Title
Essays in Health and Public Economics

Author
Robinson, Sarah R

Publication Date
2023

University of California
Santa Barbara

Essays in Health and Public Economics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Economics

by

Sarah R. Robinson

Committee in charge:

Professor Heather Royer, Chair
Professor Youssef Benzarti
Professor H.E. (Ted) Frech, III
Professor Alisa Tazhitdinova

June 2023

The Dissertation of Sarah R. Robinson is approved.

Professor Youssef Benzarti

Professor H.E. (Ted) Frech, III

Professor Alisa Tazhitdinova

Professor Heather Royer, Committee Chair

June 2023

Essays in Health and Public Economics

Acknowledgements

I could not have completed this dissertation without my advisor, Heather Royer, and my committee Youssef Benzarti, Ted Frech, and Alisa Tazhitdinova. Thank you for your support, guidance, and honest feedback every step of the way. From the big topics such as what makes a question interesting, to the small like the detailed format of my slides, I'm so thankful that you took the time to share your thoughts, teach me what I didn't yet understand, and encourage me to try again. Importantly, every one of the (many) times you told me an idea or an execution was bad, I felt confident that it was because you truly believed I could do better. I am especially indebted to Heather and Alisa for the opportunity to be their co-author – I have learned an incredible amount from each of you about the (never-ending) process of research and how to keep moving foward. Beyond research, you all have taught me so much about how to be a good member of the economics profession, and how to do this all as part of a good life.

I'm extremely grateful for the entire UCSB community, who have made this a wonderful place to spend six years. In particular, the applied micro group, who are all incredibly smart and generous with their time. Thank you to the best cohort ever, who made first year survivable and the later years much better than that. And a special thank you to my friends Woongchan Jeon, Chris Malloy, Molly Schwarz, and Jacob Gellman for being buddies to talk about research with and to forget about research with.

Thank you to my father, Larry, and my sister, Rachael, for your support and encouragement. And thank you to all of my friends for doing the tough job of being there and keeping me sane from afar!

Last but not least, thank you to my partner, Christopher Eldred, for his unconditional support; for being a witness during the lows and pushing me to celebrate during the highs, and for his invaluable perspective from outside the economics world.

Curriculum Vitæ
Sarah R. Robinson

Appointments

July 2023 – Assistant Professor of Economics, Claremont McKenna College

Education

2023	Ph.D. in Economics, University of California, Santa Barbara.
2018	M.A. in Economics, University of California, Santa Barbara.
2013	B.A. in Philosophy, Politics and Economics, Claremont McKenna College.

Publications

"Corporate Political Spending and State Tax Policy: Evidence from Citizens United" with Cailin Slattery and Alisa Tazhitdinova. Journal of Public Economics, May 2023, Vol. 221, 104859.

Working Papers

"Geographic Variation in C-Sections in the United States: Trends, Correlates and Other Interesting Facts" with Heather Royer and David Silver (Revise & Resubmit, Journal of Labor Economics)

"Do Firms Avoid Health Insurance Mandates? Evidence from the Self-Funding of Employer Plans"

"What Drives Tax Policy? Political, Institutional and Economic Determinants of State Tax Policy" with Alisa Tazhitdinova (NBER Working Paper 31268, May 2023)

Abstract

Essays in Health and Public Economics

by

Sarah R. Robinson

In this dissertation I use empirical methods to estimate the causal effects of policy on individuals and firms, as well as to investigate the mechanisms that drive differences in health insurance coverage, health outcomes, and public policy.

In the first chapter, I explore whether firms avoid health insurance mandates by self-funding their health plans. Fifty percent of the U.S. population gets health insurance through an employer, and roughly half of employers only offer one health plan. Therefore, the choices made by firms about what plan(s) to offer are critical to understanding the health insurance available to workers. This paper focuses on one dimension of the firm's decision: whether to self-fund plans (meaning the firm bears the financial risk of claims itself). I study whether firms use self-funding to avoid complying with mandates to cover specific procedures or providers. Using administrative data on the health plans offered by firms and a difference-in-differences design, I find that new mandates increase rates of self-funding among smaller firms (100-249 employees) by 3.2 percentage points, an increase of 14.5%. The mandates do not appear to affect larger firms (250+ employees), who are more likely to already be self-funded in the pre-period. These results imply that new mandates can lead to long-lasting reductions in the proportion of firms that are bound by any state health insurance regulations, including all previously mandated benefits as well as premium taxes.

The second chapter investigates the drivers of tax policy. We collect detailed data on U.S. state personal income, corporate, sales, cigarette, gasoline, and alcohol taxes over

the past 70 years to shed light on the determinants of state tax policies. We provide a comprehensive summary of how tax policy has changed over time, within and across states. We then use permutation analysis, variance decomposition, and machine learning techniques to show that the timing and magnitude of tax changes are not driven by economic needs, state politics, institutional rules, neighbor competition, or demographics. Altogether, these factors explain less than 20% of observed tax variation.

The third chapter studies geographic variation in the use of C-sections. We use U.S. natality data from 1989 to 2017 to investigate county-level geographic disparities in the use of C-section among first-birth singleton mothers. We document the existence and persistence of geographic variation in C-section across low- and high- C-section risk mothers and the sensitivity of C-section use and infant and maternal health outcomes to C-section risk across counties. Our key finding is that counties with high C-section rates perform more C-sections across the entirety of the risk distribution. Yet these higher rates of C-section are correlated with nearly equivalent or better outcomes than counties with less-intensive C-section rates. We also find that C-section use is less responsive to individual characteristics for Black than for white mothers, highlighting possible welfare-reducing disparities.

Permissions and Attributions

1. The content of chapter 2 and appendix A.2 is the result of a collaboration with Alisa Tazhitdinova.

2. The content of chapter 3 and appendix A.3 is the result of a collaboration with Heather Royer and David Silver.

Contents

Chapter 1

Do Firms Avoid Health Insurance Mandates? Evidence from the Self-Funding of Employer Plans

1.1 Introduction

Employers play a crucial role in shaping the health insurance options of workers and their families in the United States. Half of the population receives health insurance coverage from an employer (KFF, 2019). However, little is known about how employers decide what plans to offer. These decisions are important to understand because the offerings at a firm are a limited subset of all available plans: 43% of firms only offer one plan (AHRQ, 2021). Furthermore, the choices made by firms may not be optimal for workers – for example, if firms have imperfect information about worker preferences or satisfying these preferences is not profit-maximizing.

I focus on an important but understudied dimension of employer choice: whether to fully insure or to self-fund their health plan(s). Many employers offer self-funded health insurance plans, meaning the firm bears the financial risk of healthcare claims itself. In 2021, 64% of workers with employer-sponsored coverage were in self-funded plans, up from 44% in 1999 (Figure 1.1, KFF 2021). Self-funding is also more common among larger firms.[1] A potential explanation for why firms increasingly self-fund is that it allows firms to avoid state-level regulations. In particular, while plans at fully insured firms must comply with state requirements that health insurance covers specific procedures or providers ("mandated benefits"), plans at self-funded firms are exempt. However, self-funding is not particularly salient to workers, and the extent to which these high rates of self-funding reflect worker preferences is unclear. Because self-funding is consequential yet inconspicous, in 2019 New York began to require health insurance ID cards to clearly state whether the plan was fully insured or self-funded. The goal of this requirement was to ensure that all individuals "are armed with vital insurance information" and those "with state-regulated health plans receive consumer protections guaranteed by state law" (New York State DFS, 2019).

In this paper, I examine how self-funding among firms responds when states require insurance to cover new benefits. I use difference-in-differences and event study designs to estimate the causal effect of new mandated benefits. However, mandated benefits vary widely in their expected costs – while some mandates increase the cost of insurance by less than $1 per person per year (e.g., blood screening for lead), others cost more than $100 per person per year (e.g., mental healthcare). As a result, I focus on mandates that increase premiums by 1% or more and exclude mandates with negligible costs. I focus on the years 1999 to 2008 because this was a particularly active period for new mandates; the number of (costly) mandates nationwide grew from 528 to 676 (Figure 1.2).[2] In my baseline specification, I consider a binary treatment that compares states before and after

[1]In firms with 1,000 or more employees, 87% of covered workers are in self-funded plans, compared to only 21% of workers at firms with fewer than 200 workers (Figure A.1).

[2]Including mandates with negligible costs, the number of mandates grew from 1,187 to 1,647.

the passage of the first *new* mandate during my time period. As a result, the mandates that existed prior to 1999 do not directly contribute to my estimates, which are identified from variation within states. Because some states pass multiple mandates in this period or even within the same year, I also consider specifications excluding these states or with a continuous measure of treatment.

I use an administrative dataset on the welfare benefits offered by firms, the Form 5500 Series. This dataset is ideal for my setting because whether or not a firm is self-funded is observed for all firms offering health benefits. The Form 5500 must be filed by all private-sector firms with 100 or more employees, providing extensive coverage of firms participating in the "large-group" health insurance market. Excluding firms in the "small-group" market is important because the regulatory environment for these plans quite different. In particular, during the time period of study, the majority of states allowed all firms in the small-group market to waive mandated benefits (Jensen and Morrisey, 1999).[3]

My main finding is that mandates increase self-funding rates among smaller firms (100-249 employees) by 3.2 percentage points, or 14.5%.[4] This indicates substantial avoidance of mandates among these firms, highlighting the importance of accounting for self-funding when studying how mandated benefits affect workers. Because self-funded firms do not have to comply with *any* state regulations of health insurance, firms that avoid new mandates are also no longer required to offer benefits that were mandated previously or may be mandated in the future. Furthermore, self-funded firms do not pay taxes on their insurance premiums, which are levied on fully insured firms by nearly all states (at rates as high as 4%). These effects persist for at least four years after the

[3]States independently determine what size of firm is eligible for the large-group market, as high as 100 full-time equivalents but typically at 50.

[4]Controlling for firm size could bias my estimates if employment is changing as a result of the mandates. To avoid this concern, I categorize firms as small or large using their number of employees in the first year they are observed.

mandate, suggesting that self-funding in untreated states does not catch up to that of treated states in the short-run.

