive probability, and will not audit firms with $s \leq \underline{s}$. But taxpayers are uncertain about the value of $\underline{s}$; they believe it to be normally distributed with some mean to the that is further to the right than the range of **Figure 3.4**. The resulting perceived probability of audit function would generate a convex marginal cost curve over this region.[54]

Figure 3.4 shows that when the tax wedge is zero, taxpayers locate at $s = 0$. When the tax wedge is increased from zero to some positive amount (given in the lower dotted line), taxpayers shift rightward substantially, to the intersection of the dotted lines at s_1. However, when the tax wedge is increased again, the response (from s_1 to s_2) is much smaller. This occurs because the first change (from zero to t_1) is enough to cause taxpayers to shift to the more steeply increasing part of the marginal cost curve. The second change (from t_1 to t_2) is enough only to cause taxpayers to shift upward along this steeply

[54]Additionally, individuals might face a psychic cost of dishonesty $h(s)$, which would play a similar role as $p_A(s)F$. The shape of the marginal cost curve in **Figure 3.4** might be a realistic depiction of the derivative of this psychic cost function, $h'(s)$.

increasing part of curve.

Thus, there are multiple forces at play – which are not mutually exclusive – causing the shifting effect to diminish with the tax wedge. Either force, or both, could potentially explain the fairly small shifting response to HB 2117 uncovered in the empirical sections above.

3.7 Heterogeneous effects

The empirical methods so far have not uncovered a large negative effect of HB 2117 on wages paid by S corporations to their shareholders. One explanation for this small effect is that the size of the existing incentive reduces the effect of a subsequent increase. If this explanation is correct, then it may also be the case that firms that faced a smaller initial wedge prior to HB 2117 may have exhibited a larger response. In this section, I test this hypothesis. In particular, the counterfactual tax wedge (i.e., FICA) is 15.3% if wages and self-employment income from all sources (hereafter "earnings") is less than the Social Security Contribution and Benefit Base ($113,700 in 2013; hereafter, the "Social Security cap"). Above this cap, it is 2.9%, plus an additional 0.9% for tax units with earnings in excess of $200,000 ($250,000 for joint filers).[55] Thus, the theory suggests that firms owned by shareholders with counterfactual earnings in excess of the Social Security cap should exhibit a stronger response. I proxy for this status at the firm level by defining a firm to be a "high-wedge" firm in year t if less than half of its shareholders have earnings greater than the Social Security cap in year $t-1$.[56] The remaining

[55]For joint filers, the 0.9% Additional Medicare Tax depends on joint earnings, while the Social Security cap applies to individual earnings. Thus, it is possible that the lower-earning spouse of a couple subject to the Additional Medicare Tax could face a marginal tax wedge of 16.2%. As the additional wedge created by the Additional Medicare Tax is small relative to the regular Social Security and Medicare taxes, I do not exploit this variation.

[56]From 2011 onward, the share of shareholders with earnings in excess of the Social Security cap is weighted by ownership. In 2010 and earlier, ownership share was not available,

firms are "low-wedge" firms. I then repeat the same three methods – baseline difference-in-differences, propensity score reweighting, and synthetic control – separately for "low-wedge" and "high-wedge" firms.

Table 3.5 presents summary statistics for firms in the pre-period (i.e., 2010-2012), separately for low-wedge and high-wedge firms. Unsurprisingly, low-wedge firms are substantially larger and more profitable. This suggests that one should be cautious in interpreting any differential response as caused solely by the initial tax wedge. In particular, as low-wedge firms are larger, they may be more sophisticated, which may have a direct effect on the size of the response to a change in tax incentives. Furthermore, the sample size of low-wedge firms is relatively small – approximately 7,800 firm-year observations in Kansas in the pre-period – which previews the relative imprecision of the regression results presented below. Additionally, as in the summary stats for the full sample presented in **Table 3.1**, Kansas firms in this sample are somewhat larger and more profitable than non-Kansas firms, though these differences are mitigated in the border sample.

Table 3.6 presents the baseline difference-in-differences results.[57] For the sake of conciseness, I report only the sum of the coefficient estimates on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$, which represents the total estimated effect. In columns 1 and 3, firms in all states are included in the regression. In columns 2 and 4, only firms from Kansas and its four bordering states are included in the regression. In columns 3 and 4, I add the control for the change in operating income (with a coefficient that is allowed to be different between high- and low-wedge firms). Across each of the four columns, the results are quite consistent. As presented in the first row, the

so all shareholders are equally weighted.

[57]**Figures 3.13** plots the evolution of the mean dependent variable in Kansas and in control states, using all states as control states. **Figure 3.14** plots the analogous series using only border states as control states.

effect is estimated to be between -0.016 and -0.029 among high-wedge firms. This is quite close to the estimates using the entire sample, as is to be expected given that such firms make up approximately 85% of the full sample.

Of note, the estimates are somewhat more negative for low-wedge firms: between -0.038 and -0.084, with slightly more negative results after including the control for the change in operating income. This result is consistent with the theory discussed in the prior section. However, these estimate are somewhat imprecise. In particular, one cannot reject equality between the coefficients on high-wedge and low-wedge firms applying randomization inference to the difference in coefficients. Furthermore, the randomization inference method cannot reject the null hypothesis that the coefficient for either type of firm is zero. On the other hand, the randomization inference confidence intervals are sufficient to rule out very large effects among each subset of firms: using the difference-in-differences confidence intervals in the border sample, one can rule out an effect more negative than 14 percentage points for high-wedge firms and 10 percentage points for low-wedge firms. These results become somewhat more pronounced after propensity-score reweighting, as shown in **Table 3.7**. Relative to **Table 3.6**, the estimates for high-wedge firms in **Table 3.7** are somewhat closer to zero, while the estimates for low-wedge firms are more negative, with the most negative result being -0.116 in column 3. Furthermore, when including the control for the change in operating income, the randomization inference procedure does in fact reject that the effects for low-wedge and high-wedge firms are the same.[58]

Results from the synthetic controls strategy, run separately for high- and low-wedge firms, are broadly similar. **Table 3.8** presents the regression results. The top row reports results for high-wedge firms and the bottom row reports

[58]Of course, this procedure makes no correction for multiple inference.

results for low-wedge firms; column 2 imposes covariate matching while column 1 does not. For the sake of conciseness, I report only the sum of the three coefficient estimates. The results are slightly less negative for both types of firms. Furthermore, the bottom of the confidence interval shrinks to -0.13 for low-wedge firms and -0.06 for high-wedge firms. Yet, the randomization inference does not allow us to rule out equality between the two types of firms at the 10% level.

In sum, I estimate modestly larger negative effects for low-wedge firms (those with initial shareholder earnings greater than the Social Security cap) than for high-wedge firms (those with initial shareholder earnings less than the Social Security cap). This is consistent with the theory presented in Section 3.6, in that the firms facing a lower counterfactual tax wedge appeared to exhibit a larger response. However, these estimates are noisy due to the modest sample size. Furthermore, the point estimates could be explained by other factors, such as differences in sophistication and salience. Thus, these results are suggestive but far from conclusive.

3.8 Other outcomes and sensitivity analyses

3.8.1 Officer compensation

In the baseline results, the dependent variable is constructed by linking firms to shareholders through Form 1120S, Schedule K-1, and additionally linking firms to employees through Form W2. An alternative approach is to use a line item from Form 1120S directly. In particular, firms are required to report "Officer Compensation," a deductible expense, on line 7 of Form 1120S. This measure is not equal to wages paid to shareholders, in general: shareholders need not be officers, and officers need not be shareholders. Nevertheless, the officer compensation line item has the advantage of not requiring any links – and thus might be subject to less measurement error. The two measures are quite closely related. Nelson (2016) has shown that the time series of officer compensation matches fairly well the time series of wages paid to S corporation shareholders over the set of years when both are available. Additionally, in my data, the two amounts tend to match each other fairly closely within a given firm. In **Figure 3.16**, I plot a histogram of the (log) difference between wages to shareholders and officer compensation for each firm, conditional on positive wages to shareholders. The leftmost bin includes only those firms for which officer compensation equals wages to shareholders exactly – of note, this bin contains just over half the sample. For an additional 22 percent of firms, officer compensation and wages to shareholders are within 10 log points of each other.

Table 3.9 presents regression results when the dependent variable is defined as the percent change in officer compensation at the midpoint.[59] The estimates are slightly more positive than the analogous estimate using calculated wages to shareholders. For instance, the estimated total effect using all states as a control variable under the baseline difference-in-differences specification (first column, first row) is 0.002, compared to -0.032 in **Table 3.2**. The most negative estimate is -0.005 and all estimates are less than or equal to 0.015 in magnitude. Furthermore, the precision improves somewhat. Thus, the results using officer compensation are further evidence of a null effect, or a small negative effect at best.[60]

[59]Throughout this section, I report only the estimated total effect for conciseness. I also do not include the control for change in operating income.

[60]An additional reassuring observation is that (with only one exception) the same states generate the most-negative and most-positive placebo estimates, compared to the specifications with wages paid to shareholders as the dependent variable.

3.8.2 Alternative functional form

The dependent variable in the primary empirical specifications is the change in a firm's wages paid to shareholders, as measured by the percent change at the midpoint, also known as the Davis-Haltiwanger-Schuh (D.H.S.) difference. One disadvantage of this strategy is that a change from \$0 to \$1 causes the dependent variable to take on the same value as a change from \$0 to \$100,000. In this section, I consider the sensitivity of this specification choice to an alternate functional form: the change in the inverse hyperbolic sine (and modifications thereof) of wages paid to shareholders, which does not suffer from this problem (Pence (2006)).

In particular, the inverse hyperbolic sine of w is equal to $\ln(w + \sqrt{w^2+1})$. For values of w sufficiently greater than zero, this is approximately equal to $\ln(w) + \ln(2)$, but unlike the logarithm, the domain of the inverse hyperbolic sine is the entire real line. In this section, I define the dependent variable as the first difference of a small modification to the inverse hyperbolic sine, as follows:[61]

$$y_{it} = g(w_{it}; a) - g(w_{i,t-1}; a) \tag{3.10}$$

$$g(w_{it}; a) = \ln\left(w_{it} + \sqrt{w_{it}^2 + \exp(2a)}\right) \tag{3.11}$$

If the shift parameter a were equal to zero, then $g(w_{it}; a)$ would be the inverse hyperbolic sine of w_{it}. When a increases, the value of $g(0; a)$ increases (to $g(0; a) = a$), while having only a small effect on $g(w'; a) - g(w; a)$ for $w, w' >$ $\$10,000$ or so. As a result, a lower a assigns a higher value (in magnitude) to y_{it} for moves to or from $w_{it} = 0$, while having little effect on changes along the

[61]Note that this is the same functional form used to construct the change in operating income z.

"intensive margin." In this section, I will present results for $0 \leq a \leq 9$.[62] I will also consider a dependent variable equal to the first difference of a dummy for paying any positive wages to shareholders. Conceptually, this corresponds to the limiting case when $a \to -\infty$.

Table 3.10 presents the estimated total effect (that is, the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$) using these alternative functional forms. The estimate in the top panel column 1 (which uses all states as control states) implies that the share of firms paying positive wages to shareholders decreased by 0.2 percentage points in Kansas after HB 2117, from a base of about 72%. This estimate is insignificantly different from zero even with standard errors clustered by firm, and remains insignificant using RI-β. The coefficient in column 1 in bottom panel, using only border states as control states, is quite similar. Thus, there is little evidence that HB 2117 affected firms along the extensive margin of paying any wages to shareholders at all. In the remainder of the columns, the dependent variable is the first difference if $g(w; a)$. Compared to the results presented in **Table 3.2**, the coefficient estimates tend to be more negative, with the largest estimate in magnitude being -0.059. However, all estimates (coefficients and confidence intervals) shrink toward zero when a increases from 0 to 9, moving rightward in the table. Given that column 1 found no effect along the extensive margin, these differences in coefficients can be attributed to the reduction of noise arising from changes along the extensive margin. With $a =$ 9, which puts the least weight on extensive margin responses, the coefficients and confidence intervals are quite close to the baseline estimates. Thus, these alternate functional forms confirm the earlier results: HB 2117 appears not

[62]For $a \leq 9$, such changes along the intensive margin (i.e., $g(w'; a) - g(w; a)$) can be interpreted as a log change for values of w and w' near the median. For $a > 9$, this interpretation starts to break down.

to have caused a substantial reduction in wages paid to shareholders of S corporations in Kansas.

3.8.3 Robustness to alternative samples

Table 3.11 presents the regression estimates using three alternative samples.[63] In the baseline, firms are in the main sample if they are a single-shareholder firm, or if all shareholders are employees (at time $t-1$). The first row restricts this further to the set of single-shareholder firms. Such firms could plausibly respond more strongly to a change in the tax wedge, as there are no coordination problems preventing adjustment. The estimates in this row are slightly less negative than the baseline results, suggesting that the inclusion of multi-shareholder firms is not causing me to miss a larger response among single-shareholder firms.

The second row uses the sample of "Type B" firms. These are the firms that face a more complicated wage-setting problem as they have at least one shareholder that earns no wages (as of time $t-1$). In such firms, wages cannot be reduced without changing pre-tax allocations; thus, one might expect a smaller shifting response. The results presented in the second row confirm that Type B firms did not exhibit a large response: all estimates are slightly less negative than the main results for Type A firms. However, estimates for neither type of firm are significantly different than zero and the estimates for the two types of firms are not significantly different from each other; thus, these results are consistent with other explanations, including a null effect for all firms, as well.

Finally, the third row drops all firms that pay zero wages to shareholders

[63] For computational reasons, I do not present RI-β confidence intervals for the propensity-score reweighted specification. Such confidence intervals would have required re-estimating separate first-stage probit regressions for each state, for each sample.

in the base-year (that is, $t-1$). These firms are likely to have a counterfactual time-t wage of zero as well, meaning that it will typically be impossible for these firms to shift in response to HB 2117. Thus, we might see a larger effect after dropping these firms. The regression results in the first two columns support this hypothesis, though the magnitude of the shifting effect remains modest and insignificantly different than zero. In particular, the estimated total effect increases to approximately 7 percent in column 1 and 5 percent in column 2. Interestingly, the estimated total effect remains quite small using the synthetic controls method (column 3). Thus, the choice to keep firms with zero base-year wages does not qualitatively alter the results.

3.8.4 Adjusting for entry and exit

It is possible that HB 2117 increased or decreased the number of S corporations in Kansas. On the one hand, HB 2117 reduced the overall tax burden on S corporations, increasing the value of starting or continuing a marginal business, potentially increasing the number of S corporations in Kansas in the post-treatment period. On the other hand, HB 2117 made S corporations slightly tax-disfavored relative to partnerships and sole proprietorships – potentially reducing the number of S corporations in the post-treatment period.[64] Such a "sample selection" effect could bias the empirical estimates which are, at their core, a comparison of means between different sets of firms that select into the sample (Heckman (1979)).

Figure 3.15 plots the year-over-year change in the number of firms in the main estimation sample (i.e., of Type A firms). The top panel of the figure does in fact show a slight decrease in the growth rate of the number of S

[64] At its onset, HB 2117 effectively exempted the entirety of economic profits from partnerships and sole proprietorships from state income tax. By contrast, HB 2117 did not exempt the wage portion of economic profits from S corporations from state income tax.

corporations starting in 2013, when all states are used as control states. This could, potentially, represent a causal effect of HB 2117, as the law was passed in May of 2012. On the other hand, the effect is much smaller (and in fact of the opposite sign) when using solely the four border states as control states, as shown in the bottom panel of **Figure 3.15**.

Using the method of Lee (2009), one can place bounds on the estimated intensive margin treatment effect to account for the estimated extensive margin response. In particular, we first estimate the excess share of firms in control states in 2013, 2014, and 2015, relative to the other years in the sample; denote this year-specific share as p_t, for $t = 2013, 2014, 2015$. To estimate the lower bound of the treatment effect, we assume that these excess control firms have the worst potential outcomes. Thus, we drop the p_{2013} firms in the control states with the worst values of the dependent variable in 2013, and likewise in 2014 and 2015.[65] Such trimming increases the average dependent variable in control states in the post-period, and thus reduces the estimated treatment effect. To estimate the upper bound of the treatment effect, the procedure is reversed; the p_{2013}, p_{2014}, and p_{2015} firms with the *most positive* values of the dependent variable are dropped in control states. If instead there is an excess share of Kansas firms, the procedure is reversed such that the most positive and most negative Kansas firms are dropped, respectively.