In addition, I document heterogeneous treatment effects across industry groups, with larger effects in industries with higher baseline rates of self-funding and larger average deductibles. However, I do not detect an effect of mandates on self-funding among larger firms (250+ employees). This may be because larger firms were much more likely to be self-funded in the pre-period, and thus less exposed to new mandates. Large firms could also be less affected by mandates because they tend to offer more generous health insurance in general (making mandates less burdensome) and may be more likely to operate in multiple states (such that mandates in one state have only a small impact).[5]

Another way firms could avoid mandates is by ceasing to offer health coverage at all.[6] However, I do not detect a statistically significant effect of mandates on the rates of offering health coverage, and the 95% confidence interval excludes changes larger than 1-2% in either direction. Because self-funding is only defined conditional on offering health coverage, this result also reduces concern that mandates affected selection into my main analysis sample. In addition, I estimate null effects of mandates on employment at firms with health coverage. Finally, I show suggestive evidence that small firms were either less likely to offer any benefits or to operate at all, but the effects are imprecisely estimated and I am not able to distinguish between these two possibilites. Taken together, these results provide additional context for the finding in Sloan and Conover (1998) that mandated benefits reduce the probability that *individuals* are covered by insurance. In particular, these reductions may have occured because small firms ceased to offer any benefits or to operate entirely, rather than from firms dropping health coverage or

[5]Larger firms offer lower deductibles and lower out-of-pocket maximums (KFF, 2021), and Gruber (1994*b*) shows evidence that they are more likely to offer the benefits covered by mandates, conditional on being self-funded.

[6]Firms were not required to offer insurance health insurance to workers during the time period of study (though offer rates did not noticeably change when the Affordable Care Act began to require this).

reducing employment.

A potential concern with my research design is that mandated benefits may be endogenous to self-funding rates of firms – for example, policymakers could be reacting to their state's trends in self-funding when deciding whether or not to pass additional mandates. To address this concern, I provide evidence that treated and untreated states were trending similarly prior to the mandate, supporting a causal interpretation of my estimates. In addition, because I consider a staggered adoption design, my results could be biased if the treatment effect is heterogeneous across states or time (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). This issue is of particular concern in my setting, because "treatment" includes mandates with varying effects on costs (though mandates with negligible costs are excluded), as well as states that passed more than one mandate in the same year or soon after. Reassuringly, my results are qualitatively and quantitatively consistent when I use an alternative estimator that is robust to heterogeneous treatment effects and allows for a continous measure of treatment (the number of mandates passed).

Finally, another concern is that the external validity of my results could be limited to the extent that the time period I study, 1999–2008, is different from the landscape following the Affordable Care Act (ACA). The ACA included many reforms, especially to the individual market for insurance. However, the relationship between state-level regulations and self-funding for firms with 100 or more employees remains unchanged today – self-funded firms do not have to comply with state mandated benefits. The ACA may have had a chilling effect on the passage of *new* mandates, because it requires states to pay for the associated premium increase in the individual market for some types of new mandates (Office of Legislative Research, 2019).[7] However, because this

[7]This is because the federal government provides subsidies for the individual market that are based on actual premiums, so state mandates that increase premiums will also increase the size of the subisides.

rule does not apply to mandates passed prior to 2011, the mandates studied here may still be contributing to self-funding rates today. Furthermore, the rule did not completely deter the passage of new mandates, with the Centers for Medicare and Medicaid Services recently increased reporting requirements on mandates out of concern that states were not reimbursing appropriately (CMS, 2020).[8]

The welfare implications of firms avoiding mandated benefits are theoretically ambiguous, and depend on whether the mandates themselves are welfare-improving or not. On one hand, in the canonical model of Summers (1989), mandates reduce wages and employment. On the other hand, a growing body of empirical evidence suggests that the plans offered by firms may not (as predicted in the model) reflect worker preferences (e.g., Cebul et al., 2011; Fang and Gavazza, 2011; Liu and Sydnor, 2022). Thus, a richer model that takes choice frictions and labor market frictions into account could show that restricting the set of available plans – for example, by requiring plans to include specific benefits – may increase welfare. To the extent that firms avoid mandated benefits, welfare changes are attenuated in both directions: wage and employment losses will be mitigated, but any consumer protection gains (from removing "bad" options) would be diminished as well. In Section 1.2, I discuss a theoretical framework for mandated benefits, avoidance, and the associated welfare implications in more detail. Though quantifying these welfare effects is beyond the scope of this paper, in Section 1.7 I briefly discuss extensions in upcoming work.

My work contributes to several lines of literature. First, it relates to the overarching question of how firms make decisions about the health insurance plans they offer to workers. This paper is the first to study the impact of mandated benefits on the decision

[8]In addition, the ACA included a federal mandate that insurance cover preventative health services without cost-sharing, superceding state-level mandates in this category. But its broader Essential Health Benefits requirements do not apply to firms in the large-group market, whether they are fully insured or self-funded.

to self-fund at small firms – a group that is disproportionately fully insured and thus exposed to the mandates. When analyzing firms overall, Jensen, Cotter and Morrisey (1995) and Park (2000) report null effects, though their estimates rely on a few hundred firms and cross-sectional variation respectively. Using administrative data on all firms with 100 or more employees and a quasi-experimental research design, my results for firms overall are similar. However, unlike these prior studies, my data and design allow me to estimate an effect specifically for small firms, showing that the overall estimates mask important heterogeneity across firm size. Though I do not identify a statistically significant effect among larger firms, Dalton and Holland (2019) show suggestive evidence that self-funding rates may have decreased among publicly-listed firms after controlling for a rich set of corporate finance characteristics, further emphasizing that the effects are heterogeneous across firm size.

Much of the remaining literature on firm decisions has focused on how rising health-care costs jointly affect employee premium contributions and wages (Sommers, 2005; Baicker and Chandra, 2006; Clemens and Cutler, 2014; Lubotsky and Olson, 2015; Anand, 2017; Meiselbach et al., 2022). Liu and Sydnor (2022) show that many firms offer dominated plans, suggesting that firms may face choice frictions or incentives that are not aligned with workers'. Unlike these studies, my paper addresses how firms select the *design* of plans, not just how to share premium costs with workers. There are few other papers with this focus, though limited work in this area includes Moran, Chernew and Hirth (2001) and Bundorf (2002), who show that firm choices about the generosity and diversity of plans vary somewhat with worker age, gender, and income.

In addition, my work contributes to the literatures that study the effects of mandated benefits. Building on the work of Summers (1989), a wide range of studies have analyzed how mandated benefits are passed through to individuals via decreased wages, decreased job creation, decreased probability of health insurance coverage, and increased employee premium contributions (Gruber, 1994*a*; Mathur, 2010; Sloan and Conover, 1998; Bailey and Blascak, 2016). In this paper, I document an avoidance response by firms to mandates. My results show that these existing estimates of pass-through to workers are understated, as this literature is generally unable to control for the self-funding status of firms. If firms avoid mandates, the net effect on wages will be smaller than if avoidance were not available, and a portion of the impact on workers will be overlooked. With respect to the rates of firms offering health coverage, my results complement those of Jensen and Gabel (1992) and Gruber (1994*b*), who study even smaller firms with fewer than 50 or 100 employees respectively, and also do not detect any effect of mandates. This study also complements the work of Mulligan (2020) and Dillender, Heinrich and Houseman (2022), who study how employment and hours per worker responded to the ACA mandate that insurance be offered at all. I study a different context yet show a similar result, in that small firms are willing and able to reorganize aspects of their business in order to avoid costly health insurance regulations. Finally, the avoidance response documented here is an extension of that identified in other settings where firms decide whether to shift to a separate regulatory environment (e.g., shifting profits to tax havens or incorporating (Zucman, 2014; Tazhitdinova, 2020)).

1.2 Theoretical Framework

In this section, I examine the theoretical considerations for how firms may respond to mandated benefits. First, I consider the canonical model of mandated benefits from Summers (1989). Second, I expand this model to allow for avoidance. Third, I briefly discuss several factors not captured in this model, which empirical evidence suggests are quantitatively important and which have countervailing implications for welfare. Fourth, I describe how this framework applies to my empirical setting.

Consider labor demand D and labor supply S, which are both functions of the wage w, such that the equilibrium wage and employment satisfy:

$$D(w_0) = S(w_0) \tag{1.1}$$

Next consider a benefit that is valued by workers at v, but costs firms c to provide. If workers value the benefit more than it costs ($v > c$), it will be offered regardless of any mandate, leaving both firms and workers better off. However, if employees value the benefit less than it costs ($v < c$), the benefit will only be offered if mandated, in which case the new equilibrium will be given by:

$$D(w_1 + c) = S(w_1 + v) \tag{1.2}$$

As shown in Figure 1.3a, the labor demand curve shifts down by the amount that the firm must pay for the benefit, and the supply curve also shifts down to the extent that workers are willing to accept a lower wage in exchange. In the case where $v = c$ such that workers value the benefit at its cost, the mandate has no effect other than how total compensation (wages + value of benefits) is broken down between the two components: $w_1 + v = w_0$. However, as workers' value v decreases, compensation, employment, and welfare also decrease. In particular, the effects of the mandate are equivalent to those from a tax of $(c-v)$. Like a tax, the degree of pass-through to worker wages is determined by the relative wage elasticity of workers and firms.

I extend this model to allow firms to avoid the mandate, such that firms can pay a to not provide the benefit. If the firm does not avoid, then the situation remains as above in Figure 1.3a. However if the firm does avoid, the equilibrium would be given by:

$$D(w_2 + a) = S(w_2) \tag{1.3}$$

where the labor demand curve shifts down, but the labor supply curve is unchanged because the workers do not receive the benefit. Relative to no mandate, welfare is decreased, but the comparison between compliance (L_1, w_1) and avoidance (L_2, w_2) is ambiguous. In particular, it depends on the cost to provide the benefit (c), how much workers value it (v) as well as the cost to avoid (a). If the cost of avoiding is high such that $a > c - v$, then workers and firms will be better off if the firm complies. Otherwise, if $a < c - v$, as in Figure 1.3b, then the firm will choose to avoid, making itself and workers better off. Thus, in this model, the option to avoid mitigates welfare losses from the mandates themselves. Note that an econometrician who observes only wages and employment but not avoidance will underestimate the degree of pass-through to workers. In particular, she would observe workers receiving wage w_2, when wages would actually be lower, at w_1, if firms were not able to avoid. Furthermore, she may assume that workers are still receiving the value of the benefit such that compensation is $w_2 + v$, when in fact they do not and are only compensated by w_2.