Table 3.12 presents the upper bound, the baseline estimate, and the lower bound of the coefficient estimate. Column 1 uses all states as control states, while column 2 uses only border states. This table shows that the bounds on the estimates are reasonably narrow. In the first column, the upper bound is 0.000, the baseline estimate is -0.032, and the lower bound is a modest -0.062. The bounds in column 2 have a similar mean but are somewhat narrower.

[65] This ranking is computed after residualizing the dependent variable on state and year fixed effects.

Thus, it does not appear that the main results are substantially affected by selection bias.

3.9 Conclusion

Federal tax law provides an incentive for owners of S corporations to classify their operating income as "business income" rather than wages. To explore the effect of increasing this incentive, I exploit variation created by a tax reform in Kansas, known as HB 2117, which exempted the profits earned by S corporations from state income tax, but not the wages received by S corporation shareholders. Using the universe of S corporation links to shareholders and employees between 2003 and 2015, I estimate several difference-in-differences models and find that wages to shareholders fell by two to three percent in Kansas in the years after HB 2117 became law. The first model assumes the standard difference-in-differences counterfactual: that the dependent variable would have followed parallel trends with control states in the absence of treatment. The second model reweights the sample of firms according to an inverse propensity score weight in order to compare similar firms in Kansas and in control states. The third model reweights the sample of *control states* in order to match the trajectory of the dependent variable prior to treatment. The most precise specifications find the bottom of the 95% confidence interval to be about -8 percent to -6 percent.

This small result is seemingly in conflict with evidence in the literature (e.g., Auten, Splinter, and Nelson (2016) and Smith, et al. (2017)) that owners of S corporations engage in a substantial amount of shifting along this margin. Furthermore, as discussed above, shifting or reclassifying income is ubiquitous in certain other contexts, such as in anticipation of tax rate changes, or in response to a reordering of the individual and corporate top rate. I put forward several explanations for why owners of S corporations in Kansas appear not to have responded as strongly to the change in incentives created by HB 2117. One simple explanation is that the response will take time as business owners learn about the tax change. Another explanation is that the typical S corporation owner may be less sophisticated than the typical taxpayer who realizes capital gains, or than the typical taxpayer on the margin between starting a C corporation or a pass-through entity.

I show that this result is not surprising in light of the standard models of tax evasion, and simple extensions thereto. Under a broad set of assumptions, these models predict that the effect of increasing a tax wedge will be diminishing in the size of the baseline tax wedge. This occurs for two reasons. First, an increase in the tax wedge means that existing shifting becomes more valuable, and taxpayers are more cautious to protect that existing shifting. Second, the shape of the marginal cost curve may be highly convex, reflecting the shape of the probability of audit function. Consistent with either of these theories, I find that firms facing a relatively small tax wedge prior to HB 2117 reduce their wages by more than other firms. Unfortunately, the precision of these estimates drops somewhat, meaning that one cannot reject the null hypothesis that both types of firms responded to the same extent; furthermore, these point estimates are consistent with other explanations (such as heterogeneous salience of the tax change) as well.

Furthermore, this result suggests that Section 199A might not lead to a large reduction in wages paid to S corporation shareholders. However, one must be careful in generalizing the effect of HB 2117 to the effect of Section 199A as enacted in the recent tax bill. On the one hand, Section 199A has additional limitations (or "guardrails") that HB 2117 did not have: it denies the deduction for service firms owned by high-income individuals, for instance.

Perhaps, this implies that the effect of HB 2117 is an upper bound for the effect of Section 199A. On the other hand, it is possible that Section 199A will be more salient than HB 2117, which could potentially lead to a larger effect. The ultimate effects of Section 199A on wages to S corporation shareholders will therefore remain an active area for further research.

Lastly, one should note that there are many other potential margins for shifting in response to HB 2117 or, more generally, a tax wedge between wage income and business income. For instance, as found by DeBacker, Heim, Ramnath, and Ross (2017), individuals can shift from an employment relationship to an independent contracting relationship. This study does not analyze the effect of HB 2117 on margins such as these. Thus, the results in this paper are not sufficient to understand the total effect of a change in the tax wedge between wage and business income. Future research should continue to study responses along these other margins as well in order to provide comprehensive evidence on this total effect.

Table 3.1: Summary statistics

	(1) Kansas	(2) Border	(3) All non-Kansas
Number of firms (rounded)	51,400	413,300	6,962,300
Share single shareholder	0.773 [38]	0.807	0.830
Mean number of shareholders	1.295 [13]	1.243	1.216
Median total assets	61,100 [14]	39,900	42,200
Median total income	189,800 [12]	144,400	138,900
Median ordinary income	20,400 [8]	17,200	11,500
Median wages to shareholders	26,500 [12]	20,800	16,000
Share with positive wages to shs.	0.726 [14]	0.704	0.626
Median operating income	55,500 [6]	45,400	36,800
Share with positive operating income	0.814 [10]	0.802	0.758

Notes: This table presents summary statistics for "Type A" firms at $t \in \{2010, 2011, 2012\}$. Type A firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees or (2) the firm has a single shareholder. Medians are rounded to the nearest $100 in order to protect taxpayer privacy. "Border states" are Colorado, Missouri, Nebraska, and Oklahoma. The number in brackets indicates Kansas's rank for the given statistic among the 51 states (including D.C.). Operating income is defined as ordinary income plus wages paid to shareholders.

Table 3.2: Main results: Baseline difference-in-differences

			Including control for Δz	
	(1) All states	(2) Border states	(3) All states	(4) Border states
treat × 2013	-0.014	-0.015	-0.013	-0.015
	{0.005}	{0.005}	{0.005}	{0.005}
	[-0.060, 0.034]	[-0.040, 0.028]	[-0.054, 0.024]	[-0.041, 0.019]
treat × 2014	-0.007	-0.007	-0.007	-0.005
	{0.005}	{0.005}	{0.005}	{0.005}
	[-0.043, 0.038]	[-0.028, 0.029]	[-0.040, 0.019]	[-0.026, 0.022]
treat × 2015	-0.011	-0.007	-0.005	-0.005
	{0.005}	{0.005}	{0.005}	{0.005}
	[-0.058, 0.056]	[-0.037, 0.059]	[-0.046, 0.019]	[-0.033, 0.020]
$\frac{1}{1+\exp(-\Delta z)}$			0.962	0.992
			{0.001}	{0.005}
Estimated total effect:	-0.032	-0.029	-0.025	-0.026
	{0.011}	{0.011}	{0.010}	{0.011}
	[-0.161, 0.115]	[-0.105, 0.106]	[-0.141, 0.051]	[-0.100, 0.040]
Rank of KS. coef	12	10	10	5
Most positive est.	ID: 0.129	ID: 0.076	ID: 0.116	ID: 0.074
Most negative est.	ND: -0.147	ND: -0.134	DC: -0.076	DC: -0.066
Observations	14,247,300	941,700	14,247,300	941,700

Notes: This table presents baseline difference-in-difference results. The dependent variable is the DHS difference (percent change at the midpoint) in wages paid to shareholders. The top panel presents the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$. The bottom panel presents the sum of those three coefficients, which represents the estimated total effect at the end of the sample period. No stars for significance are displayed. The sample includes "Type A" firms from 2010 to 2015. "Type A" firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions. In columns 1 and 3, all states (including D.C.) are used as control states. In columns 2 and 4, the sample is restricted to Kansas and its four neighboring states. In columns 3 and 4, I include a control variable, equal to the logistic transformation of the IHS difference in operating income. Standard errors in curly braces are clustered by firm. Confidence intervals in brackets are constructed using the random inference procedure. In the second panel, I present the rank of the Kansas estimated total effect among the distribution of true and placebo estimates, out of 51 in columns 1 and 3 and 49 in columns 2 and 4. I then present the most positive and most negative placebo treatment estimates.

Table 3.3: Main results: Difference-in-differences after inverse propensity score reweighting

			Including control for Δz	
	(1) All states	(2) Border states	(3) All states	(4) Border states
treat × 2013	-0.009	-0.011	-0.009	-0.011
	{0.005}	{0.006}	{0.005}	{0.006}
	[-0.035, 0.027]	[-0.028, 0.017]	[-0.029, 0.018]	[-0.029, 0.006]
treat × 2014	-0.005	-0.006	-0.006	-0.005
	{0.005}	{0.006}	{0.005}	{0.005}
	[-0.050, 0.029]	[-0.044, 0.016]	[-0.033, 0.014]	[-0.018, 0.013]
treat × 2015	-0.005	-0.003	-0.004	-0.004
	{0.005}	{0.006}	{0.005}	{0.005}
	[-0.031, 0.033]	[-0.022, 0.037]	[-0.030, 0.015]	[-0.029, 0.009]
$\frac{1}{1+\exp(-\Delta z)}$			0.882	0.954
			{0.008}	{0.008}
Estimated total effect:	-0.019	-0.020	-0.019	-0.020
	{0.011}	{0.012}	{0.011}	{0.011}
	[-0.084, 0.083]	[-0.061, 0.066]	[-0.065, 0.027]	[-0.055, 0.021]
Rank of KS. coef	16	7	8	3
Most positive est.	ID: 0.065	DE: 0.041	NV: 0.045	DE: 0.035
Most negative est.	ND: -0.102	ND: -0.086	ND: -0.046	WY: -0.042
Observations	14,247,100	941,700	14,247,100	941,700

Notes: This table presents difference-in-difference results after reweighting the sample by inverse propensity score. The dependent variable in the "second stage" regression is the DHS difference (percent change at the midpoint) in wages paid to shareholders. The top panel presents the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$. The bottom panel presents the sum of those three coefficients, which represents the estimated total effect at the end of the sample period. No stars for significance are displayed. The sample includes "Type A" firms from 2010 to 2015. "Type A" firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions. In columns 1 and 3, all states (including D.C.) are used as control states. In columns 2 and 4, the sample is restricted to Kansas and its four neighboring states. In columns 3 and 4, I include a control variable, equal to the logistic transformation of the IHS difference in operating income. Standard errors in curly braces are clustered by firm. Confidence intervals in brackets are constructed using the random inference procedure. In the second panel, I present the rank of the Kansas estimated total effect among the distribution of true and placebo estimates, out of 51 in columns 1 and 3 and 49 in columns 2 and 4. I then present the most positive and most negative placebo treatment estimates. In the first stage, propensity scores are computed by a probit regression of (separately in each year) a dummy for being a Kansas firm on industry dummies (2-digit NAICS codes), employee size dummies, firm age dummies, dummies for being single shareholder, dummies for paying positive wages to shareholders, and modified inverse hyperbolic sine transformations of total income, ordinary income, assets, and wages to shareholders.

Table 3.4: Synthetic control regression results

	(1) Match on past dep. var	(2) Match on covariates
treat × 2013	-0.009	-0.016
	[-0.028, 0.019]	[-0.031, 0.003]
treat × 2014	-0.002	-0.006
	[-0.022, 0.016]	[-0.028, 0.019]
treat × 2015	-0.004	-0.005
	[-0.028, 0.021]	[-0.024, 0.017]
Estimated total effect	-0.014	-0.027
	[-0.066, 0.052]	[-0.078, 0.032]
Rank of KS. coef	9	6
Most positive est.	CA: 0.052	MA: 0.051
Most negative est.	NM: -0.066	WY: -0.059
Weights of top 5 states:	SD: 0.579	NE: 0.243
	KY: 0.147	OK: 0.226
	WY: 0.141	CT: 0.164
	ID: 0.077	NV: 0.108
	MT: 0.055	MS: 0.101

Notes: This table presents synthetic control results. The dependent variable is the DHS difference (percent change at the midpoint) in wages paid to shareholders, aggregated to the state-year level after computing the DHS difference. The top panel presents the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$. The bottom panel presents the sum of those three coefficients, which represents the estimated total effect at the end of the sample period. No stars for significance are displayed. The underlying sample includes "Type A" firms from 2004-2015. "Type A" firms are defined as having one of the following two characteristics at time $t - 1$: (1) all shareholders are employees or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions. In column 1, the synthetic control weights are constructed solely to minimize squared error between the dependent variable in Kansas and in the synthetic control state prior from 2004 to 2012. In column 2, matching on covariates (Equation (3.7)) is additionally imposed. Confidence intervals are constructed using the random inference procedure. In the bottom panel, I present the rank of the Kansas estimate for the total effect among the distribution of true and placebo estimates, out of 51. I then present the most positive and most negative placebo treatment estimates. Finally, I present the weight of the top 5 states making up the synthetic control state.

Table 3.5: Heterogeneous effects: Summary statistics

	High-wedge firms			Low-wedge firms		
	(1) Kansas	(2) Border	(3) All non-Kansas	(4) Kansas	(5) Border	(6) All non-Kansas
Number of firms (rounded)	43,600	359,400	5,749,800	7,800	53,900	1,212,500
Share single shareholder	0.781	0.813	0.840	0.728	0.767	0.785
	[38]					
Mean number of shareholders	1.265	1.221	1.189	1.466	1.395	1.342
	[13]			[9]		
Median total assets	49,900	32,700	32,700	195,200	151,400	145,800
	[14]			[22]		
Median total income	164,600	128,100	117,900	666,400	587,700	553,400
	[8]			[22]		
Median ordinary income	18,300	15,600	10,000	55,500	50,100	30,300
	[7]			[15]		
Median wages to shareholders	24,000	18,000	12,100	126,800	125,700	123,200
	[11]			[31]		
Share with positive wages to shs.	0.724	0.700	0.613	0.736	0.731	0.688
	[14]			[20]		
Median operating income	48,300	40,300	30,700	217,400	208,400	182,700
	[6]			[18]		
Share with positive operating income	0.815	0.803	0.757	0.809	0.799	0.765
	[10]			[16]		

Notes: This table presents summary stats analogous to **Table 3.1**, separately for high-wedge and low-wedge Type A firms. High-wedge firms, presented in columns 1-3, are firms for which fewer than half of shareholders have earnings (at time $t-1$) greater than the Social Security cap. Such firms faced an initially higher tax wedge between profits and wages. Low-wedge firms, presented in columns 4-6 are firms for which at least half of shareholders meet that restriction. Such firms faced an initially lower tax wedge between profits and wages. See also the notes to **Table 3.1**.

Table 3.6: Heterogeneous effects: Baseline difference-in-differences for firms below and above Social Security cap

			Including control for Δz	
	All states	Border states	All states	Border states
High-wedge firms				
Estimated total effect:	-0.029	-0.027	-0.016	-0.016
	{0.012}	{0.012}	{0.010}	{0.011}
	[-0.155, 0.104]	[-0.098, 0.091]	[-0.133, 0.028]	[-0.090, 0.024]
Low-wedge firms				
Estimated total effect	-0.054	-0.038	-0.074	-0.084
	{0.024}	{0.025}	{0.017}	{0.018}
	[-0.186, 0.178]	[-0.141, 0.187]	[-0.192, 0.087]	[-0.157, 0.079]
P-values for tests of equality				
Clustered by firm	0.350	0.707	0.001	0.000
Using RI-β	0.549	0.653	0.745	0.653
Observations	14,247,300	941,700	14,247,300	941,700

Notes: This table presents the total estimated effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$), analogous to the bottom panel **Tables 3.2** and **3.3**, separately for high-wedge and low-wedge Type A firms. High-wedge firms, presented in the top panel, are firms for which fewer than half of shareholders have earnings (at time $t - 1$) greater than the Social Security cap. Such firms faced an initially higher tax wedge between profits and wages. Low-wedge firms, presented in columns 2 and 4 are firms for which at least half of shareholders meet that restriction. Such firms faced an initially lower tax wedge. In columns 3 and 4, I add a control for the change in operating income, whose coefficient is allowed to be different for high- and low-wedge firms. In columns 1 and 3, all states are used as control states. In columns 2 and 4, only border states are used as control states. No stars for significance are displayed. See also the notes to **Tables 3.2** and **3.3**.

Table 3.7: Heterogeneous effects: Propensity-score reweighted result for firms below and above Social Security cap

			Including control for Δz	
	All states	Border states	All states	Border states
High-wedge firms				
Estimated total effect:	-0.015	-0.019	-0.003	-0.009
	{0.013}	{0.013}	{0.012}	{0.012}
	[-0.077, 0.073]	[-0.060, 0.064]	[-0.119, 0.074]	[-0.080, 0.058]
Low-wedge firms				
Estimated total effect	-0.068	-0.045	-0.116	-0.105
	{0.026}	{0.027}	{0.020}	{0.020}
	[-0.297, 0.099]	[-0.183, 0.118]	[-0.205, -0.010]	[-0.175, 0.007]
P-values for tests of equality				
Clustered by firm	0.070	0.398	0.000	0.000
Using RI-β	0.157	0.408	0.039	0.041
Observations	14,245,700	941,700	14,245,700	941,700

Notes: This table presents estimates analogous to **Table 3.6**, using a sample that is reweighted by inverse propensity scores, as in **3.3**. The first-stage probit models are estimated separately on the samples of high- and low-wedge firms. See also the notes to **Tables 3.6** and **3.3**.