In this simple model, the cost of providing the benefit (c) as well as the value of the benefit to workers (v) are observed and fixed. However, c may depend on time-variant characteristics of the market for the benefit, as well as firm preferences beyond the direct financial cost. In addition, firms may be able to lower wages by more or less than v in exchange for offering the benefit. While fully characterizing a model that incorporates these features is beyond the scope of this paper, I generalize the simple model to allow for them. Thus, I consider firms as avoiding the mandate if:

$$a < c(\theta, \eta) - \phi(v, \varepsilon) \tag{1.4}$$

As before, a is the cost of avoiding. I allow the cost of complying with the mandate $c(\cdot)$ to be a function that varies with θ, a measure of adverse selection into the benefit, as well as η, a measure of the firm's preferences. The firm weighs these costs against the amount that could be passed through to workers, $\phi(\cdot)$, which is a function of how much workers value the benefit v and some error ε rather than simply the value alone.

Furthermore, theoretical and empirical work suggest that these additional factors play a significant role. As a result, a richer model accounting for these factors would show that restricting the set of available plans (for example, by mandating plans to cover specific benefits) can increase welfare, in which case firm avoidance of the mandates would be welfare-diminishing. I briefly discuss these factors below. In particular, they each explain how workers could value the benefit more than it would cost to provide, yet the benefit would not exist absent a mandate (or in the presence of a mandate that can be avoided). Of course, if workers do not value the benefit, then mandates will only decrease welfare. On the other hand, it is not clear why policymakers would implement a mandate for a benefit that workers do not value more than its cost. Thus, these factors may also explain why policymakers consider mandated benefits to be worthwhile, at the same time as these benefits are not provided by the free market.

Adverse selection: The welfare losses from mandates described above may be offset by welfare gains if mandates reduce adverse selection. Prior to the mandate, workers may have selected into firms offering the benefit based on their likelihood of using it. The cost of providing the benefit (c) would reflect this selection, such that workers at other firms may not find it worthwhile even if it were efficient to insure them (Akerlof, 1970; Rothschild and Stiglitz, 1976). Thus, with adverse selection, reducing the choice of plans available could theoretically increase welfare by decreasing the cost of providing it (Einav and Finkelstein, 2011; Ericson and Sydnor, 2017; Marone and Sabety, 2022). In this case, if firms also select into avoiding mandates, avoidance would reduce welfare.

Individual choice frictions: In the simple model described above, firms know how much their workers value a given benefit, and in the absence of a mandate will provide benefits that are valued more than they cost. But if employees do not maximize their own preferences, it will be difficult for firms to gather the information needed to do so on their behalf.[9] Empirical work has documented that individuals frequently fail to make rationalizable choices with respect to health insurance and healthcare, and thus employers may misperceive workers' true preferences (v) with some error (ε). For example, employees frequently select dominated plans and generally seem not to understand how insurance works (Bhargava, Loewenstein and Sydnor, 2017; Brot-Goldberg et al., 2017). Individuals also experience frictions when using healthcare: they respond to small increases in out-of-pocket prices by reducing high-value care, which ultimately increases mortality (Chandra, Gruber and McKnight, 2010; Choudhry et al., 2011; Baicker, Mullainathan and Schwartzstein, 2015; Brot-Goldberg et al., 2017; Chandra, Flack and Obermeyer, 2021; Gross, Layton and Prinz, 2022).

Firm choice frictions: Even if firms do know the preferences of workers, it can be difficult to actually offer a health insurance plan that is best suited to these preferences. Firms encounter substantial search frictions when making a selection from the wide range of potential plans (Cebul et al., 2011), similar to the frictions that individuals face when comparing many options (Kling et al., 2012).[10] As a result, the cost of providing a benefit

[9]More specifically, the plan, employee premium contribution, and wage *bundle* may not the bundle that workers most prefer, among the options the firm is willing to provide.

[10]These frictions compound the baseline difficulty that firms face in aggregating heterogeneous worker preferences into a small number of offered plans (Goldstein and Pauly, 1976; Moran, Chernew and Hirth, 2001; Bundorf, 2002).

(c) may depend on firm factors such as search costs that are beyond the direct financial cost.

Firm incentives: Finally, maximizing worker preferences may not be profit-maximizing. Liu and Sydnor (2022) show that a large fraction of firms offer a dominated plan – meaning a plan that is financially worse for every employee, regardless of how much healthcare they use, compared to another plan offered at the same firm. This finding, that firms offer plans that none of their workers should take, is difficult to reconcile with the standard frictionless model. Liu and Sydnor (2022) go on to provide suggestive evidence that firms may differentially favor high-deductible plans. Firm preferences of this type would be reflected in the cost to the firm of providing certain benefits (c). In addition, Fang and Gavazza (2011) show that employee turnover and frictions in the labor market lead firms to underinvest in employee health (raising expenditures in retirement). Thus, frictions in the labor market could be reflected in the degree to which firms have incentives to respond to worker preferences (v).

Altogether, these factors suggest that in a richer theoretical model, restricting plan options by mandating benefits may in fact increase (and thus avoidance would decrease) welfare. Finally, to apply this model to avoidance via self-funding, it is important to note that self-funding exempts firms from *all* mandated benefits, including previously enacted ones. In addition, the costs to a firm of self-funding are unlikely to vary substantially with the number of mandated benefits. As a result, it may be most appropriate to think of a new mandate as an increase in the cost of complying with mandated benefits (c), and potentially an increase in the value of mandated benefits to workers (v), without much change in the cost of avoiding (a). If so, firms with low avoidance cost may already be self-funded by the time I observe them, and not affected by new mandates at all. Thus, we may expect the firms avoiding new mandates (marginal firms) to have higher costs of avoidance, compared to the firms that are already self-funded (inframarginal firms). Additionally, to the extent that new mandates change the cost of complying with mandated benefits, we may expect that firms facing higher marginal compliance costs (firms for whom it is more costly to provide new benefits) to be the firms with an avoidance response.

1.3 Institutional Setting

Employer-sponsored health insurance is the primary way that working-age adults receive health coverage in the U.S. About two-thirds of all firms offer health coverage, including 99% of firms with 200 or more workers, and these figures have remained steady since 1999 (KFF, 2021). Health coverage plays an extremely important role in how workers are paid – it is typically the most expensive non-wage component, constituting 8% of total compensation (BLS, 2020). Firms offering health coverage can structure their plans to be fully insured or self-funded. The choice of whether or not to self-fund has substantial implications for a firm's obligations and finances. There are four main differences:

Payments: Under full insurance, the firm pays monthly premiums to an insurance carrier. Premiums are experience-rated, meaning that they are customized to the firm based on firm characteristics and its claims history.[11] However, premiums are negotiated and set for the length of the contract (a few years). Under self-funding, the firm pays a fixed fee to an insurer to administer the plan, but pays employee healthcare claims itself.

Financial risk: Under full insurance, the insurer bears risk – if healthcare claims are unusually high in a month or year, the firm continues to pay the same premium. On the other hand, a self-funded firm bears the financial risk of high claims. (Self-funded firms

[11] I focus exclusively on firms with 100 or more employees. Firms with fewer employees (fewer than 50 or 100, depending on the state) participate in the small-group market. Those plans were a mix of medically underwritten, experience-rated, community-rated prior to the ACA, and community-rated after the ACA (Hall and McCue, 2021).

may also purchase stoploss coverage, which can limit the amount that the firm pays in claims.)

Plan design: Fully insured firms have less control over the design of health plans. Self-funded plans can customize the benefits covered, cost-sharing arrangements, and even the provider network to a greater degree.

Compliance: The plans that fully insured firms buy from insurers must comply with all federal and state regulations. However, self-funded plans are covered by the Employee Retirement Income Security Act of 1974 (ERISA), which pre-empts state regulations (McCuskey, 2022). As a result, self-funded plans do not have to comply with any state health insurance regulations. For example, at the onset of the COVID-19 pandemic, many states passed laws requiring coverage of telehealth services. But because these laws were at the state level, they only applied to fully insured firms, and self-funded firms were not required to comply.[12]

To employees, the experience of using their health coverage is approximately the same under the two types of plan – the appearance of insurance cards and the process of finding, using, and paying for care are almost identical. Insurance cards for self-funded plans typically include the name of the plan administrator and a statement that they provide administrative services only. Figure A.2 includes an example health insurance card for a self-funded plan that is administered by Anthem Blue Cross. (Similarly, self-funding is also not particularly salient to healthcare providers.)

The primary drawback to firms of self-funding is the assumption of financial risk. Each individual's healthcare claims are uncertain, and the distribution of claims is highly skewed: among ages 18-64, the top 1% of individuals account for 23% of healthcare spending (Ortaliza et al., 2021). As a result, self-funded firms can face much higher costs (relative to if they were fully insured) if only a few of their employees experience

[12] Both fully insured and self-funded plans must comply with federal regulations.

large health shocks. Thus, the variance of these firms' potential claims is an important consideration in their financial planning. Large firms will face a smaller variance, and are correspondingly more likely to self-fund, because they can spread the risk across more employees. As shown by Dalton and Holland (2019), firms that have less difficulty accessing liquidity or lower opportunity cost of investments are also more likely to self-fund, because higher than anticipated claims are less burdensome to them.[13]

The primary appeal to firms of self-funding is the flexibilty in plan design. In particular, because self-funded firms can avoid state regulations, they are excluded from state laws that require health insurance to cover specific benefits or providers. Many factors may influence how appealing this option is to firms – for example, mandated benefits may be inconsistent with a firm's "values," or the firm may expect many more mandates to occur in the future. However, I focus on the most tangible factor in how appealing self-funding is to firms: the cost of adding the existing mandated benefits to their health coverage plan.

1.4 Data

1.4.1 Form 5500 Series

I use the Form 5500 Series, an administrative dataset on employee welfare benefit plans. This form is submitted on an annual basis to jointly satisfy reporting requirements with the Department of Labor and Internal Revenue Service; firms face penalties for non-filing and responses are subject to audit. I focus on the time period 1999 (?rst available)

[13]Note that the variance of claims can affect the choices of firms without any assumption about the firm's risk preferences – for example, a risk-neutral firm facing a non-linear budget constraint due to limited liquidity.

to 2008 (to avoid any anticipation of the 2010 Affordable Care Act).[14]

Firms are required to file the Form 5500 for each employee welfare benefit plan with 100 or more participants. An employee welfare benefit plan includes one or more of: health, dental, vision, life insurance/death benefits, disability (temporary or long-term), supplemental unemployment, severance, prepaid legal, scholarship, apprenticeship and training, or housing.[15] The mapping of benefits to plans is at the firm's discretion: firms may choose to file information about a single plan that covers all of their benefits, or file separately for multiple plans (e.g., one health plan and one dental plan). In order to weight all firms equally regardless of what they choose, I aggregate the set of plans to the firm level.[16]

Each filing includes the number of plan participants, which are measured as of the end of the plan year, reflect individuals rather than full-time equivalents, and do not include family (e.g., spouse and dependents enrolled in an employee's health plan). Because I do not observe the number of employees directly, I define the firm's number of employees as the number of participants in its largest plan (across all welfare benefits).[17] As a result, this measure may exclude employees who are not eligible for any of the benefits listed above, such as part-time workers. A few firms report implausibly large numbers of participants, so I exclude the top 1% of firms by size.

The Form 5500 data do not directly specify whether a plan is fully insured or self-funded, but firms are required to provide details about insurance contracts and how benefits are paid for. Among firms that offer health benefits, I follow guidance from the

[14]As discussed in Section 1.1, the ACA made no changes to the relationship between mandated bene?ts and self-funding for firms with 100 or more employees.

[15]I focus on welfare benefit plans, though the Form 5500 also separately collects information on pension benefits (defined benefit and defined contribution).

[16]I also exclude a small number of multi-employer plans, i.e., where one plan covers employees at more than one firm.

[17]For example, if a firm has 100 participants in one health plan, 200 participants in a second health plan, and 300 participants in a single life insurance plan, I will count them as having 300 employees.

Department of Labor to identify self-funded plans (DOL, 2021). The most important factor in identifying fully insured plans is the presence of any health insurance contract details, and the most important factor in identifying self-funded plans is indication that benefits are paid from general assets or a trust.

The main outcome of interest is whether or not the firm is self-funded, conditional on offering health coverage. For almost all observations (93%), this outcome is equal to either zero (if the firm is fully insured) or one (if the firm is self-funded). However, for the remaining observations, the outcome is a fraction between zero and one. This occurs when a firm divides their health plans across multiple Form 5500s filings, only some of which are self-funded. In this case, I identify self-funding at the plan level and then use the participant-weighted average of self-funding across the firm's plans. I also consider as outcomes the (log) number of employees at firms with health coverage; whether the firm offers any health coverage; and whether firms report any welfare benefits at all (e.g., life insurance or long-term disability) through the Form 5500, though I cannot distinguish firms that offer no benefits from firms that are not operating.

The Form 5500 data also includes limited information about the firm such as industry (6-digit NAICS code) and address. A limitation of this dataset is that only one address is observed per firm. As a result, I treat each firm as though all of its employees are in the same state as its headquarters, an assumption which is more plausible for smaller firms. While precise data on the prevalence of multi-state firms is not easily available, size is a strong determinant of the number of establishments at a firm. Firms with 100-299 employees have 3 establishments on average, while 300+ employee firms have 41 establishments on average (U.S. Census Bureau, 2019).

For the sample of firms offering health coverage, Figure 1.4 shows the distribution of firm size in 1999, the first year of the sample. Most firms are small; the median number of employees is 342. Furthermore, size is an important determinant of self-funding rates

– as shown in Figure 1.5, larger firms are much more likely to be self-funded. As a result, my empirical approach allows the response of smaller firms to differ from the response of larger firms. I follow the U.S. Census Bureau's classification of firm sizes in selecting a cutoff of 250 employees, but my results are similar if I use higher or lower cutoffs.

Table 1.1 shows summary statistics for firms overall, for firms with 100-249 employees, and for firms with 250 or more employees. Because firm size could be endogenous to the passage of mandated benefits, when separating firms by size I use employment in the first year the firm is observed. Statistics are shown for the sample of firms offering health coverage as well as the broader sample of all firms in the dataset. 88% of firms offer health benefits.[18] Of these firms, 26% are self-funded overall, with self-funding rates of 22% among smaller firms and 29% among larger firms.

Over the course of 10 year period, firms are observed for 5.5 years on average, which reflects an improbable amount of entry and exit, and may be due to the fact that firms often change their Employee Identification Number (EIN). To better connect firms over time, I allow for changes in EIN if the firm name and address remain the same. Even so, the low tenure indicates that this methodology does not perfectly capture changes. As a result, I refrain from using firm fixed effects throughout my analysis.

1.4.2 State Benefit Mandates

Data on mandates comes from the Blue Cross Blue Shield Assocation (Laudicina, Gardner and Holland, 2013). This report identifies state-level mandates that specific procedures, providers, or persons be covered by insurance, as well as the year in which

[18]The higher rate of offering health coverage among smaller firms in Table 1.1 is an artifact of categorizing firms by size in their first year. Around 7% of observations are from firms that were small initially but became large. When I instead split firms by contemporaneous size, 82.1% of smaller firms and 91.1% of larger firms offer health coverage.

each mandate was passed.[19] All states have at least one mandate, ranging from 8 mandates in Idaho to 39 mandates in Maryland in 1998.

There is substantial heterogeneity in the cost of incorporating these mandates into a health plan. In particular, many mandates are expected to have negligible effects on insurance premiums, while a few mandates are expected to be quite costly. Mandated benefits with very low costs may be used by a small number of people, be associated with low spending per person, or both. Despite this, prior work in this area has mostly focused on effects from the *total number* of mandates, implicitly treating every mandate as equally costly.

To account for this heterogeneity, I use cost estimates from several sources (described below) to exclude mandates with negligible effects on premiums. These sources use claims data for fully insured plans, calculate the total spending related to the mandated benefits, and average the spending across all plan participants. Thus, the cost estimates are relative to zero spending on these benefits, and will overstate the marginal cost increase if firms offered the benefit prior to the mandate. On the other hand, the cost estimates are measured among fully insured plans, and may be understated if the firms that would experience high spending on the mandated benefits switch to being self-funded. The estimates do account for potential moral hazard in that they measure spending after the benefits have been mandated, rather than before.

The primary source of mandates costs comes from the Council of Affordable Health Insurance (CAHI), a research and advocacy association of insurance carriers, in 2009. CAHI provides cost estimates across all states, which is important because the exact coverage and language of each mandate can vary from state to state (e.g., for mandated infertility treatment benefits, the number of IVF cycles that are covered may vary from

[19] Mandates typically take effect within a year of passage, and firms switching to self-funding likely do so at the beginning of their benefit year.

state to state.) As a result, CAHI estimates whether each mandate will increase costs by: less than 1%, 1-3%, 3-5%, and 5-10%. One limitation of this approach is that the “less than 1%” category continues to mask a significant amount of heterogeneity in costs. Therefore, I supplement using reports from three states who study the costs associated with their own mandates and provide numerical estimates: Connecticut in 2009, Massachussets in 2013, and Rhode Island in 2014. These states are representative in terms of the number of mandates in 1998 as well as the number of additional mandates passed. However, my results are similar if I only rely on the cost estimates from CAHI to identify costly mandates.

Figure 1.6 shows the methodology for identifying costly mandates. There are 60 mandates that are newly enacted in one or more states in my time period of study. Of these, I include the 19 mandates that are estimated by CAHI to cost at least 1% of premiums. For the remaining 41 mandates, I look for any estimate across the three state reports that the mandate will cost more than \$50 per person per year, and I find an additional 2 (Diabetic Supplies & Education and Home Health Care). Thus, I include in my analysis 21 mandates that at least one source has identified as costly, and I exclude the remaining 39 mandates that no source has identified as costly. The costly mandates have mean (median) estimated costs of \$63 (\$28), while the excluded mandates have estimated of costs of \$11 (\$2).

Figure 1.7 shows the distribution of mandates across states over time. From 1998 to 2008, the average number of mandates in a state increased from 10.3 to 13.2. This growth occurred throughout the distribution, as shown by the median and interquartile range. The distribution for all mandates, including thoes with negligible cost estimates, can be found in Figure A.3. Finally, the list of mandates and how frequently they contribute to my causal estimates are show in Table A.1.

1.5 Empirical Framework

I estimate the impact of mandated benefits using a difference-in-differences design with a two-way fixed effects specification. First, I discuss a specification considering all firms together:

$$Y_{it} = \beta Mandate_{st} + \gamma_s + \delta_t + \varepsilon_{it} \tag{1.5}$$

where Y_{it} represents the outcome for firm i in year t, such as whether the firm is self-funded. State fixed effects γ_s ensure that estimates are identified from variation within states, rather than cross-sectional comparisons. Year fixed effects δ_t control for idiosyncratic time effects. In all specifications, standard errors are clustered at the state level (level of treatment).

The treatment variable, $Mandate_{st}$, is an indicator equal to one in the years after state s passes any costly mandate. By considering a binary treatment, I am comparing states before and after their passage of the first new mandate during my time period. Because my results are identified from variation within states, mandates that existed prior to 1999 do not contribute to the estimated effects. While most states pass only one mandate at a time, some states pass multiple mandates within a year. Figure 1.8 shows variation by state in the intensity of treatment: 31 states pass a single mandate at the time of treatment; 16 states including D.C. added between two and four mandates; and 4 states do not pass any mandates in this time period (are never treated).[20] I include all states in my main specification, such that my estimates are the effect of "changing the number of mandates for the first time" rather than the effect of one additional mandate. However, my results are similar if I exclude the states that added more than one mandate. In addition, all but 18 states go on to pass more mandates in subsequent years, but my

[20] The number of mandates passed is not strongly related to the number of mandates in existence as of 1998 (Figure A.4), with a correlation of only -0.095.

results are similar when excluding these states from the analysis.