Table 3.8: Heterogeneous effects: Baseline synthetic control results for firms facing low and high counterfactual tax wedges

	(1) Match on past dep. var	(2) Match on covariates
High-wedge firms		
Estimated total effect:	-0.006	-0.002
	[-0.058, 0.058]	[-0.044, 0.065]
Weights of top 5 states:	SD: 0.482	OK: 0.311
	KY: 0.226	IA: 0.158
	MT: 0.169	MS: 0.148
	WY: 0.082	NE: 0.126
	IA: 0.035	MO: 0.119
Low-wedge firms		
Estimated total effect:	-0.057	-0.057
	[-0.129, 0.020]	[-0.128, 0.038]
Weights of top 5 states:	HI: 0.377	IA: 0.376
	AR: 0.164	WV: 0.153
	NE: 0.129	MD: 0.134
	WV: 0.124	LA: 0.116
	ND: 0.104	UT: 0.108
P-value for diff using RI-β	0.118	0.120

Notes: This table presents estimates the total estimated effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$), analogous to **Tables 3.4**, separately for high-wedge and low-wedge. High-wedge firms, presented in the top panel, are firms for which fewer than half of shareholders have earnings (at time $t-1$) greater than the Social Security cap. Low-wedge firms, presented in the bottom panel are firms for which at least half of shareholders meet that restriction. In column 1, the synthetic control weights are calculated solely to minimize squared error between the dependent variable in Kansas and in the synthetic control state from 2004 to 2012. In column 2, covariate matching (Equation (3.7)) is additionally imposed. See also the notes to **Table 3.4**.

Table 3.9: Robustness: Officer compensation instead of wages to shareholders

	(1) Baseline diff-in-diff	(2) Propensity score reweighted	(3) Synthetic control
All states as controls			
Estimated total effect:	0.002	0.013	0.003
	{0.012}	{0.012}	
	[-0.140, 0.122]	[-0.062, 0.104]	[-0.062, 0.064]
Rank of KS. coef	28	34	17
Most positive est.	ID: 0.142	ID: 0.076	CA: 0.065
Most negative est.	ND: -0.120	ND: -0.091	NM: -0.061
Observations	14,247,300	14,247,100	24
Border states			
Estimated total effect:	-0.005	0.007	
	{0.012}	{0.013}	
	[-0.101, 0.119]	[-0.049, 0.096]	
Rank of KS. coef	21	29	
Most positive est.	ID: 0.096	ID: 0.056	
Most negative est.	ND: -0.124	ND: -0.089	
Observations	941,700	941,700	

Notes: This table repeats the main analysis using a dependent variable constructed from officer compensation as reported on Line 7 of Form 1120S, rather than wages to shareholders constructed using Form W2 and Form 1120S, Schedule K-1. Specifically, the dependent variable is the the DHS difference (percent change at the midpoint) in officer compensation. In each panel, I report only the total estimated effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$). In the top panel, all states are used as control states. In the bottom panel, only the four states bordering Kansas (Colorado, Missouri, Nebraska, and Oklahoma) are used as control states. I report the baseline difference-in-differences result in column 1, the inverse propensity score reweighted result in column 2, and the synthetic control result in column 3, except that I do not run synthetic controls when the sample is restricted to Kansas and its four bordering states. Standard errors in braces in columns 1-2 are clustered by firm. Confidence intervals in brackets are calculated via random inference on the coefficient estimate. See also the notes to **3.2, 3.3** and **3.4**.

Table 3.10: Functional form robustness: alternative definition of dependent variable

	F.D. in dummy	First difference in $g(w)$			
	(1)	(2)	(3)	(4)	(5)
		$a = 0$	$a = 3$	$a = 6$	$a = 9$
All states as controls					
Estimated total effect:	-0.0022	-0.059	-0.052	-0.045	-0.033
	{0.0039}	{0.042}	{0.030}	{0.020}	{0.010}
	[-0.0401, 0.0286]	[-0.504, 0.387]	[-0.384, 0.301]	[-0.263, 0.216]	[-0.136, 0.129]
Rank of KS. coef	26	20	17	13	11
Most positive est.	ID: 0.0379	ID: 0.446	ID: 0.332	ID: 0.218	ID: 0.104
Most negative est.	ND: -0.0308	ND: -0.446	ND: -0.353	ND: -0.261	ND: -0.161
Observations	14,247,300	14,247,300	14,247,300	14,247,300	14,247,300
Border states					
Estimated total effect:	-0.0031	-0.056	-0.047	-0.037	-0.025
	{0.0042}	{0.045}	{0.032}	{0.021}	{0.010}
	[-0.0303, 0.0241]	[-0.356, 0.349]	[-0.265, 0.277]	[-0.175, 0.205]	[-0.082, 0.128]
Rank of KS. coef	15	11	11	12	10
Most positive est.	ID: 0.0272	ID: 0.300	ID: 0.218	ID: 0.137	ID: 0.057
Most negative est.	ND: -0.0272	ND: -0.406	ND: -0.324	ND: -0.242	ND: -0.153
Observations	941,700	941,700	941,700	941,700	941,700

Notes: This table presents the total estimated effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$) for alternative dependent variables. In column 1, the dependent variable is the first difference in a dummy for paying positive wages to shareholders. In the remainder of the columns, the dependent variable is a first difference in $g(w; a) = \ln(w + \sqrt{w^2 + \exp(2a)})$, where $a = 0, 3, 6, 9$ from columns 2-5, respectively. In the top panel, all states are used as control states. In the bottom panel, only Colorado, Missouri, Nebraska, and Oklahoma are used as control states. See also the notes to **Table 3.2**.

Table 3.11: Robustness to alternative samples

	(1) Baseline diff-in-diff	(2) Propensity score reweighted	(3) Synthetic control
Single-shareholder firms			
Estimated total effect: All states	-0.015	-0.015	-0.010
	{0.012}	{0.012}	
	[-0.134, 0.118]		[-0.059, 0.038]
Estimated total effect: Border states	-0.013	-0.013	
	{0.013}	{0.013}	
	[-0.088, 0.107]		
Type B firms			
Estimated total effect: All states	-0.016	0.007	-0.014
	{0.016}	{0.017}	
	[-0.203, 0.060]		[-0.063, 0.036]
Estimated total effect: Border states	-0.011	0.005	
	{0.017}	{0.018}	
	[-0.204, 0.090]		
Drop zero-wage firms			
Estimated total effect: All states	-0.068	-0.057	-0.011
	{0.013}	{0.016}	
	[-0.202, 0.117]		[-0.081, 0.068]
Estimated total effect: Border states	-0.057	-0.045	
	{0.014}	{0.017}	
	[-0.138, 0.112]		

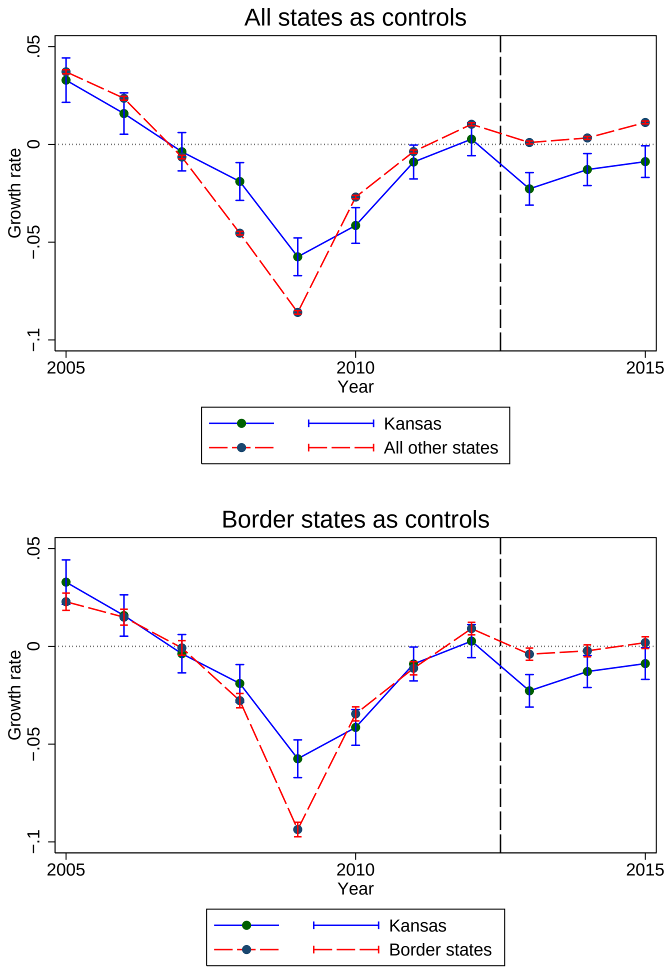
Notes: This table repeats the main analysis on different samples. In each panel, I report only the total estimated effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$). In the top panel, I restrict the sample to single-shareholder firms, meaning that I drop firms that have multiple shareholders that each receive a wage at time $t-1$. In the second panel, the sample is switched from the set of Type A firms to the set of Type B firms. Type B firms have at least two shareholders (at time $t-1$), and at least one of those shareholders did not receive a Form W2 from the firm. In the third panel, I drop all firms that pay zero wages to shareholders in the base year (time $t-1$). I report the baseline difference-in-differences result in column 1, the inverse propensity score reweighted result in column 2, and the synthetic control result in column 3. In all panels, the top row uses firms in all states, while the bottom row restricts the sample to Kansas and the four border states. I do not estimate synthetic control specifications when the sample is restricted to Kansas and the border states. Standard errors in braces in columns 1-2 are clustered by firm. Confidence intervals in brackets are calculated via random inference on the coefficient estimate. See also the notes to **3.2, 3.3** and **3.4**.

Table 3.12: Lee (2009) bounds on treatment effect

	(1) All states as control	(2) Border states as control
Upper bound	0.000	-0.012
Baseline estimate	-0.032	-0.029
Lower bound	-0.062	-0.045
Excess KS firms (2013)	-0.010	-0.003
Excess KS firms (2014)	-0.004	0.004
Excess KS firms (2015)	-0.002	-0.001

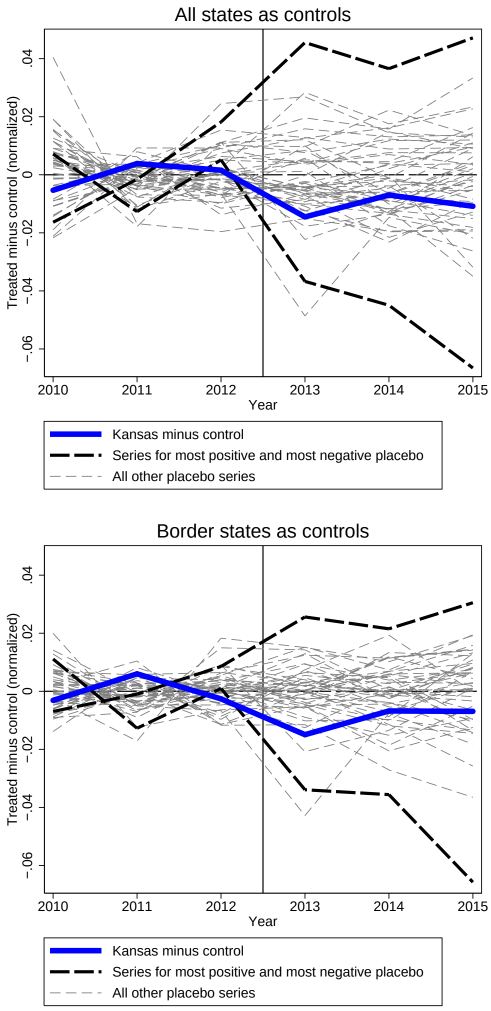
Notes: This table bounds the estimates in the bottom panel of **Table 3.2** using the method of Lee (2009). The excess mass of firms in 2013, 2014, and 2015 in Kansas is obtained by estimating the following regression (at the state-year level), where f_{st} is the number of firms in the sample in state s at time t, $treat_s$ is a dummy for Kansas and $d2013_t$, $d2014_t$, and $d2015_t$ are dummies for 2013, 2014, and 2015 respectively: $\ln(f_{st}) - \ln(f_{s,t-1}) = \alpha_s + \theta_t + \delta_{2013} treat_s \times d2013_t + \delta_{2014} treat_s \times d2014_t + \delta_{2015} treat_s \times d2015_t + u_{st}$. A negative estimate of excess Kansas firms indicates that the excess firms are in control states. When the excess is positive, the lower bound is computed by trimming the top δ_{2013} firms in Kansas in 2013, the top $\delta_{2013} + \delta_{2014}$ firms in Kansas in 2014, and the top $\delta_{2013} + \delta_{2014} + \delta_{2015}$ firms in Kansas in 2015, where the ranking is measured by the residualized dependent variable. When the excess is negative, the lower bound is computed by trimming the bottom $|\delta_{2013}|$, $|\delta_{2013}+\delta_{2014}|$, $|\delta_{2013}+\delta_{2014}+\delta_{2015}|$ firms in control states in 2013, 2014, and 2015 respectively. The upper bound is computed analogously.

Figure 3.1: Evolution of baseline dependent variable (DHS difference in wages paid to shareholders) in Kansas and control states, 2004-2015



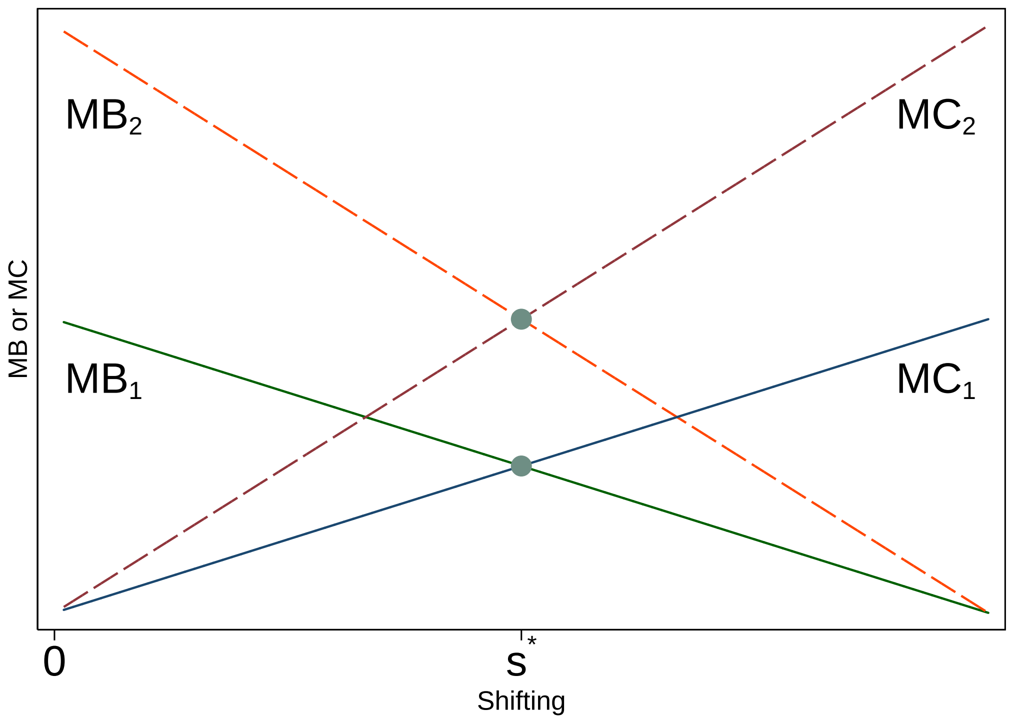
Notes: This figure plots the evolution of the mean dependent variable y_{ist}, the DHS difference (percent change at the midpoint) of wages paid to shareholders, in Kansas and in control states, controlling for the change in operating income. In particular, these are the coefficients from a regression of y_{ist} on a full set of control-by-year and treat-by-year dummies. The sample includes "Type A" firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. In the top panel, all states are used as control states. In the bottom panel, only border states (Colorado, Missouri, Nebraska, and Oklahoma) are used as control states. See text for additional "data quality" sample restrictions.

Figure 3.2: True treatment effect relative to distribution of placebo treatment effects: baseline



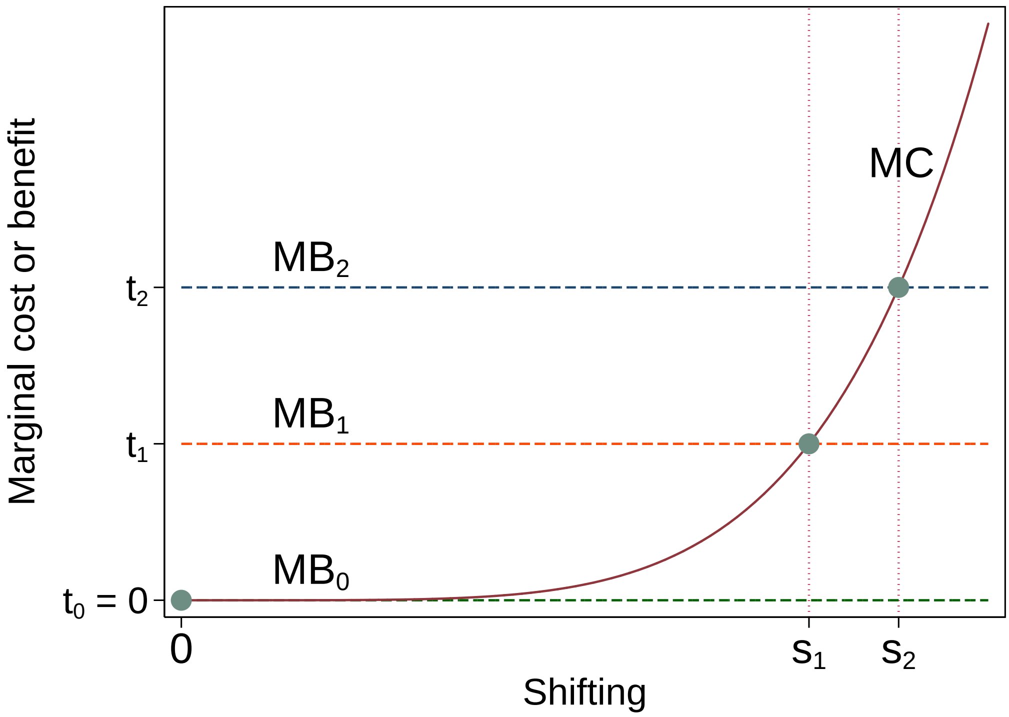
Notes: This figure plots the Kansas treatment effect relative to placebo treatment effects. The thick black line plots the mean dependent variable (the percent change at the midpoint of wages paid to shareholders) in Kansas relative to the mean dependent variable in control states, normalized such that this difference equals zero, on average, in the pre-period (2010-2012). The remaining series plot the mean dependent variable in each state j relative to the mean dependent variable in all other states, again normalized to equal zero in the pre-period. The darkest dashed lines represent the series corresponding to the most negative and most positive placebo treatment estimates. The top panel uses all states as control states. The bottom panel uses only border states (i.e., the set of states bordering placebo-treated state j) as control states. The sample includes Type A firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional “data quality” sample restrictions.

Figure 3.3: Marginal cost and marginal benefits of shifting: no fixed cost



Notes: This figure illustrates two potential solutions to the model presented in section 3.6. It plots two sets of marginal cost and marginal benefit curves. The solid lines correspond to the curves when the tax wedge is t_1. The dotted lines correspond to the curves when the tax wedge is $t_2 > t_1$. The solution is given by the intersection of the two curves, which is given by s^* in both cases. See the text for further discussion.

Figure 3.4: Marginal cost and marginal benefits of shifting: zero probability of reversal



Notes: This figure illustrates three potential solutions to the model presented in section 3.6. It plots a possible marginal cost curve against three candidate marginal benefit curves, MB_0, MB_1, and MB_2. With no tax wedge ($t = MB_0 = 0$), the optimal amount of shifting is zero. When the tax wedge increases to $t_1 = MB_1$, optimal shifting is s_1. When the tax wedge increases further to $t_2 = MB_2$, optimal shifting increases, but only slightly, to s_2. See the text for further discussion.

3.10 Appendix

3.10.1 Construction of Data

I construct the data as follows. First, I retrieve from the Business Return Transaction File (BRTF) all Forms 1120S filed for tax years 2003 through 2015, except for (1) those that are marked as "Amended, Corrected, Supplemental, Tentative or Revised" according to the computer condition code or (2) those with tax identification numbers (TINs) coded as "unmatchable."[66] Second, I retrieve from the Information Returns Master File (IRMF) all Forms 1120S, Schedule K-1 associated with these (payer) TINs and tax years. If multiple Schedules K-1 are found for the same combination of payer TIN (typically EIN), shareholder TIN (typically SSN), and tax year, I keep the one which was entered into the database most recently. Additionally, I use Form 1041, Schedule K-1 to identify the TINs of owners of trusts that are themselves shareholders of S corporations (indicated on Form 1120S, Schedule K-1).[67] Let L_{SSN} denote the set of shareholder, tax-year combinations identified by Form 1120S, Schedule K-1, supplemented by Form 1041, Schedule K-1. Let $L_{SSN,EIN}$ denote the set of recipient TIN, payer TIN, tax-year combinations identified by Form 1120S, Schedule K-1, supplemented by by Form 1041, Schedule K-1.

I also retrieve certain Forms W2 from the IRMF. In particular, I retrieve all Forms W2 for the shareholder TIN, tax year combinations in L_{SSN}. As with Form 1120S, Schedule K-1, I only keep the most recently entered Form W2 in cases in which there are multiple Forms W2 for a given payer TIN, recipient TIN, tax year combination.[68] I identify the wages paid to shareholders

[66]All data was retrieved in August 2017.

[67]I do not attempt to find the beneficial owners of trusts identified as having a SSN as their Tax Identification Number type.

[68]Note that the payer TIN for the Form W2 need not correspond to the 1120S payer TIN, in general.

by restricting these Forms W2 to those payer TIN, recipient TIN, tax year combinations in $L_{SSN,EIN}$.

Additionally, for the sake of calculating shareholders' earnings relative to the Social Security cap, I retrieve from the Individual Return Transaction File (IRTF) all Forms 1040, Schedule SE for shareholder TIN, tax year combinations identified by Form 1120S, Schedule K-1. As before, in the (unusual) case of duplicates, I keep the most recently-entered form.[69] Total earnings are calculated by summing up self-employment income (which I define as line 4a of Part I of Section B of Form 1040, Schedule SE) and W2 wages. Lastly, I retrieve industry from the BRTF entity file and the number of employees from Form 941.[70]

I perform several additional data cleaning steps. These steps are: (1) dropping firm-year observations with missing Form 1120S, Schedule K-1, (2) dropping firm-year observations in which there are no shareholders from the same state (as identified by Schedule K-1) as the firm (as identified by Form 1120S), (3) dropping firm-year observations whose income does not match between Form 1120S and its Schedule K-1, (4) dropping firm-year observations in which the number of shareholders from Form 1120S does not match the number of associated Schedules K-1, (5) dropping firm-year observations in which wage expenses or officer compensation is reported on Form 1120S, but no Forms W2 are found, and (6) dropping firm-year observations in which one shareholder receives positive income and another receives negative income.[71]

[69] For Form 1040, Schedule SE, this corresponds to the cycle posting date.

[70] Form 941 must be filed quarterly, in general. I define employment as the average of the four quarters within a given tax year. To deal with a common data entry error in which firms mistakenly write the compensation amount on the employment line, I drop firms (i.e., treat employment as zero) when total compensation averages less than \$500 per employee.

[71] For step (3), I allow a mismatch of up to 5% of income. Additionally, while the tax year on the Form 1120S should correspond to the tax year on the Schedule K-1, I allow for misreporting in this respect – i.e., I do not drop firms if the income from Form 1120S in year t to match the income from Schedule K-1 for year $t + 1$. For step (6), I drop such firms because income must be allocated on a pro-rata basis; thus, a situation in which one

Figure 3.17 plots the share of raw observations dropped in each year at each step. The upper-left panel shows step (1); the upper-right panel shows the effect of step (2) on the subset of firms remaining after step (1); and so forth. Each panel of the figure shows that the trends in Kansas mostly mirror the trends in other states, which is reassuring for the difference-in-differences empirical strategies. Additionally, there are visible spikes in the share dropped due to missing Schedules K-1 in 2005 and (to a lesser extent) 2010. Income matching improves dramatically after 2005, perhaps because the coding of duplicate Forms 1120S improved starting in 2005. The share dropped due to a mismatch in the number of shareholders also improves over time, falling from about 6% in the beginning of the sample to less than 1% by the end. The share dropped due to the presence of a wage or officer compensation deduction without the accompanying Forms W2 is fairly stable between 4% and 6%.

In **Table 3.19**, I report summary statistics for the set of firms that are dropped due to data quality, relative to the firms that are not, using observations from 2010-2015.[72] By construction, we generally must restrict ourselves to variables that are found on the 1120S. The table shows that dropped firms tend to be smaller, in terms of total income, ordinary income, assets, and officer compensation. However, these differences are relatively modest. Of course, we must interpret statistics for the dropped firms with caution, as the data from these firms may be less reliable.

An alternative approach to linking S corporation to their shareholders would be to use the so-called Document Locator Number (DLN), as Cooper, et al. (2016) do. The DLN is an identifier that uniquely links a tax filing;

shareholder receives positive income and another receives negative should never occur. The one exception is in the shareholder's final year of ownership, in which a "closing-of-the-books" election could produce such a result. Thus, I do not apply step (6) to firm-year observations which comprise a shareholder's final year of ownership of a given firm.

[72]In this table, I do not require firms to be "Type A" firms, nor do I require firms to be present at time $t-1$ in order to remain in the sample at time t.

in principle, this should link Form 1120S to its Schedule K-1, as they are required to be included in the same filing. Unfortunately, the DLN match performs quite poorly prior to 2007. In particular, the number of Schedules K-1 matched to Form 1120S using the DLN more than doubles between 2006 and 2007; the number of Schedules K-1 matched according to payer TIN and tax year exhibits no such break. For this reason, I use the payer TIN and tax year match throughout the paper, subject to the data quality selection described above.

3.10.2 Health insurance included in Box 1 wages

As a general rule, individuals who are 2% owners (or more) of an S corporation must include in Box 1 W2 wages the health insurance premiums paid (or reimbursed) by the firm. These premiums are not subject to FICA and, in general, the shareholder will be entitled to an above-the-line deduction for these amounts (the self-employed health insurance deduction). Thus, in general, the shareholder will arrive at the same outcome as the usual rule for employer health insurance premiums: exclusion from income for income tax purposes and FICA purposes.

Interestingly, HB 2117 repealed the self-employed health insurance deduction for the purpose of calculating Kansas income. This meant that, after HB 2117, owners of S corporations were effectively required to include health insurance premiums in their Kansas income. Thus, the tax favorability of health insurance premiums fell between the pre-period (when they were not taxed in Kansas) and the post-period (when they were taxed in Kansas). Conceivably, this change could lead to a response: shifting away from health insurance premiums (e.g., by reducing its actuarial value) toward true wages or toward profits. This decision is likely fairly inelastic, perhaps due to behavioral frictions (e.g., Handel (2013)). But, since shifting from health insurance toward profits would be observationally equivalent to shifting from wages toward profits, this mechanism would tend to bias my estimates downward: I would be capturing shifting along a separate margin beyond the margin under study. Given that I estimate a small negative effect to begin with, this downward bias does not threaten the interpretation of the results – that the effect of HB 2117 on wages to shareholders was small at best.

Nevertheless, it would be ideal to construct a definition of wages that does not include health insurance premiums. There are two candidate strategies for doing so. However, both have substantial drawbacks, and thus I do not use either of these strategies. First, one could make use of the fact that the health insurance premiums are not included in Medicare wages as recorded on the W2. Thus, one could define "wages" as equal to Medicare wages less retirement contributions (such as to 401(k) accounts).[73] However, the Medicare wages field in the source data appears to be substantially less clean than the (Box 1) wages field. In the underlying data, I performed an auxiliary analysis on the subset of W2s where Medicare wages are not equal to Box 1 wages (and which do not receive a Form 1065, Schedule K-1 or Form 1120S, Schedule K-1). I find that this difference is not equal to retirement contributions for 27 percent of firms – which is implausibly high, given the relatively narrow circumstances (other than retirement contributions) which would make Medicare wages not equal wages.[74]

Second, one could make use of the self-employed health insurance deduction field on the 1040. However, taxpayers do not report which business the health

[73] We would want to subtract retirement contributions – which are included in Medicare wages – because the correct concept of "wages" is the amount of true wages that is included in Kansas taxable income.

[74] Such circumstances include W2s received by nonresident aliens and by students employed by their school.

insurance deduction is attributable to. E.g., if a shareholder (and/or his or her spouse) owns more than one business (S corporation or otherwise), then I would not know to which firm to attribute the health insurance deduction. In the data, I find that about 60 percent of the shareholders in the main sample that claim a self-employed health insurance deduction have this problem: they (or their spouse) have some other source of business income. For both of these reasons, I use Box 1 wages as my definition of wages despite the imperfection of doing so.

Table 3.13: Functional form approxmation: Log difference, DHS difference, and I.H.S. difference, starting from $x_{t-1} = 30,000$

	(1) Log difference	(2) D.H.S. difference	(3) I.H.S. difference (a=9)
$-30000	.	.	-4.040
$-1	.	.	-2.020
$1	-10.309	-2.000	-2.020
$1000	-3.401	-1.871	-1.897
$15000	-0.693	-0.667	-0.645
$25000	-0.182	-0.182	-0.175
$30000	0.000	0.000	0.000
$37500	0.223	0.222	0.217
$45000	0.405	0.400	0.396
$60000	0.693	0.667	0.680
$90000	1.099	1.000	1.083

Notes: This table shows values of the difference $f(x_t, x_{t-1})$ for three types of differences, holding x_{t-1} fixed at $30,000, and varying x_t. In the first column, $f(x_t, x_{t-1}) = \ln(x_t) - \ln(x_{t-1})$. In the second column, $f(x_t, x_{t-1})$ is given by the Davis, Haltiwanger, Schuh (DHS) difference, $f(x_t, x_{t-1}) = \frac{x_t - x_{t-1}}{0.5*(x_t + x_{t-1})}$. In the third column, $f(x_t, x_{t-1})$ is given by the difference in the modified inverse hyperbolic sine: $f(x_t, x_{t-1}) = g(x_t; a) - g(x_{t-1}; a)$, where $g(x, a) = \ln(x + \sqrt{x^2 + \exp(2a)})$, and $a = 9$. I do not allow the DHS difference to be defined when x_t or x_{t-1} is less than zero.

Table 3.14: Statistics of firms that do not pay a wage to shareholders

	(1) Zero wage firms	(2) Positive wage firms
Claims officer comp. deduction	0.084	0.927
Median assets	12,600	77,700
Has a non-shareholder employee	0.343	0.653
Has positive operating income	0.530	0.942
Has zero operating income	0.051	0.000
Has positive total income (line 6)	0.842	0.992
Has zero total income (line 6)	0.130	0.005

Notes: This table presents summary statistics for firms Type A firms, 2010-2015, that pay zero wages to shareholders (column 1) and firms that pay positive wages to shareholders (Column 2). Type A firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees or (2) the firm has a single shareholder. See text for additional "data quality" restrictions. "Claims officer comp. deduction" is a dummy indicating whether the firm reports a positive amount of officer compensation on Form 1120S. Operating income is defined as ordinary income plus wages to shareholders. Median assets is rounded to the nearest hundred in order to protect taxpayer confidentiality.

Table 3.15: Effect of HB 2117 on operating income

	(1) Baseline diff-in-diff	(2) Propensity score reweighted	(3) Synthetic control
Linear: Δz			
Estimated total effect: All states	-0.042	0.010	-0.005
	{0.023}	{0.025}	
	[-0.148, 0.485]	[-0.170, 0.383]	[-0.122, 0.095]
Estimated total effect: Border states	-0.010	0.015	
	{0.024}	{0.023}	
	[-0.121, 0.460]	[-0.071, 0.391]	
Logistic: $\frac{1}{1+\exp(\Delta z)}$			
Estimated total effect: All states	-0.007	0.000	0.005
	{0.003}	{0.003}	
	[-0.030, 0.082]	[-0.025, 0.064]	[-0.012, 0.031]
Estimated total effect: Border states	-0.003	0.001	
	{0.003}	{0.003}	
	[-0.020, 0.079]	[-0.016, 0.067]	

Notes: This table repeats the main analysis for the change in operating income (ordinary income plus wages to shareholders). In the top panel, the dependent variable is Δz, the IHS difference in operating income (the first difference of $\ln(z + \sqrt{z^2 + \exp(2 \times 9)})$). In the second panel, the dependent variable is the logistic transformation of Δz, i.e., $\frac{1}{1+\exp(\Delta z)}$; this functional form is used in the main specifications as a control variable. The three columns report the baseline difference-in-differences estimate for the total effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$), the inverse propensity score reweighted result, and the synthetic control result. In all panels, all states are used as control states in the top row, and only border states are used as control states in the bottom row. I do not estimate synthetic control specifications using only border states as controls. Standard errors in braces in columns 1-2 are clustered by firm. Confidence intervals in brackets are calculated via random inference on the coefficient estimate. See also the notes to **3.2, 3.3** and **3.4**.