In order to interpret β as the causal effect of mandates on firm outcomes, the standard difference-in-differences parallel trends assumption must hold. In my setting, this requires that mandates are uncorrelated with other unobserved time-varying determinants of firm outcomes related to the health insurance plans they offer to workers. In other words, in the absence of mandates, there would have been no change in firm outcomes among treated states relative to states that did not pass mandates. To test the plausibility of this assumption, I expand Equation 1.5 to an event study framework with leads and lags:

$$Y_{it} = \sum_{k \neq -1} \beta_k 1\{t - \tau = k\} + \gamma_s + \delta_t + \varepsilon_{it} \tag{1.6}$$

For $k < -1$, the β_k coefficients estimate anticipatory responses of firm outcomes k years before the state passes any costly mandate, relative to the year immediately prior. If these lead coefficients are very close to zero, then treated and control firms were trending similarly prior to the mandate, lending support to the assumption that they would have continued to do so in the absence of any mandates. Conversely, for $k > -1$, the β_k coefficients estimate the response k years after the mandate, and allow me to examine how the response of firms evolves over time.[21]

Next, I consider a specification where the treatment effect can differ between small and large firms. Allowing this possibility is important because firm size is a strong determinant of self-funding rates (see Figure 1.5). Equation 1.7 describes the difference-

[21]When estimating Equation 1.6, I consider event times -5 to 3, where 0 is the year of treatment. Event times of -6 and earlier are binned into one indicator, and event times of 4 and later are binned into another indicator (not shown). For comparison, when estimating Equation 1.5, I exclude event times of -6 and earlier, as well as 4 and later. My results are similar when considering narrower event windows.

in-differences specification:

$$Y_{it} = \beta_1 Mandate_{st} * Small_i + \beta_2 Mandate_{st} * Large_i + \theta Small_i + \gamma_s + \delta_t + \varepsilon_{it} \quad (1.7)$$

$Small_i$ is an indicator equal to one if the firm has fewer than 250 employees in the first year the firm is observed. $Large_i$ is an indicator equal to one if the firm has 250 or more employees in the first year of observation. I control for (initial) firm size, so the θ coefficient captures baseline differences in outcomes between small and large firms.[22] Then, I interact firm size with the treatment variable, $Mandate_{st}$. Thus, β_1 is the effect of mandates on firm outcomes for small firms only, and β_2 is the effect of mandate on firm outcomes for large firms only. I also expand this specification to an event study:

$$Y_{it} = \sum_{k \neq -1} \beta_{1k} 1\{t-\tau = k\} Small_i + \beta_{2k} 1\{t-\tau = k\} Large_i + \theta Small_i + \gamma_s + \delta_t + \varepsilon_{it} \quad (1.8)$$

For $k < -1$, the β_{1k} coefficients allow testing for pre-trends among small firms, and the β_{2k} coefficients allow separate testing for pre-trends among large firms. For $k > -1$, the β_{1k} coefficients estimate the treatment effect over time for small firms, and the β_{2k} coefficients estimate these dynamic effects separately for large firms.

To test whether my results are sensitve to alternative specifications, I consider models with additional fixed effects and time-varying covariates. In particular, I allow the idiosyncratic state and time effects to vary by the industry of the firm. I replace the state fixed effects with state-by-industry or state-by-sector fixed effects, where industry is the 6-digit NAICS code for the firm (in the first year of observation) and sector is the 3-digit NAICS code. Similarly, I replace the year fixed effects with year-by-industry or year-by-sector fixed effects. These are potentially important because health insurance outcomes

[22] Because changes in firm size could be endogenous to new mandates, I do not include contemporaneous firm size in my baseline specification. However, my results are similar when I include it.

vary across industries (Figures 1.16-1.20), so a single set of state and time fixed effects may not be sufficient for capturing geographic variation or time shocks. I also consider models that control for the number of contemporaneous employees at the firm, as well as the number of non-costly (excluded) mandates in each state for a given year.

Recent work has shown that two-way fixed effect specifications, as those described above, can lead to biased estimates if the treatment effect is heterogeneous between groups or over time (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). Heterogeneous treatment effects are particularly likely in my setting, because "treatment" includes mandates with varying effects on costs (though mandates with negligible costs are excluded), as well as states that passed more than one mandate in the same year or soon after. Therefore, I also estimate effects using an alternative estimator that is robust to heterogeneous treatment effects. I use the estimator from de Chaisemartin and D'Haultfœuille (2022) for several reasons. First, this estimator allows for dyamic effects, akin to an event study, where the treatment effect may grow or shrink over time. Similarly, placebo effects can also be estimated to test the parallel trends assumption. Finally, unlike other robust estimators, the de Chaisemartin and D'Haultfœuille (2022) estimator also allows for continuous treatment, which is ideal for my setting because some states pass more than one mandate.

1.6 Results

In Section 1.6.1, I first show results on whether mandates have any effect on firm rates of offering health coverage. Then in Section 1.6.2, I study the effect of mandates on self-funding rates. Finally in Section 1.6.3, I analyze the effect of mandates on employment and whether the firm reports any welfare benefits at all.

1.6.1 Offering Health Coverage

One way that firms could avoid complying with mandated benefits is by ceasing to offer health coverage at all. Furthermore, if mandates do affect the rates of offering health coverage, estimates of the effect on self-funding may be biased due to selection (because self-funding is only observed for firms that offer coverage). Thus, I first estimate the effect of mandates on firm rates of offering health coverage. For this analysis, my sample is the set of all firms that report offering any welfare benefits through the Form 5500 in a given year.

Figure 1.9 shows estimates from the event study and difference-in-differences specifications where the treatment effect is not allowed to vary across firm size (Equations 1.5 and 1.6). In the pre-period, the coefficients are not statisically distinguishable from zero, but they are trending downward in a way that suggests treated and control states were evolving differently prior to the mandate. However, this tendency is reduced when the treatment effect is allowed to vary across small and large firms. Figure 1.10 shows estimates of Equations 1.7 and 1.8. In the pre-period, for both small and large firms, the estimates are statistically indistinguishable from zero and are not strongly trending up nor down.[23]

In the post-period, I am unable to detect an effect of mandates on rates of offering insurance among small or large firms. For both types of firms, the difference-in-differences coefficient is -0.2 percentage points and not statisically distinguishable from zero. For small firms, the 95% confidence interval excludes decreases larger than 1.9 percentage points and increases larger than 1.5 percentage points. For large firms, the 95% confidence interval excludes decreases larger than 1.3 percentage points and increases larger than 1.0 percentage point. Relative to the mean rate of offering health coverage, these confidence

[23]For small (large) firms, an F-test that all of the pre-period coefficients are equal to zero has a p-value of 0.07 (0.48).

intervals exclude effects larger than 1-2% for both smaller and larger firms. (Table A.2 shows the detailed difference-in-differences regression results for firms overall and split by size.)

These results suggest that new costly mandates did not have a significant effect on the rates of firms offering health coverage to workers. In addition, the results reduces any potential concern that the sample of firms for which I observe self-funding status could be changing in response to mandates.

1.6.2 Self-Funding

Next, I restrict my sample to the set of firms offering any health coverage and study the effect of mandates on whether or not firms self-fund their health plans. Figure 1.11 shows estimates from the event study and difference-in-differences specifications for firms overall. While there does not appear to be evidence of pre-trends, and the coefficients in the post-period show an upward trend, the treatment effect is not statistically different from zero.

However, Figure 1.12 shows that the effect of mandates varies dramatically by firm size. For smaller firms with 100-249 employees, mandates increase self-funding rates by 3.2 percentage points, representing a 14.5% increase. Self-funding rates increase sharply in the year that the mandate is passed, and rise slightly again in the following year. The effects persist for at least four years following treatment. In contrast, I am unable to detect any response among larger firms, and the 95% confidence inverval excludes increases larger than 0.5 percentage points (1.7%) and decreases larger than 1.7 percentage points (5.9%). (Table A.3 shows the detailed difference-in-differences regression results for firms overall and split by size.) For both small and large firms, coefficients in the pre-period are close to zero and neither trending up nor down.[24] I conduct a series of robustness

[24]For small (large) firms, an F-test that all four of the pre-period coefficients are equal to zero has

checks to test whether my results are sensitive to alternative samples or specifications:

Additional fixed effects and controls: Figure 1.13 shows results where the state and year fixed effects are allowed to vary within sector (3-digit NAICS) or within industry (6-digit NAICS). Also shown are results from regressions that control for the number of (contemporaneous) employees at the firm, as well as the number of non-costly (excluded) mandates in each state for a given year. The estimates for both small and large firms are remarkably stable across all specifications.

Excluding states with additional mandates: I consider several restrictions on the set of states used. Results are similar if I exclude states that pass additional mandates after the treatment year (Figure A.5); exclude states that pass two or more mandates in the treatment year (Figure A.6, though only statistically significant for small firms at the 10% level); or exclude states that passed mandates in any of the four years prior to the treatment year (Figure A.7). Because many states pass mandates in 1999 (Figure A.8), I also show estimates excluding these states, and find quantitatively similar though imprecise results (Figure A.9).

Excluding additional mandates: In Figure A.10 I consider only the mandates identifies by CAHI as costly. The estimates are similar in magnitude, albeit imprecise.

Varying cutoff between small and large firms: Figure A.11 shows that my results are similar when the cutoff between smaller and larger firms is higher or lower than 250. In particular, the effect for smaller firms is quite stable across cutoffs. For larger firms, all but one of the 95% confidence intervals include zero.