Table 3.16: Balance of covariates after propensity score reweighting (all states)

	Unweighted means		Propensity score reweighted	
	(1) Kansas	(2) Other states	(3) Kansas	(4) Other states
Industries				
Agriculture, etc.	0.044	0.017	0.017	0.017
Mining, Oil & Gas	0.021	0.005	0.005	0.005
Utilities	0.001	0.001	0.001	0.001
Construction	0.140	0.138	0.131	0.132
Manufacturing	0.038	0.037	0.037	0.036
Wholesale Trade	0.036	0.046	0.045	0.045
Retail Trade	0.102	0.103	0.103	0.101
Transp. & Warehousing	0.032	0.032	0.033	0.033
Information	0.012	0.016	0.015	0.016
Finance	0.060	0.041	0.040	0.041
Real Estate	0.082	0.091	0.092	0.093
Professional Svcs	0.169	0.180	0.182	0.183
Mgmt of Companies	0.003	0.003	0.003	0.003
Admin & Support	0.035	0.037	0.038	0.038
Education	0.004	0.007	0.007	0.007
Health Care	0.095	0.095	0.100	0.101
Arts & Entertainment	0.011	0.021	0.022	0.022
Accom. & Food	0.036	0.043	0.045	0.044
Other Services	0.073	0.079	0.080	0.080
Public Admin.	0.000	0.000	0.000	0.000
Employee Bins				
No employees	0.273	0.332	0.328	0.335
Empl. bin 1 (+)	0.165	0.175	0.179	0.176
Empl. bin 2 (+)	0.104	0.101	0.100	0.099
Empl. bin 3 (+)	0.156	0.137	0.136	0.136
Empl. bin 4 (+)	0.155	0.132	0.133	0.131
Empl. bin 5 (+)	0.147	0.123	0.125	0.123
Date of Entry Dummies				
First observed in 2003	0.627	0.595	0.483	0.477
First observed in 2004	0.061	0.067	0.059	0.058
First observed in 2005	0.059	0.062	0.059	0.058
First observed in 2006	0.051	0.055	0.059	0.059
First observed in 2007	0.049	0.051	0.062	0.063
First observed in 2008	0.038	0.042	0.060	0.061
First observed in 2009	0.030	0.033	0.057	0.058
First observed in 2010	0.025	0.029	0.050	0.050
First observed in 2011	0.023	0.025	0.042	0.043
First observed in 2012	0.017	0.020	0.034	0.035
First observed in 2013	0.014	0.014	0.024	0.025
First observed in 2014	0.006	0.008	0.013	0.014
Other variables				
Log total income	12.581	12.326	12.285	12.294
Log ordinary income	10.413	10.112	10.104	10.137
Log assets	11.701	11.465	11.427	11.427
Log wages to shareholders	10.731	10.536	10.518	10.526
Single shareholder	0.769	0.823	0.832	0.831
Dummy for pos. wages to shs.	0.732	0.638	0.626	0.629

Notes: This table presents mean values of covariates in Kansas and all other states among Type A firms from 2010 to 2015. Type A firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees or (2) the firm has a single shareholder. All covariates are measured at time $t-1$. The first two columns report the means for the unweighted sample. Columns 3 and 4 report the means for the inverse-propensity-score-weighted sample. The propensity scores are estimated via year-specific probit regressions that include the displayed covariates as regressors (excluding redundant covariates). In the table, "log" is a shorthand for $\ln(y + \sqrt{y^2 + \exp(2 \times 9)}$. See text for additional sample restrictions.

Table 3.17: Balance of covariates after propensity score reweighting (border states)

	Unweighted means		Propensity score reweighted	
	(1) Kansas	(2) Other states	(3) Kansas	(4) Other states
Industries				
Agriculture, etc.	0.044	0.024	0.027	0.027
Mining, Oil & Gas	0.021	0.015	0.017	0.017
Utilities	0.001	0.001	0.001	0.001
Construction	0.140	0.153	0.145	0.146
Manufacturing	0.038	0.034	0.033	0.033
Wholesale Trade	0.036	0.033	0.033	0.033
Retail Trade	0.102	0.094	0.093	0.092
Transp. & Warehousing	0.032	0.028	0.029	0.029
Information	0.012	0.013	0.013	0.013
Finance	0.060	0.051	0.052	0.052
Real Estate	0.082	0.085	0.086	0.086
Professional Svcs	0.169	0.186	0.187	0.188
Mgmt of Companies	0.003	0.002	0.002	0.002
Admin & Support	0.035	0.038	0.038	0.038
Education	0.004	0.005	0.006	0.005
Health Care	0.095	0.096	0.101	0.101
Arts & Entertainment	0.011	0.014	0.014	0.014
Accom. & Food	0.036	0.040	0.040	0.040
Other Services	0.073	0.078	0.079	0.079
Public Admin.	0.000	0.000	0.000	0.000
Employee Bins				
No employees	0.273	0.282	0.279	0.281
Empl. bin 1 (+)	0.165	0.212	0.209	0.209
Empl. bin 2 (+)	0.104	0.109	0.107	0.107
Empl. bin 3 (+)	0.156	0.143	0.143	0.143
Empl. bin 4 (+)	0.155	0.136	0.138	0.138
Empl. bin 5 (+)	0.147	0.119	0.123	0.123
Date of Entry Dummies				
First observed in 2003	0.627	0.604	0.491	0.490
First observed in 2004	0.061	0.065	0.057	0.057
First observed in 2005	0.059	0.061	0.058	0.058
First observed in 2006	0.051	0.054	0.058	0.058
First observed in 2007	0.049	0.050	0.062	0.063
First observed in 2008	0.038	0.042	0.060	0.060
First observed in 2009	0.030	0.033	0.057	0.057
First observed in 2010	0.025	0.028	0.048	0.048
First observed in 2011	0.023	0.023	0.040	0.040
First observed in 2012	0.017	0.019	0.033	0.033
First observed in 2013	0.014	0.013	0.023	0.024
First observed in 2014	0.006	0.008	0.013	0.013
Other variables				
Log total income	12.581	12.386	12.403	12.403
Log ordinary income	10.413	10.309	10.350	10.358
Log assets	11.701	11.404	11.422	11.420
Log wages to shareholders	10.731	10.604	10.636	10.634
Single shareholder	0.769	0.800	0.804	0.804
Dummy for pos. wages to shs.	0.732	0.714	0.709	0.710

Notes: This table presents mean values of covariates in Kansas and its four bordering states (Colorado, Missouri, Nebraska, and Oklahoma) among Type A firms from 2010 to 2015. Type A firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees or (2) the firm has a single shareholder. All covariates are measured at time $t-1$. The first two columns report the means for the unweighted sample. Columns 3 and 4 report the means for the inverse-propensity-score-weighted sample. The propensity scores are estimated via year-specific probit regressions that include the displayed covariates as regressors (excluding redundant covariates). In the table, "log" is a shorthand for $\ln(y + \sqrt{y^2 + \exp(2 \times 9)})$. See text for additional sample restrictions.

Table 3.18: Functional form robustness: alternative definition of dependent variable after propensity score reweighting

	F.D. in dummy	First difference in $g(w)$			
	(1)	(2)	(3)	(4)	(5)
		$a = 0$	$a = 3$	$a = 6$	$a = 9$
All states as controls					
Estimated total effect:	0.0019	-0.016	-0.021	-0.027	-0.025
	{0.0044}	{0.047}	{0.034}	{0.021}	{0.010}
	[-0.0200, 0.0198]	[-0.263, 0.254]	[-0.203, 0.195]	[-0.143, 0.136]	[-0.079, 0.081]
Rank of KS. coef	32	25	23	15	10
Most positive est.	ID: 0.0219	ID: 0.248	ID: 0.182	ID: 0.116	NV: 0.054
Most negative est.	ND: -0.0179	ND: -0.270	ND: -0.216	ND: -0.163	ND: -0.106
Observations	14,247,100	14,247,100	14,247,100	14,247,100	14,247,100
Border states					
Estimated total effect:	0.0001	-0.025	-0.025	-0.026	-0.021
	{0.0046}	{0.049}	{0.035}	{0.022}	{0.011}
	[-0.0165, 0.0152]	[-0.199, 0.196]	[-0.150, 0.156]	[-0.112, 0.117]	[-0.072, 0.079]
Rank of KS. coef	25	15	14	10	7
Most positive est.	ID: 0.0166	ID: 0.174	ID: 0.124	DE: 0.087	DE: 0.052
Most negative est.	DC: -0.0151	ND: -0.221	ND: -0.182	ND: -0.143	ND: -0.099
Observations	941,700	941,700	941,700	941,700	941,700

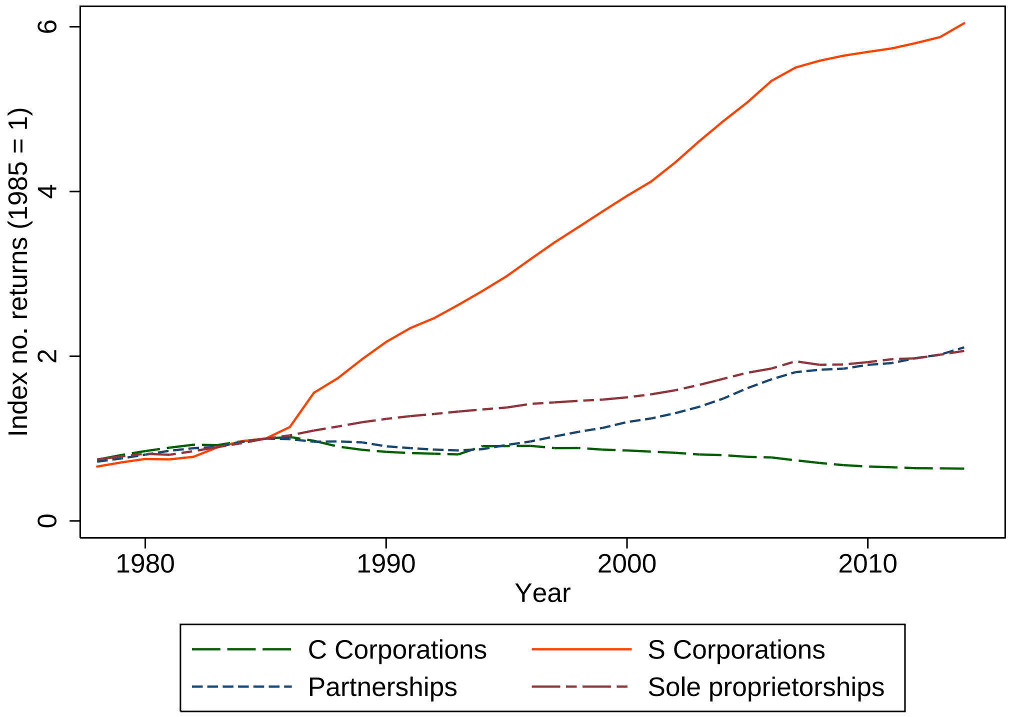
Notes: This table presents the total estimated effect (the sum of the coefficients on $treat \times 1(t = 2013)$, $treat \times 1(t = 2014)$, and $treat \times 1(t = 2015)$) for alternative dependent variables, after reweighting the sample by inverse propensity scores. In column 1, the dependent variable is the first difference in a dummy for paying positive wages to shareholders. In the remainder of the columns, the dependent variable is a first difference in $g(w; a) = \ln(w + \sqrt{w^2 + \exp(2a)})$, where $a = 0, 3, 6, 9$ from columns 2-5, respectively. In the top panel, all states are used as control states. In the bottom panel, only Colorado, Missouri, Nebraska, and Oklahoma are used as control states. See also the notes to **Table 3.3**.

Table 3.19: Characteristics of firms that are dropped due to data quality

	(1) Not dropped	(2) Dropped
Total assets		
25th percentile	1,500	0
50th percentile	39,700	23,000
75th percentile	202,700	195,600
Ordinary income		
25th percentile	-1,200	-1,400
50th percentile	7,700	5,600
75th percentile	48,400	43,000
Total income		
25th percentile	16,400	18,500
50th percentile	113,300	101,100
75th percentile	358,300	333,000
Officer compensation		
25th percentile	0	0
50th percentile	2,000	1,500
75th percentile	48,000	32,900

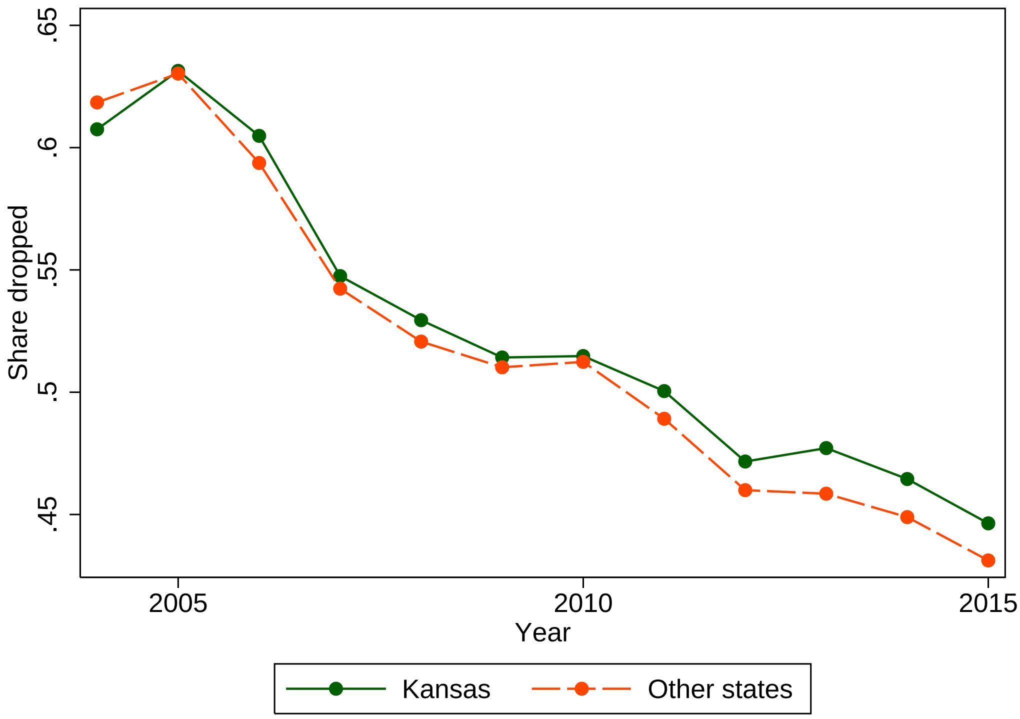
Notes: This table shows characteristics of firms that are dropped due to data quality (second column) relative to firms that are not dropped due to data quality. Restrictions related to balance (i.e., requiring firms to be present at time $t-1$ and t) and shareholder-employees (i.e., requiring firms to be "Type A" firms) are not imposed here. This data includes observations from 2010-2015. See text for a discussion of data quality sample restrictions.

Figure 3.5: Evolution of choice of business entity in the United States: C corporations, S corporations, partnerships, and sole proprietorships



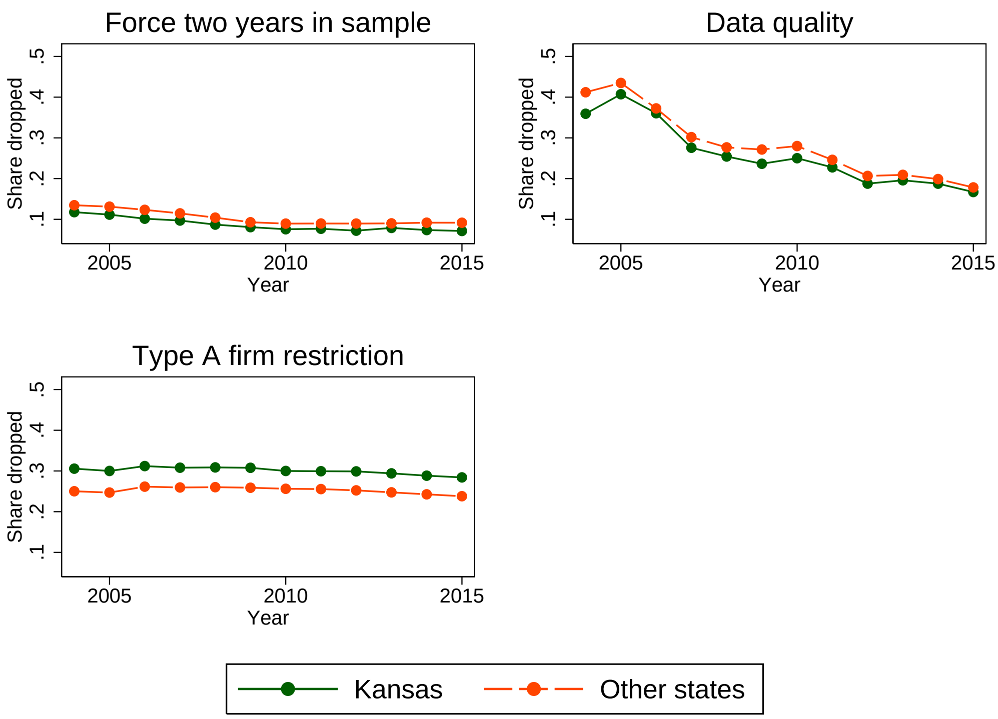
Notes: This figure plots the number of active returns of four different entity types from 1978 through 2014, as reported by Joint Committee on Taxation (2017). Each series is normalized such that the value is set to one in 1985, the year prior to the passage of the Tax Reform Act of 1986.

Figure 3.6: Share of observations dropped in each year



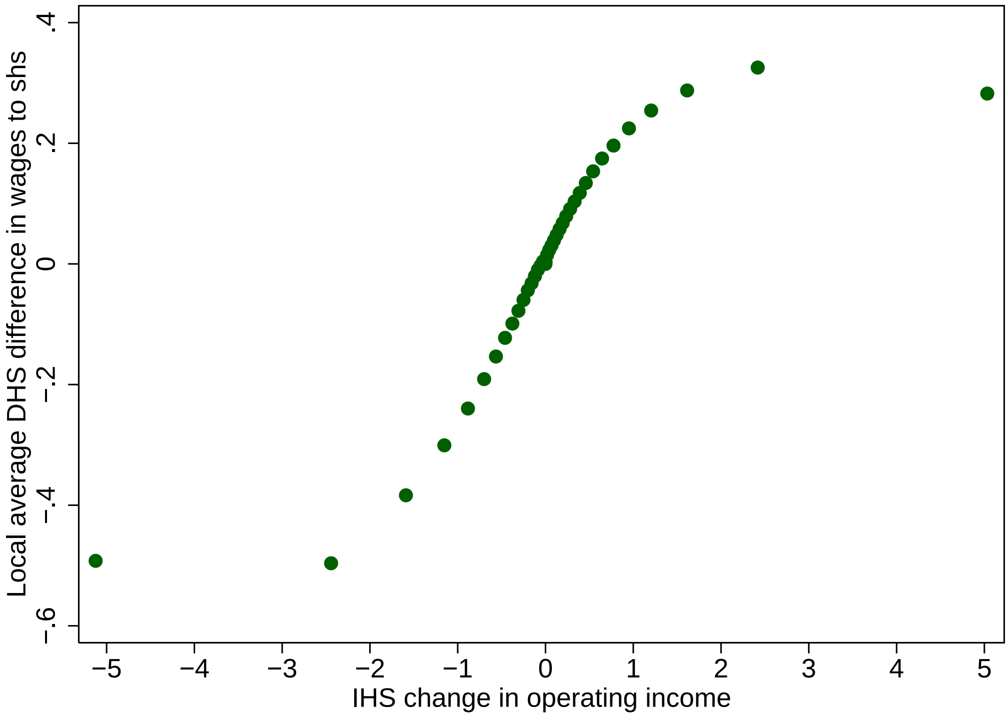
Notes: This figure plots the share of observations that are dropped from the raw data in each year for one of the following two reasons. Firms are dropped if (1) they are not present, or have a data problem described in the *Data Appendix*, at times $t-1$ or t or (2) they have at least one shareholder that is not an employee at time $t-1$.

Figure 3.7: Share of observations dropped in each year: separately by sample restriction



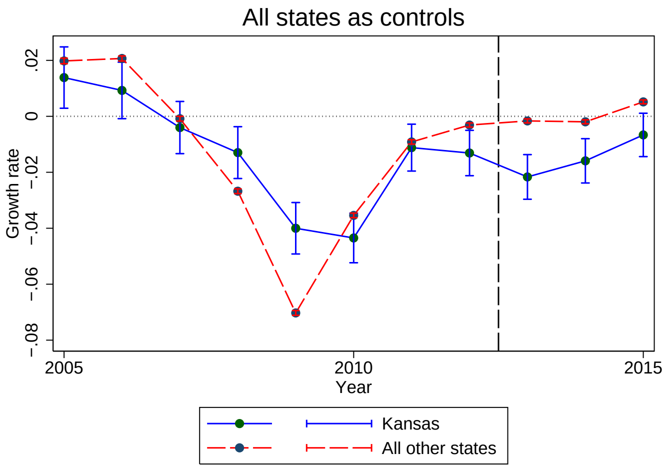
Notes: This figure plots the share of observations dropped in each of three steps. The upper-left panel plots the share of firms that are dropped because they are not present in the raw data at time $t-1$. The upper-right panel plots the share of remaining firms that are dropped because they have a data problem described in in the *Data Appendix*, at times $t-1$ or t. The lower panel plots the share of remaining firms that are dropped because they have at least one shareholder that is not an employee at time $t-1$ (i.e., they are not Type A firms).

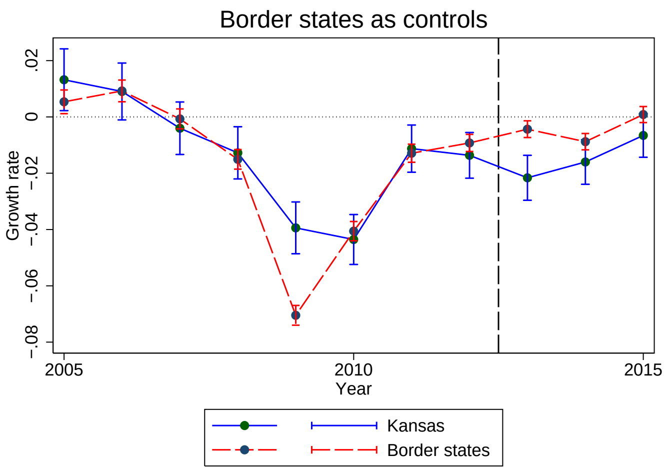
Figure 3.8: Relationship between DHS difference in wages to shareholders and the IHS difference in operating income



Notes: This figure plots a binned scatter plot of the DHS (Davis, Haltiwanger, Schuh) difference in wages to shareholders as a function of the modified inverse hyperbolic sine (IHS) difference in operating income. See text for a definition of both of these functional forms. The sample includes all observations from 2010-2015.

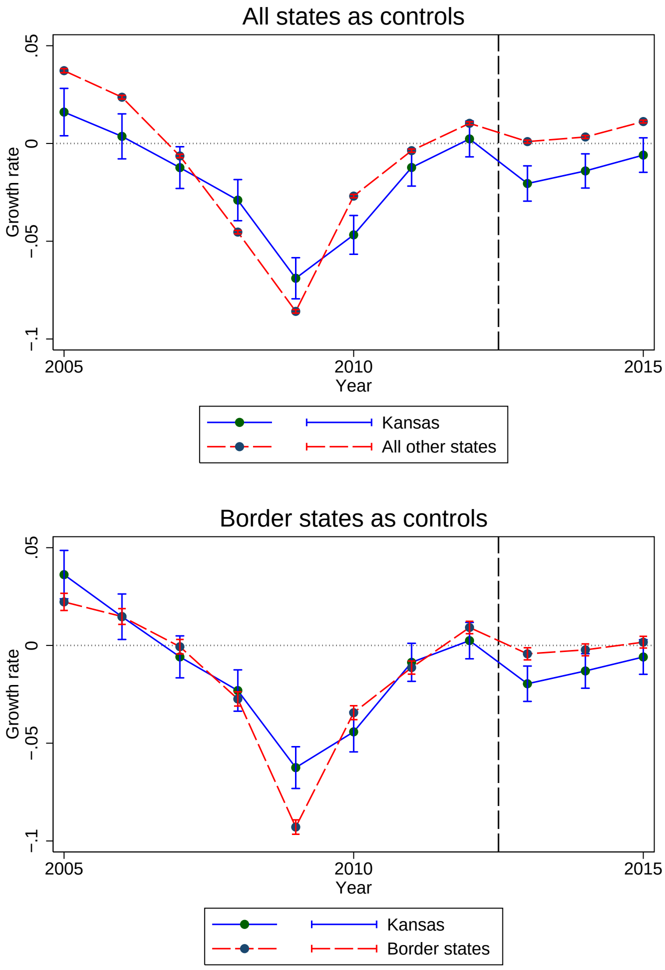
Figure 3.9: Evolution of baseline dependent variable (DHS difference in wages paid to shareholders) in Kansas and control states, 2004-2015, controlling for the change in operating income





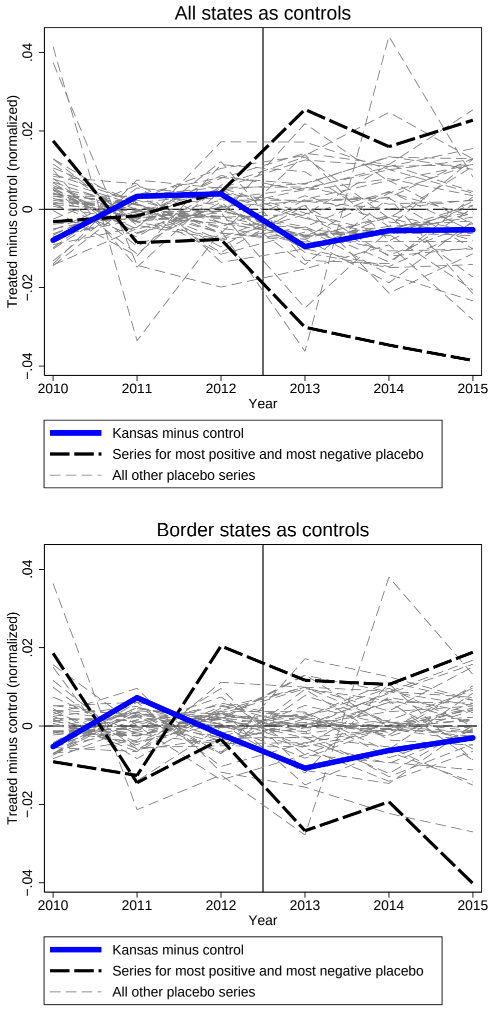
Notes: This figure plots the evolution of the mean dependent variable y_{ist}, the DHS difference (percent change at the midpoint) of wages paid to shareholders, in Kansas and in control states, controlling for the change in operating income. In particular, these are the coefficients from a regression of y_{ist} on a full set of control-by-year and treat-by-year dummies, with the change in operating income included as a control variable. Thus, the points on the graph can be interpreted as the average growth rate when the change in (nominal) operating income is zero. The sample includes "Type A" firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. In the top panel, all states are used as control states. In the bottom panel, only border states (Colorado, Missouri, Nebraska, and Oklahoma) are used as control states. See text for additional "data quality" sample restrictions.

Figure 3.10: Evolution of baseline dependent variable (DHS difference in wages paid to shareholders) in Kansas and control states, 2004-2015, after propensity-score reweighting



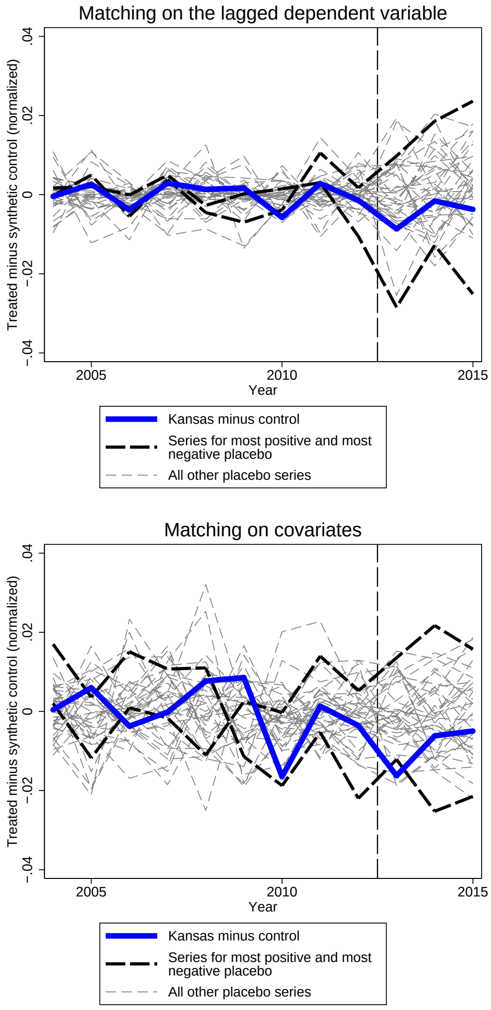
Notes: This figure plots the evolution of the dependent variable y_{ist}, the percent change at the midpoint of wages paid to shareholders, after reweighting the sample according to inverse propensity score weights (as described in the notes to **Table 3.3**). The sample includes "Type A" firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions.

Figure 3.11: True treatment effect relative to distribution of placebo treatment effects: after propensity-score reweighting



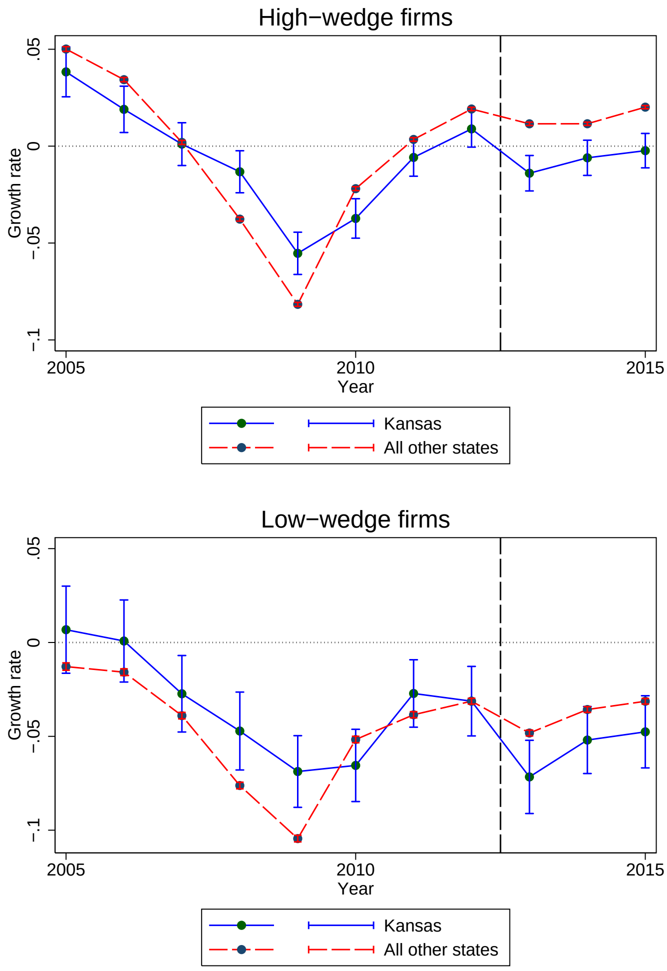
Notes: This figure plots the Kansas treatment effect relative to placebo treatment effects after reweighting the sample according to inverse propensity score weights (as described in the notes to **Table 3.3**). The thick black line plots the mean dependent variable (the percent change at the midpoint of wages paid to shareholders) in Kansas relative to the mean dependent variable in control states, normalized such that this difference equals zero, on average, in the pre-period (2010-2012). The remaining series plot the mean dependent variable in each state j relative to the mean dependent variable in all other states, again normalized to equal zero in the pre-period. The darkest dashed lines represent the series corresponding to the most negative and most positive placebo treatment estimates. The top panel uses all states as control states. The bottom panel uses only border states (i.e., the set of states bordering placebo-treated state j) as control states. The sample includes Type A firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions.