Robust to heterogeneous treatment effects: In order to address concerns that two-way fixed effects specifications may be biased when effects are heterogeneous across groups

---

a p-value of 0.02 (0.01). However, when I narrow the window of time around the treatment year, such that only three pre-period coefficients are estimated, the effect for small firms remains stable and the parallel trends test p-value rises to 0.54. For large firms, the p-value rises to 0.08 when I consider two pre-period coefficients (Table A.4).

or time, I turn to the robust estimator from de Chaisemartin and D'Haultfœuille (2022). The results are shown in Figure 1.14. I do not detect evidence of pre-trends for small or large firms. For small firms, self-funding rates rise after the passage of new mandates. Note that the coefficients plotted in the figure are analagous to those from an event study design, in that they estimate the effect of "changing the number of mandates for the first time" k periods after the change. However, they are estimated by comparing the evolution of outcomes in states that add new mandates only to those that have not yet added mandates (rather than to all states, including already-treated states). In addition, this estimation strategy allows me to calculate the average effect of one additional mandate in a way that is robust to heterogeneous treatment effects. I show these average effects for small and large firms. For small firms, I estimate that an additional mandate increases self-funding by 2.0 percentage points and the 95% confidence interval does not include zero. This estimate is smaller in magnitude than the baseline difference-in-differences estimate, which makes intuitive sense because the baseline estimates average across states with one or more mandates at the time of treatment. For large firms, the post-period coefficients show a small upward trend, but are neither the individual coefficients nor the average effect for one mandate are statistically distinguishable from zero.

In Figure 1.15, I show that the effect of mandates on self-funding is heterogeneous across industry groups. For both small and large firms, I interact the treatment variable with indicators for nine industry groups. (Figure A.12 shows how the sample of small firms that offer health is distributed across these industry groups). Small firms in agriculture, fishing, and forestry industries have the largest response, though it is also the least precisely estimated, followed by firms in other service industries and then firms in construction industries. I examine how these treatment effects are related to characteristics of small firms in these industries in 1999. These treatment effects are negatively correlated with rates of offering health coverage (Figure 1.16, -20% correlation) and positively correlated with rates of self-funding (Figure 1.17, 70% correlation). To study the relationship with additional characteristics of these industries, I use estimates of the total premium, employee premium contribution, and deductible in each industry group in 2002 from the Medical Expenditure Panel Survey – Insurance Component (AHRQ, 2021).[25] The effect of mandates on self-funding is negatively correlated with both the average total premium and the average employee premium contribution (Figures 1.18 and 1.19, correlations -44% and -18%), and positvely correlated with the average deductible (Figure 1.20, 69% correlation).

1.6.3 Employment & Reporting Any Welfare Benefits

Finally, I study other two additional margins along which firms may respond to mandates. Figure 1.21 shows the estimated effects of mandates on (log) employment among firms that offer health coverage. The difference-in-differences coefficients are not statistically different from zero, though imprecisely estimated. For large firms, the 95% confidence intervals exclude increases larger than 2.4% and decreases larger than 3.7%. For small firms, the 95% confidence intervals exclude increases larger than 3.2% and decreases larger than 5.4%, though the event study coefficient in the year immediately following the mandate shows a statistically significant decline.

Figure 1.22 shows the estimated effect of mandates on whether firms report any welfare benefits at all. For this outcome, the sample is firms that *ever* report any welfare benefits. Firms could not report benefits if they do not offer any of the welfare benefits associated with the Form 5500. Alternatively, firms would not report benefits if they are no longer in operation. A limitation of the data is that I am not able to distinguish between these two mechanisms. I do not find any evidence of a response for large firms. I find suggestive evidence that small firms cease to offer any benefits or operate, though

[25] Note that these characteristics are for all firms in the industry group, rather than only small firms.

the difference-in-differences coefficient is only statistically significant at the 10% level and the effect appears concentrated in the year immediately following the mandate.

1.7 Discussion & Conclusion

In this paper, I study the effect of state mandated benefits on self-funding for the health insurance offered by firms. I document that new mandates increase self-funding rates among smaller firms (100-249 employees) by 3.2 percentage points, representing a 14.5% increase. These findings indicate a substantial degree of avoidance among these firms, i.e., that they switch to self-funding so that they are not required to comply with the mandates. A limitation of the data is that I do not observe any details about the specific benefits covered by a plan – as a result, I cannot rule out the possibility that firms switch to self-funding but still begin to cover the benefit. However, treated and control firms trend very similarly prior to the mandate, and self-funding rates at treated firms rise right after the mandate is passed. These findings suggest that the mandates have a causal effect on the attractiveness of self-funding to firms. Because costly mandates make flexibility in plan design and exemption from state regulations more valuable, I consider my results to be suggestive evidence that firms are avoiding providing the mandated benefit.

A potential secondary mechanism worth discussing is one of reduced adverse selection. If the mandate succeeds in bringing coverage of the benefit to additional workers, this may reduce adverse selection in the market for that benefit. It is plausible that this reduced adverse selection would itself make self-funding more appealing to firms. For example, if coverage for infertility treatment is covered by more firms after the mandate, each individual firm may find that the use of this benefit is lower and more predictable, making self-funding less financially risky. Future work on how specific plan benefits evolve after a mandate is passed can provide futher insight into the role that these mechanisms play.

I am unable to detect an effect of mandates on larger firms (250+ employees). Size is a significant determinant of self-funding status, where larger firms are much more likely to be self-funded at baseline. As a result, larger firms are more likely to be inframarginal with respect to new mandates (cannot become more self-funded than they already are). There are a few additional potential explanations for why I do not find an effect among larger firms. First, conditional on being fully insured, larger firms may also be more likely to have seriously considered switching – as firms grow, their benefits administration becomes more sophisticated, and these firms may have already decided that self-funding is not advantageous for them. In contrast, smaller firms may be prompted to seriously consider self-funding by the mandates. Interestingly, the effects for small firms persist for at least four years after the mandate, suggesting that there is no "catch-up" by firms not exposed to new mandates in this period.

Second, larger firms offer more generous health insurance in general. Plans at larger firms have lower deductibles and lower out-of-pocket maximums (KFF, 2021), and Gruber (1994*b*) shows evidence that they are more likely to offer the benefits covered by mandates, conditional on being self-funded. If larger firms are already offering the benefit, they would not have any change in their incentive to switch to self-funding. Finally, a limitation of this data is that I only observe one address per firm (headquarters address). To the extent that larger firms are more likely to be present in multiple states, then mandates in one state will affect a smaller proportion of their total workforce, and thus treatment for these firms has a lower intensity. However, self-funding is particularly advantageous for multi-state firms (because they can design one plan without concern for differing state regulations), and so I expect that these firms are also more likely to be self-funded at baseline, and thus inframarginal with respect to new mandates.

I document heterogeneous effects on self-funding for small firms across industry groups. In particular, the treatment effect is larger for industries with higher rates of self-funding, suggesting that these industries find self-funding more appealing overall. Furthermore, the treatment effect is also larger for industries with higher average deductibles. This result would be consistent with the idea that industries with low deductibles can respond to the cost of mandated benefits by raising deductibles, but industries with high deductibles cannot and thus are more likely to use self-funding as a margin of adjustment.

I do not detect an effect of mandates on the rates at which firms offer health coverage, among small or large firms. My results are fairly precise, where the 95% confidence interval excludes increases or decreases of 1-2%. Because self-funding is only defined for the set of firms that offer health benefits, it is important to document that the selection into this sample is not changing as the result of the treatment. I also find no evidence that mandates affected employment among firms that offer health coverage. These findings provide additional context for understanding prior work on how health insurance coverage responds to mandates. In particular, Sloan and Conover (1998) find that a larger number of mandates reduces the probability that individuals are covered by insurance. This paper suggests that these reductions may not occur by firms dropping health coverage specifically, or by firms that offer coverage reducing employment. Rather, these effects may occur because small firms drop all kinds of welfare benefits or cease to operate.

Firms' avoidance of state mandates through self-funding presents a significant challenge for policy. When firms self-fund, they are no longer required to comply with *any* state regulations for health insurance. So, when considering whether or not to mandate a new benefit, policymakers must account for a variety of effects on compliance. A firm switching to self-funding may not offer the benefit currently being mandated; it can drop benefits associated with older mandates; and it will not be required to comply with any future mandates. Thus, policymakers need to consider a dynamic problem where stricter policies today may lead to reduced regulatory scope tomorrow. For example, when states mandated greater coverage of telehealth at the beginning of the COVID-19 pandemic, all firms that had decided to self-fund previously were exempted (even if no new firms becamse self-funded).

Furthermore, this firm avoidance response has implications for all of the other types of state-level regulations beyond mandated benefits. Nearly all states require fully insured firms to pay taxes on their insurance premiums, as high as 4%, but these do not apply to self-funded firms. States may also pass laws along other dimensions, such as California's 2017 law protecting consumers from surprise medical bills, which did not apply to self-funded firms (though they were impacted by a similar a federal law that came into effect in 2022). When firms switch to self-funding to avoid mandated benefits, they also become exempt from all these other regulations. In addition, my results raise the possibility that these other types of regulations could also affect firms' decision to self-fund.

A welfare analysis of firm avoidance of mandates would require, first, a comprehensive study of the ways that employees might be impacted by mandates. While such an exercise is beyond the scope of this paper, in upcoming work I will extend this analysis to study how firms use multiple margins of adjustment to respond to mandates and other increases in healthcare costs. In particular, I will study how self-funding, plan design, employee premium contributions, and wages are jointly adjusted by firms. As a result, I will be able to estimate the total pass-through of mandates to workers, how it is distributed across the various margins of adjustment, and the extent to which additional factors (such as the firm's labor market power) affect which channel the firm chooses.

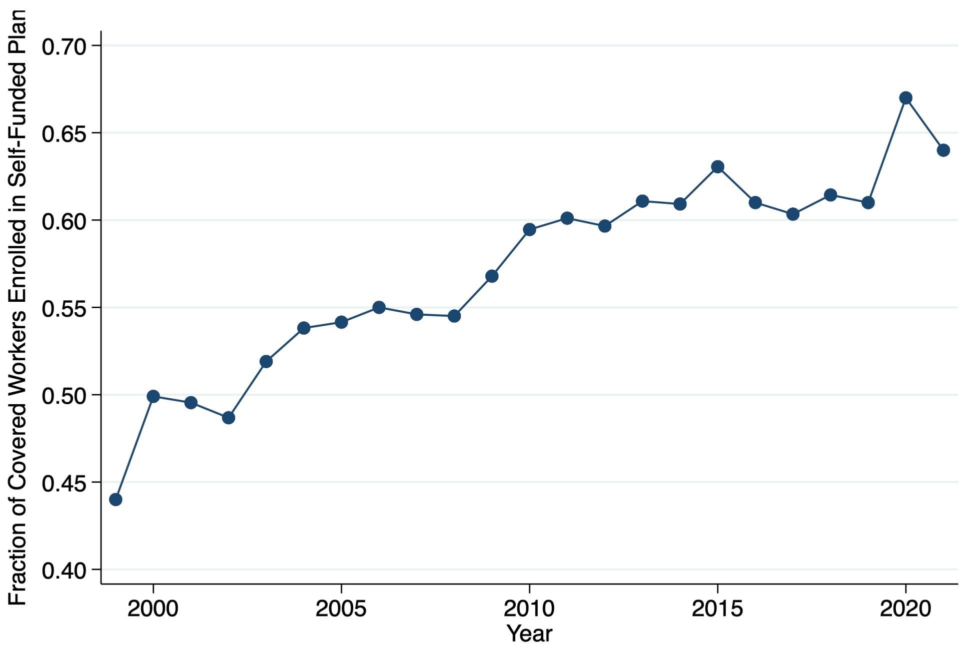
1.8 Tables & Figures

Table 1.1: Summary Statistics 1999-2008

	All Firms	100-249	250+
(a) All firms:			
Observations	354,510	143,468	211,042
Firms	64,025	29,095	34,930
Years in sample	5.54 (3.32)	4.93 (3.10)	6.04 (3.41)
Employees (1^{st} year)	1,031.13 (5,379.52)	161.40 (40.98)	1,915.87 (7,538.25)
Employees	1,536.62 (6,833.25)	239.78 (1,514.93)	2,418.22 (8,657.66)
Offers health	0.88 (0.33)	0.92 (0.27)	0.85 (0.36)
(b) Firms that offer health:			
Observations	309,977	131,344	178,633
Firms	57,513	28,966	28,547
Years in sample	5.39 (3.21)	4.53 (2.99)	6.26 (3.19)
Employees (1^{st} year)	1,032.43 (5,386.18)	161.50 (40.97)	1,916.15 (7,542.96)
Employees	1,618.47 (7,045.16)	234.24 (1081.61)	2,636.26 (9,100.79)
Self-funded	0.26 (0.42)	0.22 (0.41)	0.29 (0.43)

Notes: This table provides summary statistics for (a) all firms that report offering any welfare benefits through the Form 5500, and (b) firms that offer health coverage. Statistics are shown separately for firms that have +/− 250 employees in the first year they are observed. When categorizing by *contemporaneous* size, 82.1% of smaller firms and 91.1% of larger firms offer health coverage.

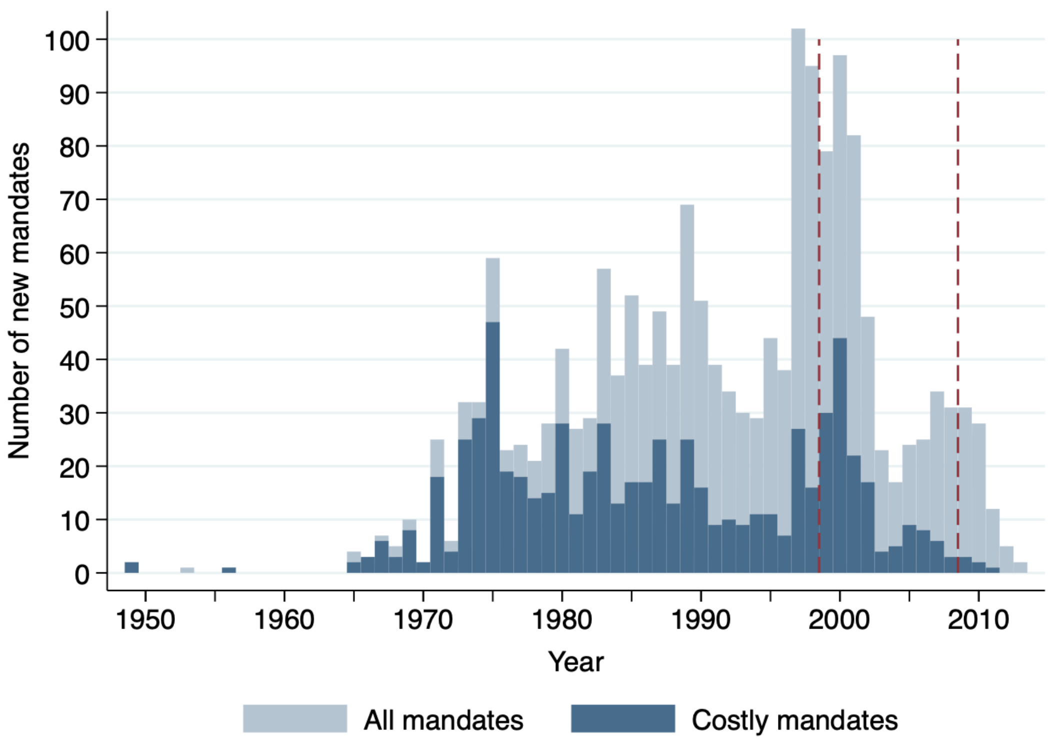
Figure 1.1: Prevalence of Self-Funding Over Time



Notes: This figure shows the percent of workers who are enrolled in self-funded plans, among all workers who are covered by an employer-sponsored health plan.

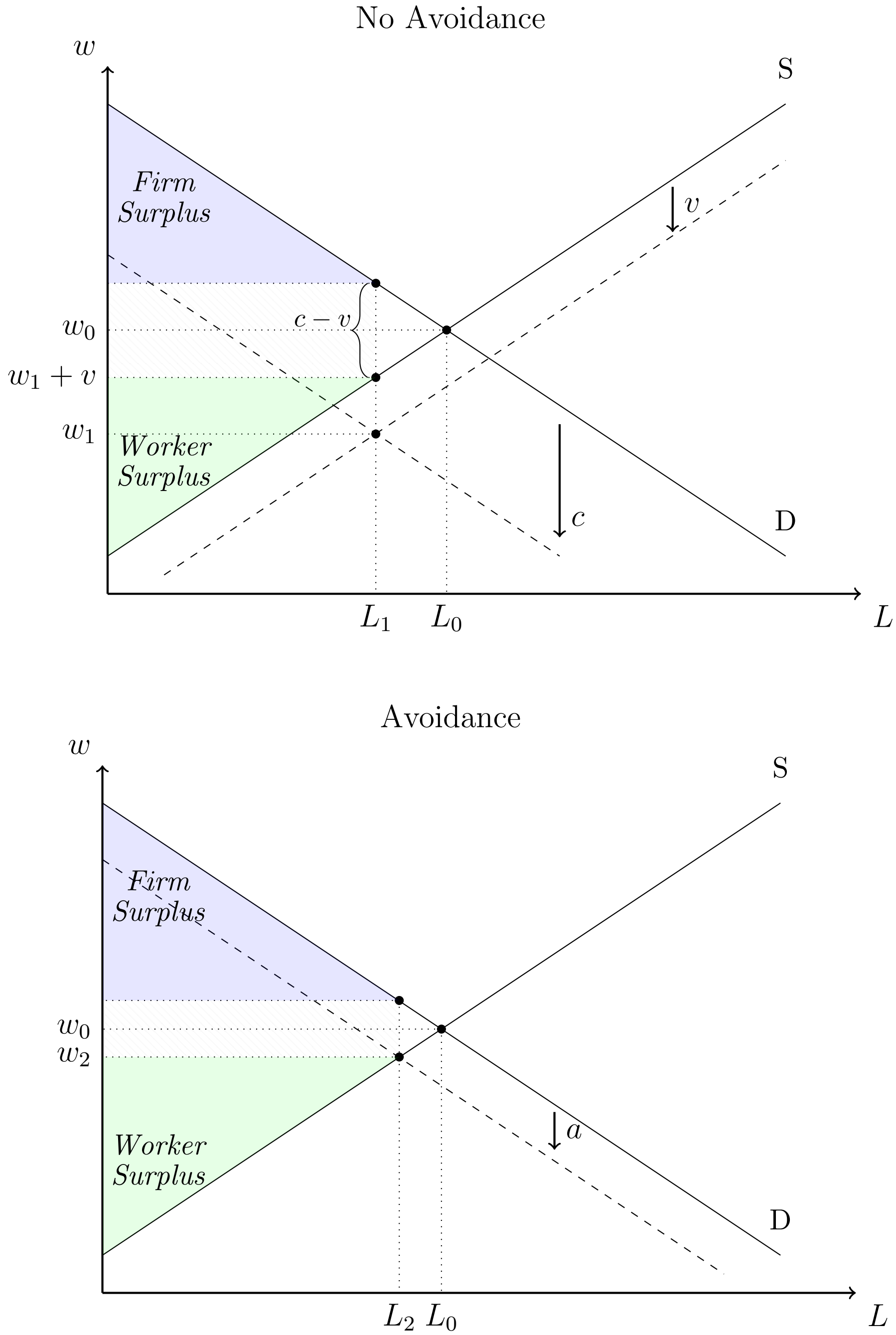
Source: Kaiser Family Foundation Employer Health Benefits Survey

Figure 1.2: New Mandated Benefits Over Time



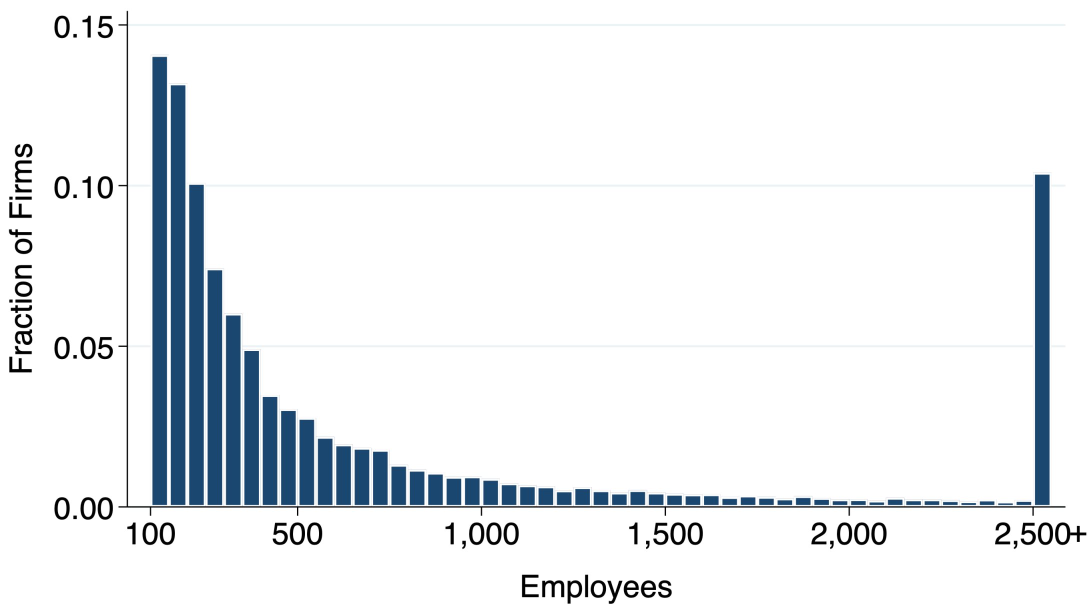
Notes: This figure shows the number of new mandated benefits in the U.S. in each year. The total number of new mandates is shown, as well as the number of new mandates that are expected to raise premiums by 1% or more. This paper focuses on costly mandates passed in years 1999-2008.