Figure 3.12: Synthetic Control: comparison of treatment effect to placebo treatment effects



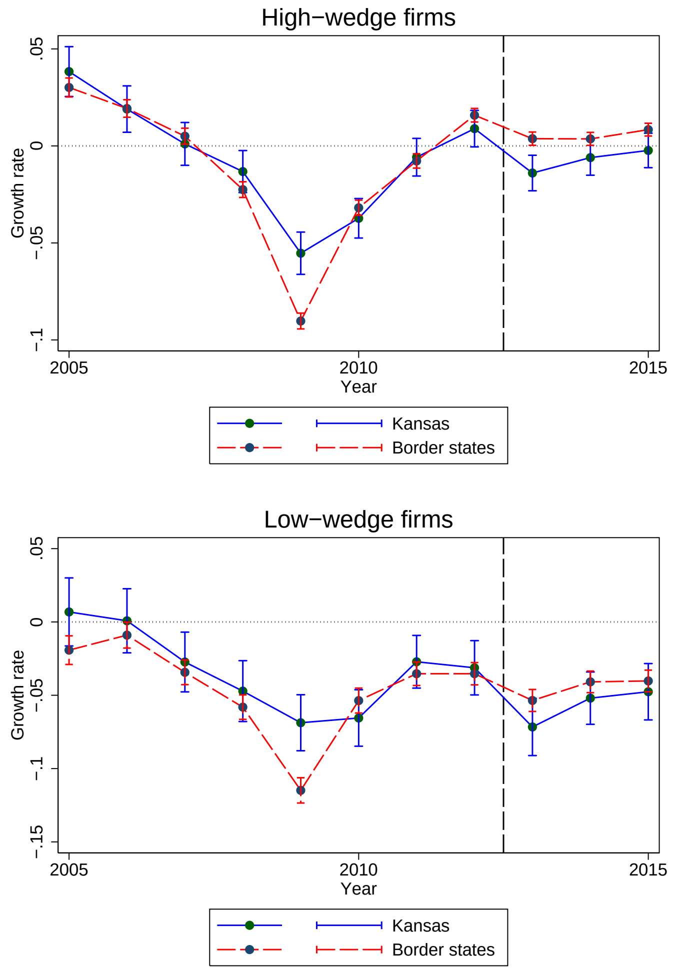
Notes: This figure plots the Kansas treatment effect relative to placebo treatment effects under the synthetic control method. The thick black line plots the mean dependent variable (the percent change at the midpoint of wages paid to shareholders) in Kansas relative to the mean dependent variable in the synthetic control state, normalized such that this difference equals zero, on average, in the pre-period (2005-2012). The remaining series plot the mean dependent variable in each state j relative to the mean dependent variable in their synthetic control state, again normalized to equal zero in the pre-period. The darkest dashed lines represent the series corresponding to the most negative and most positive placebo treatment estimates. The top panel constructs the synthetic control weights in order to match the path of the dependent variable in the pre-period. The bottom panel additionally imposes covariate matching (Equation 3.7). The sample includes Type A firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions.

Figure 3.13: Heterogeneous effects: Evolution of baseline dependent variable (DHS difference in wages paid to shareholders) in Kansas and control states, 2004-2015, separately for high-wedge and low-wedge firms.



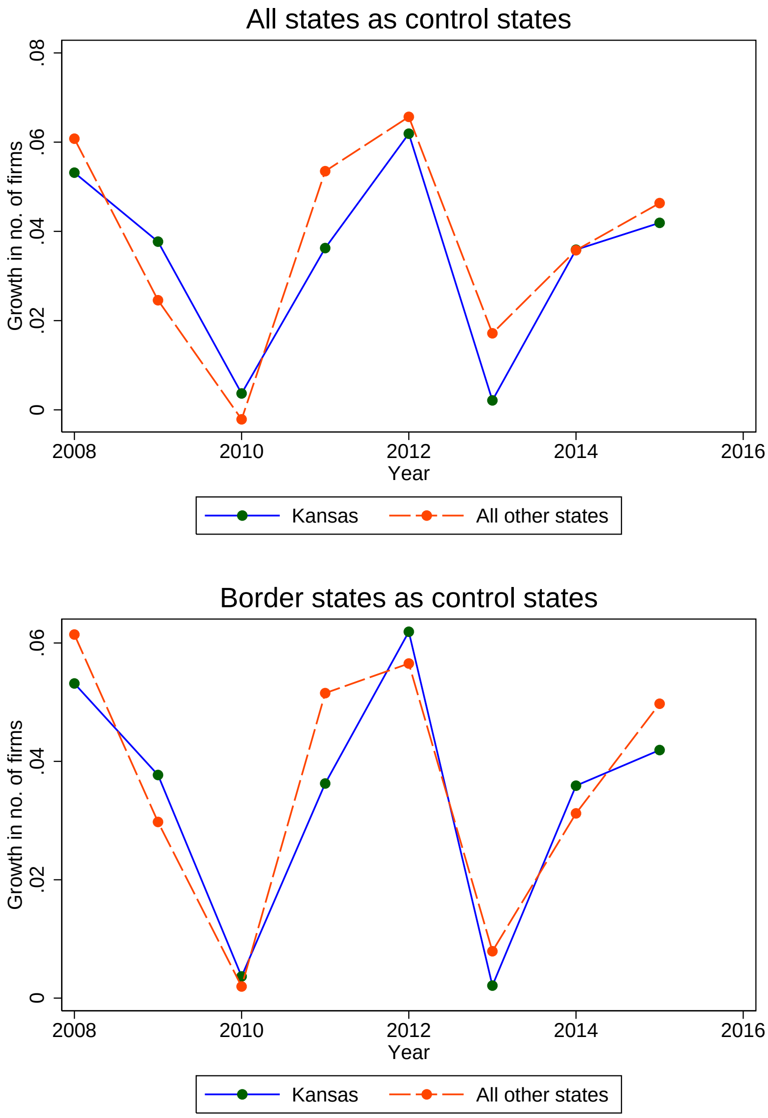
Notes: This figure plots the evolution of the dependent variable y_{ist}, the percent change at the midpoint of wages paid to shareholders, separately for high-wedge and low-wedge firms. High-wedge firms, presented in the top panel, are firms for which fewer than half of shareholders have earnings (at time $t-1$) greater than the Social Security cap. Such firms faced an initially higher tax wedge between profits and wages. Low-wedge firms, presented in the bottom panel, are firms for which at least half of shareholders meet that restriction. Such firms faced an initially lower tax wedge. The sample includes "Type A" firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions.

Figure 3.14: Heterogeneous effects: Evolution of baseline dependent variable (DHS difference in wages paid to shareholders) in Kansas and control states, 2004-2015, separately for high-wedge and low-wedge firms, using only border states as control states.



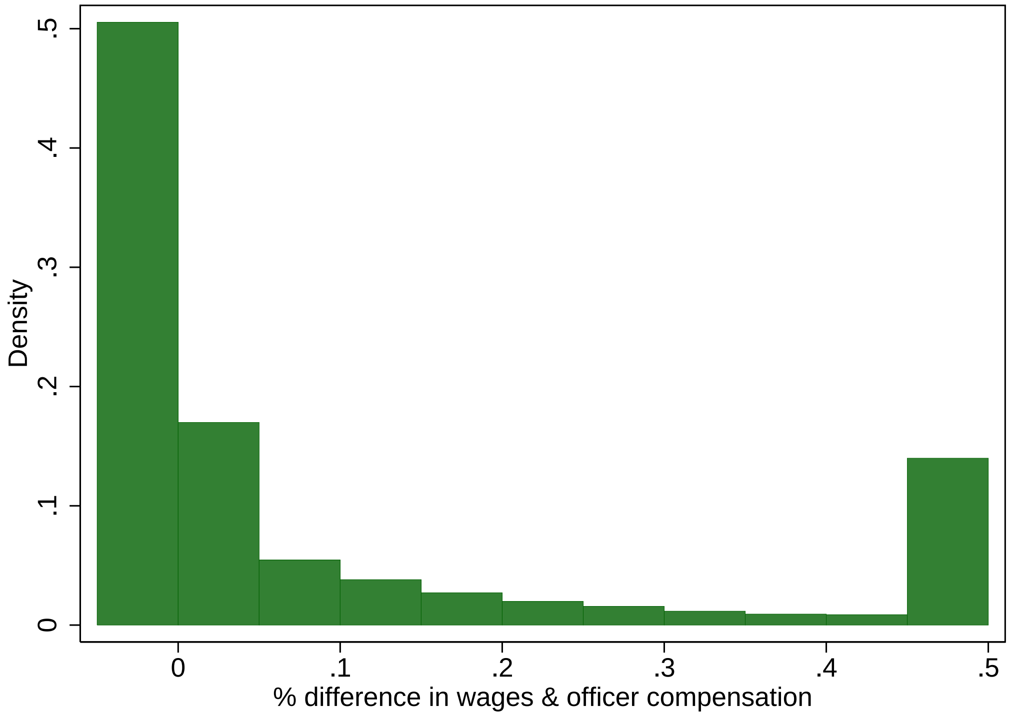
Notes: This figure plots the evolution of the dependent variable y_{ist}, the percent change at the midpoint of wages paid to shareholders, separately for high-wedge and low-wedge firms. Only the four states bordering Kansas (Colorado, Missouri, Nebraska, and Oklahoma) are included as control states. High-wedge firms, presented in the top panel, are firms for which fewer than half of shareholders have earnings (at time $t-1$) greater than the Social Security cap. Such firms faced an initially higher tax wedge between profits and wages. Low-wedge firms, presented in the bottom panel are firms for which at least half of shareholders meet that restriction. Such firms faced an initially lower tax wedge. The sample includes "Type A" firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions.

Figure 3.15: Growth rate in number of firms in main sample



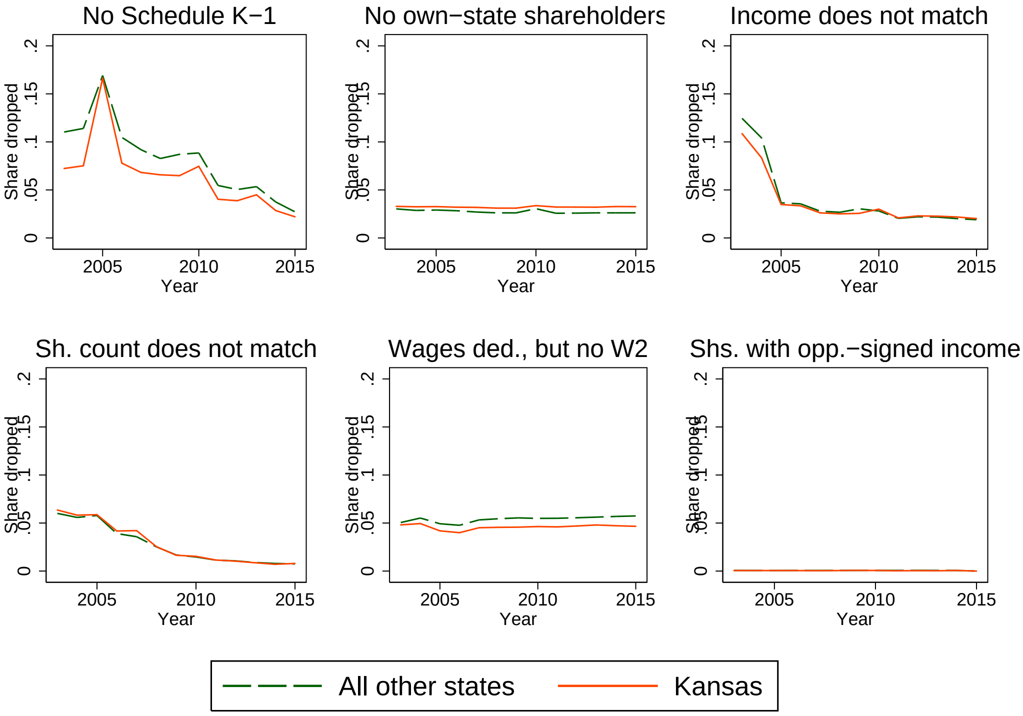
Notes: This figure plots the first difference in the log number of firms in Kansas and in all other states among Type A firms. "Type A" firms, defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. See text for additional "data quality" sample restrictions.

Figure 3.16: Histogram of relative difference between wages to shareholders and officer compensation (as directly reported on Form 1120S)



Notes: This figure plots a histogram of the share of firm-year observations, based on the absolute (log) difference between wages to shareholders and officer compensation. The sample includes Type A firms from 2010-2015, conditional on paying positive wages to shareholders. "Type A" firms are defined as having one of the following two characteristics at time $t-1$: (1) all shareholders are employees, or (2) the firm has a single shareholder. The leftmost bin includes only firms where officer compensation equals wages to shareholders exactly. The rightmost bin includes firms where the log difference is greater than 0.5; this bin also includes firms where officer compensation is reported to be zero. See text for additional "data quality" sample restrictions.

Figure 3.17: Share of firms dropped due to data quality in each year: separately by data quality step



Notes: This figure plots the share of firms dropped in each year, for each of the following four reasons sequentially: (1) there is no Schedule K-1 found, (2) there are no shareholders (as listed on the Schedule K-1) from the same state as the firm (as indicated by Form 1120S), (3) the sum of income from Schedule K-1 does not match Form 1120S, (4) the number of shareholders from Form 1120S does not match the number of Schedules K-1, (5) the firm reports a deduction for wages or officer compensation paid, but no Forms W2 issued by the firm are found, or (6) at least one shareholder reports positive ordinary income while another reports negative ordinary income from the firm. Step (1) is presented in the upper-left corner, step (2) in the upper-middle, step (3) in the upper-right, step (4) in the lower-left, step (5) in the lower-middle, and step (6) in the lower-right. The reported shares correspond to the share of firms dropped among those firms that have survived the previous steps.

Bibliography

Abadie, A., Athey, S., Imbens, G.W., & Wooldridge, J.M. (2014). Finite Population Causal Standard Errors. NBER Working Paper No. 20325. Retrieved October 6, 2017 from `http://www.nber.org/papers/w20325`.

Abadie, A. & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. The American Economic Review, 93, 113-132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association, 105, 493-505.

Auerbach, A. & Poterba, J. (1988). Capital Gains Taxation in the United States: Realizations, Revenue, and Rhetoric. Brooking Papers on Economic Activity, 1988, 595-637.

Auerbach, A. & Slemrod, J. (1997). The Economic Effects of the Tax Reform Act of 1986. Journal of Economic Literature, 35, 589-632.

Auten, G., Splinter, D., & Nelson, S. (2016). Reactions of High-Income Taxpayers to Major Tax Legislation. National Tax Journal, 69, 935-964.

Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica, 77, 1229-1279.

Baicker, K. (2005). The spillover effects of state spending. Journal of Public Economics, 89, 529-544.

Bailey, M.A. (2005). Welfare and the Multifaceted Decision to Move. American Political Science Review 99, 125-135.

Borjas, G.J. (1999). Immigration and Welfare Magnets. Journal of Labor Economics, 17, 607-637.

Borjas, G.J. (2017). The Labor Supply of Undocumented Immigrants. Labour Economics, 46, 1-13.

Bound, J. & Holzer, H.J. (2000). Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s. Journal of Labor Economics, 18, 20-54.

Brown, C.C. & Oates, W.E. (1987). Assistance to the poor in a federal system. Journal of Public Economics, 32, 307-330.

Brueckner, J.K. (2000). Welfare Reform and the Race to the Bottom: Theory and Evidence. Southern Economic Journal, 66, 505-525.

Bull, N. & Burnham, P. (2008). Taxation of Capital and Labor: The Diverse Landscape of Entity Type. National Tax Journal, 61, 397-419.

Burman, L.E., Clausing, K.A., & O'Hare, J.F. (1994). Tax Reform and Realizations of Capital Gains in 1986. National Tax Journal, 47, 1-18.

Burman, L.E. & Randolph, W.C. (1994). Measuring Permanent Responses to Capital-Gains Tax Changes in Panel Data. The American Economic Review, 84, 794-809.

Cameron, A.C. & Miller, D.L. (2015). A Practioner's Guide to Cluster-Robust Inference. Journal of Human Resources, 50, 317-372.

Carroll, R. & Joulfaian, R. (1997). Taxes and Corporate Choice of Organizational Form. OTA Paper 73. Retrieved December 12, 2016 from `https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/WP-73.pdf`.

Chodorow-Reich, G. (2017). Geographic Cross-Sectional Fiscal Multipliers: What Have We Learned? Working Paper. Retrieved February 2, 2018 from https://economicdynamics.org/meetpapers/2017/paper_1214.pdf.

Chodorow-Reich, G. & Karabarbounis, L. (2016). The Limited Macroeconomic Effects of Unemployment Benefit Extensions. NBER Working Paper No. 22163. Retrieved February 2, 2018 from http://www.nber.org/papers/w22163.

Coglianese, J. (2015). Do Unemployment Insurance Extensions Reduce Employment? Working Paper. Retrieved February 2, 2018 from http://scholar.harvard.edu/files/coglianese/files/coglianese_2015_ui_extensions.pdf.

Cohodes, S.R., Grossman, D.S., Kleiner, S.A., & Lovenheim, M.F. (2016). The Effect of Child Health Insurance Access on Schooling: Evidence from Public Insurance Expansions. Journal of Human Resources, 51, 556-588.

Congressional Budget Office. (2012). Unemployment Insurance in the Wake of the Recent Recession. Retrieved February 2, 2018 from https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/reports/UnemploymentIns_One-col.pdf.

Congressional Budget Office. (2016). The Budget and Economic Outlook: 2016 to 2026. Retrieved March 8, 2016 from https://www.cbo.gov/publication/51129.

Conley, T.G. and Taber, C.R. 2011. Inference with "Difference in Differences" with a Small Number of Policy Changes. Review of Economics and Statistics, 93, 113-125.

Cooper, M., McClelland, J., Pearce, J., Prisinzano, R., Sullivan, J., Yagan, D., Zidar, O., & Zwick, E. (2015). Business in the United States: Who Owns It and How Much Tax Do They Pay? Office of Tax Analysis Working Paper 104. Retrieved December 12, 2016 from https://www.treasury.gov/resource-center/tax-policy/tax-analysis/Documents/WP-104.pdf.

Correia, S. (2016). Linear Models with High-Dimensional Fixed Effects: An Efficient and Feasible Estimator. Working Paper. Retrieved February 1, 2018 from http://scorreia.com/research/hdfe.pdf.

Currie, J. & Gruber, J. (1996a). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. Quarterly Journal of Economics, 111, 431-466.

Currie, J. & Gruber, J. (1996b). Saving Babies: The Efficacy and Cost of

Recent Changes in the Medicaid Eligibility of Pregnant Women. Journal of Political Economy, 104, 1263-1296.

Cutler, D.M. & Gruber, J. 1996. Does Public Insurance Crowd out Private Insurance? Quarterly Journal of Economics, 111, 391-430.

Daly, M.C., Hobijn, B., Sahin, A., & Valletta, R.G. (2012). A Search and Matching Approach to Labor Markets: Did the Natural Rate of Unemployment Rise? Journal of Economic Perspectives, 26, 3-26.

Davis, S.J., Haltiwanger, J.C., and Schuch, S. (1996). *Job Creation and Destruction*. Cambridge: MIT Press.

DeBacker, J., Heim, B.T., Ramnath, S.P., & Ross, J.M. (2017). The Impact of State Taxes on Pass-Through Businesses: Evidence from the 2012 Kansas Income Tax Reform. Working paper. Retrieved October 11, 2017 from `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2958353`

Department of Health and Human Services. (2014). 2014 Actuarial Report: On the Financial Outlook for Medicaid. Report to Congress. Retrieved January 22, 2016 from `https://www.medicaid.gov/medicaid-chip-program-information/by-topics/financing-and-reimbursement/downloads/medicaid-actuarial-report-2014.pdf`.

Dowd, T., Landefeld, P., & Moore, A. (2017). Profit shifting of U.S. multinationals. Journal of Public Economics, 148, 1-13.

Dowd, T. & McLelland, R. (2017). The Bunching of Capital Gains Realizations. Working Paper. Retrieved May 29, 2017 from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2899093

Dube, A., Lester, T.W., Reich, M. (2010). Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties. Review of Economics and Statistics, 92, 945-964.

Dube, A., Lester, T.W., Reich, M. (2016). Minimum Wage Shocks, Employment Flows, and Labor Market Frictions. Journal of Labor Economics, 34, 663-704.

Elsby, M.W., Hobijn, B., & Sahin, A. (2010). The Labor Market in the Great Recession. Brooking Papers on Economic Activity, 1-69.

Farber, H.S., Rothstein, J., & Valletta, R.G. (2015). The Effect of Extended Unemployment Insurance Benefits: Evidence from the 2012-2013 Phase-Out. American Economic Review, 105, 171-176.

Farber, H.S. & Valletta, R.G. (2015). Do extended unemployment benefits lengthen unemployment spells? Evidence from recent cycles in the US labor market. Journal of Human Resources, 50, 873.

Farhi, E. & Werning, I. (2016). Fiscal Multipliers: Liquidity Traps and Currency Unions. Handbook of Macroeconomics, 2, 2417-2492.

Farooq, A. & Kugler, A. (2016). Beyond Job Lock: Impacts of Public Health Insurance on Occupational and Industrial Mobility. NBER Working Paper No. 22118. Retrieved March 5, 2018 from `http://www.nber.org/papers/w22118`.

Finkelstein, A., Hendren, N. & Luttmer, E.F. (2015). The Value of Medicaid: Interpreting Results from the Oregen Health Insurance Experiment. NBER Working Paper 21308. Retrieved January 22, 2016 from `http://www.nber.org/papers/w21308`.

Fisher, R.A. (1935). *The Design of Experiments.* New York: Hafner Publishing Company.

Fiva, J.H. (2009). Does welfare policy affect residential choices? An empirical investigation accounting for policy endogeneity. Journal of Public Economics, 99, 529-540.

Frean, M., Gruber, J., Sommers, B.D. (2017). Premium subsidies, the mandate, and Medicaid expansion: Coverage effects of the Affordable Care Act. Journal of Health Economics, 53, 72-86.

Garthwaite, G., Gross, T., Notowidigdo, M.J. (2014). Public Health Insurance, Labor Supply, and Employment Lock. Quarterly Journal of Economics, 129, 653-696.

Gelbach, J.B. (2004). Migration, the Life Cycle, and State Benefits: How Low Is the Bottom? Journal of Political Economy, 112, 1091-1130.

Goolsbee, A. (2000). What Happens When You Tax the Rich? Evidence from Executive Compensation. The Journal of Political Economy, 108, 352-378.

Goodman, L. (2017). The Effect of the Affordable Care Act Medicaid Expansion on Migration. Journal of Policy Analysis and Management, 36, 211-238.

Gordon, R. & Slemrod, J. (2000). Are 'Real' Responses to Taxes Simply Income Shifting Between Corporate and Personal Tax Bases? in Slemrod, J. (ed), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, Russel Sage Foundation and Harvard University Press, 2000, 240-280.

Government Accountability Office. (2009). TAX GAP: ACtions Needed to Address Noncompliance with S Corporation Tax Rules. Report to the Committe on FInance, U.S. Senate. Retrieved December 14, 2016 from `http://www.gao.gov/new.items/d10195.pdf`

Gross, T. & Notowidigdo, M.J. (2011). Health insurance and the consumer bankruptcy decision: Evidence from expansions of Medicaid. Journal of Public Economics, 95, 767-778.

Gruber, J. & Simon, K.. 2008. Crowd-out 10 years later: Have recent public insurance expansions crowded out private health insurance? Journal of Health Economics, 27, 201-217.

Hagedorn, M., Karahan, F., Manovskii, I., & Mitman, K. (2016). Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects. Working Paper. Retrieved February 2, 2018 from `kurtmitman.com`.

Hagedorn, M., Manovskii, I., & Mitman, K. (2016). The Impact of Unemployment Benefit Extensions on Employment: The 2014 Employment Miracle. Retrieved February 2, 2018 from `kurtmitman.com`.

Heckman, J.J. (1979). Sample Selection Bias as a Specification Error. Econometrica, 47, 153-161.

Handel, B.R. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. American Economic Review, 103, 2643-2682.

Internal Revenue Service. (2016). SOI Tax Stats: Business Tax Statistics. Retrieved December 12, 2016 from `https://www.irs.gov/uac/soi-tax-stats-business-tax-st`

Johnston, A.C. & Mas, A. (2016). Potential Unemployment Insurance Duration and Labor Supply: The Individual and Market-Level Response to a Benefit Cut. NBER Working Paper No. 22411. Retrieved February 2, 2018 from http://www.nber.org/papers/w22411.

Joint Committee on Taxation. (2017). Present Law and Data Related To The Taxation of Business. JCX-42-17. Retrieved September 26, 2017 from https://www.jct.gov/publications.html?func=startdown&id=5021

Kaestner, R., Kaushal, & Van Ryzin, G. (2003). Migration consequences of welfare reform. Journal of Urban Economics, 53, 357-376.

Kaiser Family Foundation. (2016a). Medicaid/CHIP Upper Income Eligibility for Children, 2000-2016. Retrieved January 22, 2016 from http://kff.org/medicaid/state-indicator/medicaidchip-upper-income-eligibility-limits-for-child

Kaiser Family Foundation. (2016b). Medicaid Income Eligibility Limits for Other Childless Adults, 2011-2016. Retrieved January 22, 2016 from http://kff.org/medicaid/state-indicator/medicaid-income-eligibility-limits-for-other-n

Kaiser Family Foundation. (2016c). Medicaid Income Eligibility Limits for Parents, 2002-2016. Retrieved January 22, 2016 from http://kff.org/medicaid/state-indicator/medicaid-income-eligibility-limits-for-parents/

Kaiser Family Foundation. (2018). Status of State Action on the Medicaid Expansion Decision. Retrieved March 5, 2018 from https://www.kff.org/health-reform/state-indicator/state-activity-around-expanding-medicaid-under-th ?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

Kroft, K. & Notowidigdo, M.J. (2016). Should unemployment insurance vary with the unemployment rate? Theory and evidence. The Review of Economic Studies, 83, 1092-1124.

Lalive, R., Landais, C., & Zweimuller, J. (2015). Market Externalities of Large Unemployment Insurance Extension Programs, 105, 3564-3596.

Landais, C., Michaillat, P., & Saez, E. (2010). A Macroeconomic Theory of Optimal Unemployment Insurance. NBER Working Paper No. 16526. Retrieved February 2, 2018 from http://www.nber.org/papers/w16526.

Lee, D.S. (2009). Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects. Review of Economic Studies, 76, 1071-1102.

Lo Sasso, A.T. & Buchmueller, T.C. (2004). The effect of the state children's health insurance program on health insurance coverage. Journal of Health Economics, 23, 1059-1082.

MacKinnon, J.G. & Webb, M.D. (2016). Difference-in-Differences Inference with Few Treated Clusters. Working Paper. Retrieved May 29, 2017 from https://www.dal.ca/content/dam/dalhousie/pdf/faculty/science/economics/seminars/MacKinnon_Webb.pdf

Marinescu, I. (2017). The general equilibrium impacts of unemployment insurance: Evidence from a large online job board. Journal of Public Economics, 150, 14-29.

Mazerov, M. (2016). Kansas' Tax Cut Experience Refutes Economic Growth Predictions of Trump Tax Advisors. Retrieved December 29, 2016 from http://www.cbpp.org/sites/default/files/atoms/files/moore_paper_as_published_with_proposed_mm_updates_js_r1.pdf

McCloskey, J.A. (2017). The Kansas Tax Experiment: Impact of 2012 Kansas Tax Reform on Output, Employment, & Establishments. Working Paper. Retrieved January 26, 2018 from https://jesmccloskey.github.io/pdf/mccloskey_kstax.pdf

McKinnish, T. (2005). Importing the Poor: Welfare Magnetism and Cross-Border Welfare Migration. Journal of Human Resources, 40, 57-76.

McKinnish, T. (2007). Welfare-induced migration at state borders: New evidence from micro-data. Journal of Public Economics, 91, 437-450.

Meyer, B.D. (2000). Do the Poor Move to Receive Higher Welfare Benefits? Working Paper. Retrieved January 22, 2016 from https://www.researchgate.net/profile/Bruce_Meyer/publication/23551262_Do_the_Poor_Move_to_Receive_Higher_Welfare_Benefits/links/542095890cf241a65a1e42c2.pdf

Michaillat, P. (2012). Do Matching Frictions Explain Unemployment? Not in Bad Times. American Economic Review, 102, 1721-1750.

Mitman, K. & Rabinovich, S. (2014). Do Unemployment Benefit Extensions Explain the Emergence of Jobless Recoveries? Working Paper. Retrieved February 2, 2018 form https://www.bde.es/f/webpi/SES/seminars/2016/files/sie1626.pdf.

Molloy, R, Smith, C.L., & Wozniak, A. (2011). Internal Migration in the United States. Journal of Economic Perspectives, 25, 173-196.

Nakamura, E. & Steinsson, J. (2014). Fiscal Stimulus in a Monetary Union: Evidence from US Regions. American Economic Review, 3, 753-792.

Nelson, S. (2016). Paying Themselves: S Corporation Owners and Trends in S Corporation Income, 1980-2013. Office of Tax Analysis Working Paper 107. Retrieved December 12, 2016 from https://www.treasury.gov/

resource-center/tax-policy/tax-analysis/Documents/WP-107.pdf

Oates, W.E. (1999). An Essay on Fiscal Federalism. Journal of Economic Literature, 37, 1120-1149.

Office of Management and Budget. (2013). Revised Delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and Guidance on the Uses of the Delineations of These Areas. Memorandum. Retrieved January 22, 2016 from https://www.whitehouse.gov/sites/default/files/omb/bulletins/2013/b-13-01.pdf

Passel, J.S. & Cohn, D. (2014). Unauthorized Immigrant Totals Rise in 7 States, Fall in 14 States: Decline in Those From Mexico Fuels Most State Decreases. Washington, DC: Pew Research Center. Retrieved January 22, 2016 from http://www.pewhispanic.org/2014/11/18/unauthorized-immigrant-totals-rise-in

Pence, K.M. (2006). The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving. B.E. Journal of Economic Analysis & Policy, 5, 1-24.

Pissarides, C.A. (2010). *Equilibrium Unemployment Theory.* 2000: MIT Press, Cambridge, MA.

Rothstein, J. (2011). Unemployment Insurance and Job Search in the Great

Recession. Brookings Papers on Economic Activity, 143-196.

Ruggles, S., Genadek, K., Goaken, R., Grover, J., & Sobek, M. (2015). Integrated public use microdata series: Version 6.0 [machine-readable database]. Minneapolis, MN: University of Minnesota.

Saavedra, L.A. (2000). A Model of Welfare Competition with Evidence from AFDC. Journal of Urban Economics, 47, 248-279.

Saez, E. (2004). Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications. In J. Poterba (Ed.), *Tax Policy and the Economy, Volume 18* (pp. 117-173). Cambridge, MA: MIT Press.

Sahm, C.R., Shapiro, M.D., & Slemrod, J. (2012). Check in the Mail or More in the Paycheck: Does the Effectiveness of Fiscal Stimulus Depend on How It Is Delivered? American Economic Journal: Economic Policy, 4, 216-250.

Sammartino, F. & Weiner, D. (1997). Recent Evidence on Taxpayers' Response to the Rate Increases in the 1990's. National Tax Journal, 50, 683-705.

Schaffer, M.E. (2010). xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML, and k-class regression for panel data models. http://ideas.repec.org/c/boc/bocode/s456501.html

Schwartz, A.L. & Sommers, B.D. (2014). Moving for Medicaid? Recent Eligibility Expansions Did Not Induce Migration From Other States. Health Affairs, 33, 88-94.

Slemrod, J. (1995). Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986. The American Economic Review, 85, 175-180.

Slemrod, J. (1996). High-Income Families and the Tax Changes of the 1980s: The Anatomy of Behavioral Response. In M. Feldstein & J. Poterba, J. (Eds.), *Empirical Foundations of Household Taxation* (pp. 169-192). Chicago: University of Chicago Press.

Slemrod, J. (1998). Methodological Issues in Measuring and Interpreting Taxable Income Elasticities. National Tax Journal, 51, 773-778.

Slemrod, J. & Yitzhaki, S. (2002). Tax Avoidance, Evasion and Administration. In A. Auerbach and M.Feldstein (Eds.), *Handbook of Public Economics, Volume 3* (pp. 1423-1470). Amsterdam, Netherlands: Elsevier Science B.V.

Smith, M., Yagan, D., Zidar, O., & Zwick, E. Capitalists in the Twenty-First Century. Working Paper. Retrieved September 12, 2017 from `https://eml.berkeley.edu/~yagan/Capitalists.pdf`.

Summers, L. (2010). The Economic Case for Extending Unemployment Insurance. Blog post. Retrieved February 2, 2018 from `https://obamawhitehouse.archives.gov/blog/2010/07/14/economic-case-extending-unemployment-insurance`.

Tornqvist, L., Vartia, P., and Vartia, Y.O. 1985. How Should Relative Changes Be Measured? The American Statistician, 39, 43-46.

Turner, T. & Blagg, B. (2015). The Short-term Effects of the Kansas Income Tax Cuts on Employment. Public Finance Review, 1-20.

Wildasin, D.E. 1991. Income Redistribution in a Common Labor Market. American Economic Review, 81, 757-774.

Yitzhaki, S. (1987). On the Excess Burden of Tax Evasion. Public Finance Review, 15, 123-137.