Figure 1.3: Simple Theoretical Framework



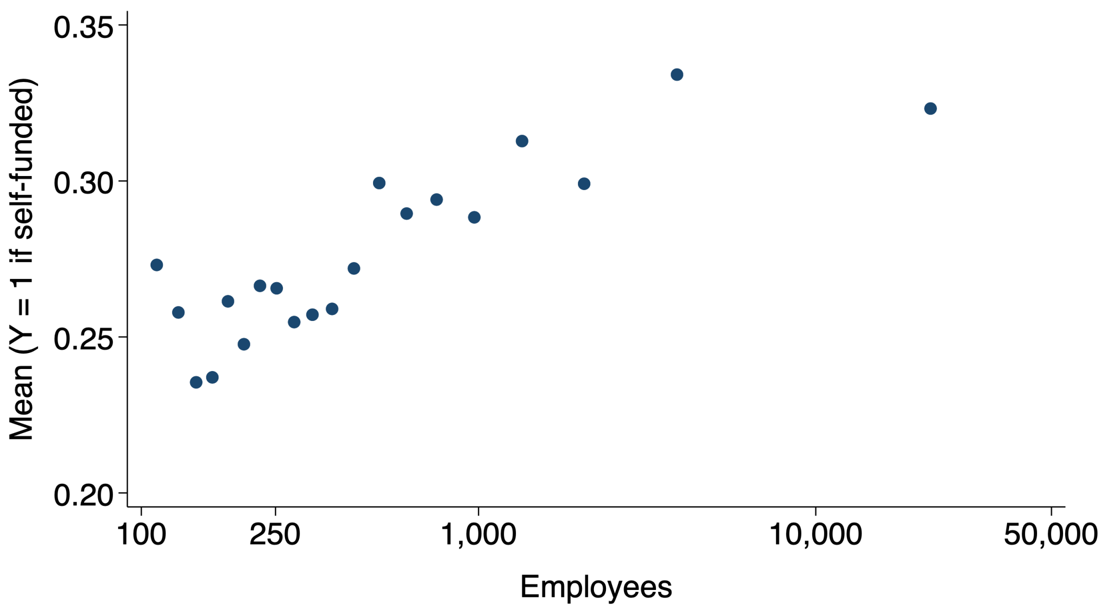
Notes: This figure shows (a) a simple model of mandated benefits as in Summers (1989), for a mandate that costs firms c to provide and is valued by workers at v. I extend this model in (b) to allow firms to avoid complying with the mandate by paying a.

Figure 1.4: Distribution of Firm Size



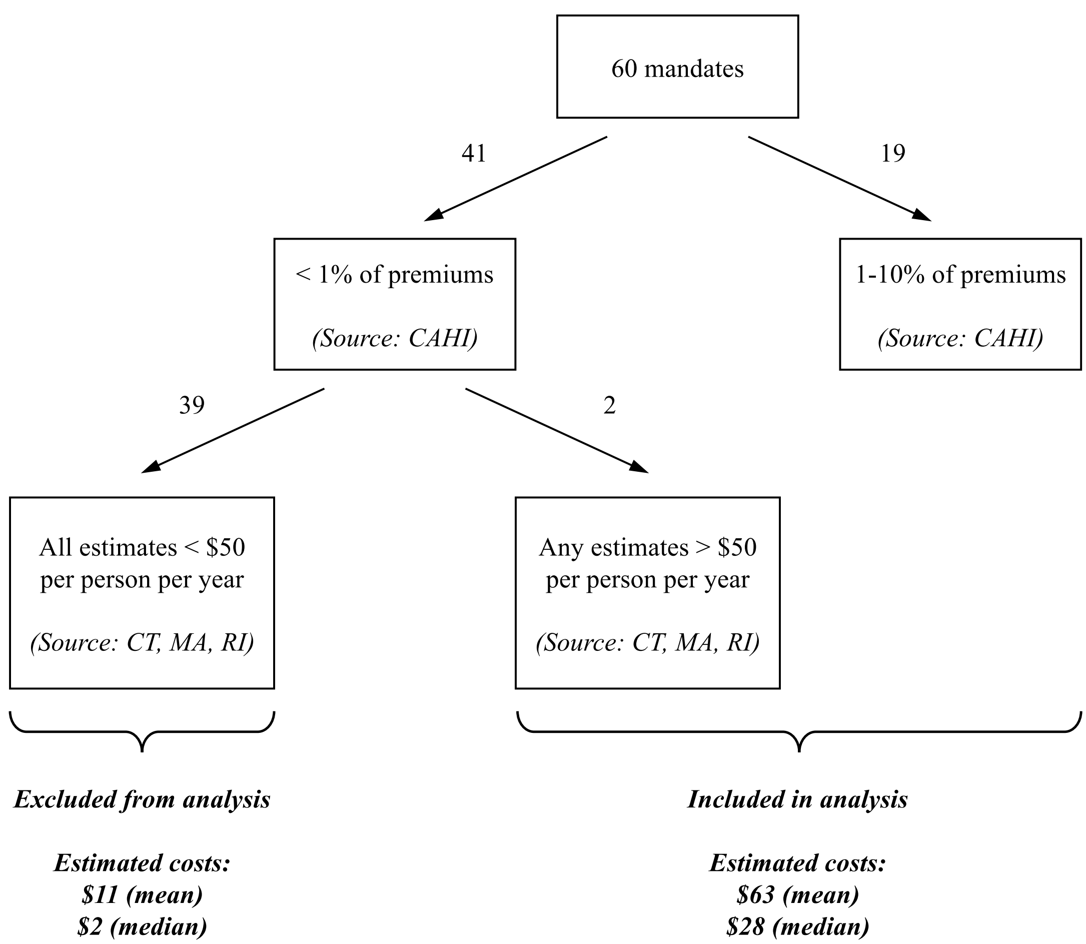
Notes: This figure shows a histogram of firm size in the year 1999 (first year of sample), conditional on offering health coverage. The last bin includes all firms with 2,500 or more employees.

Figure 1.5: Self-Funding Rates by Firm Size



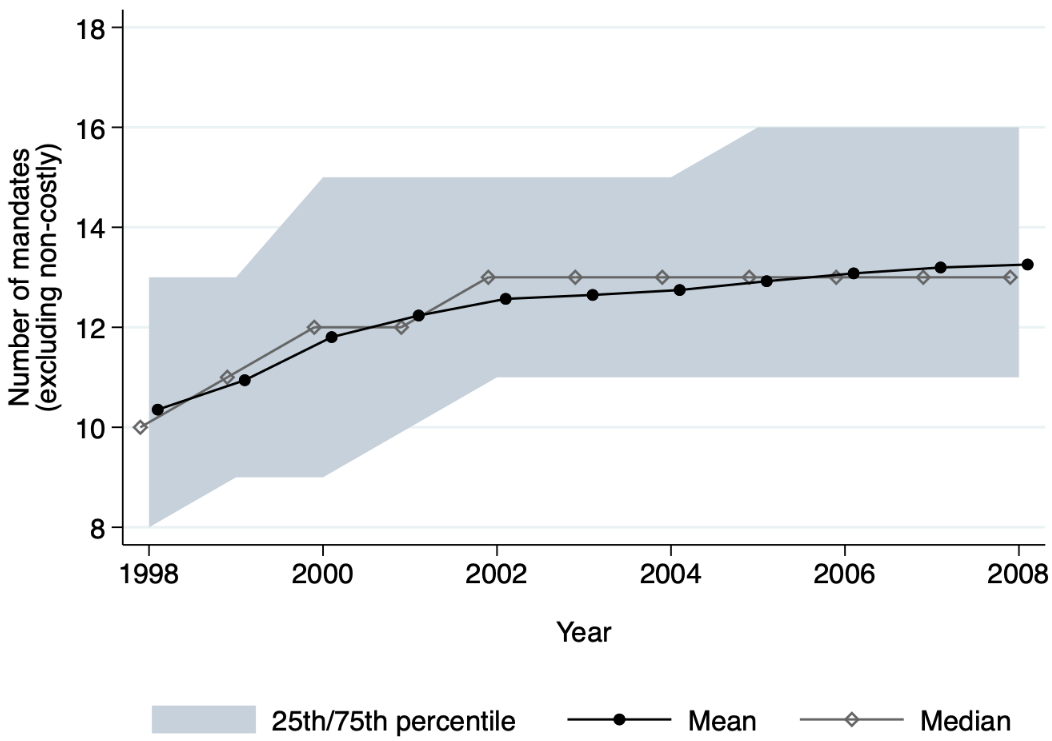
Notes: This figure shows a binscatter of firm size and self-funding rates in the year 1999 (first year of sample), conditional on offering health coverage.

Figure 1.6: Identifying Costly Mandates



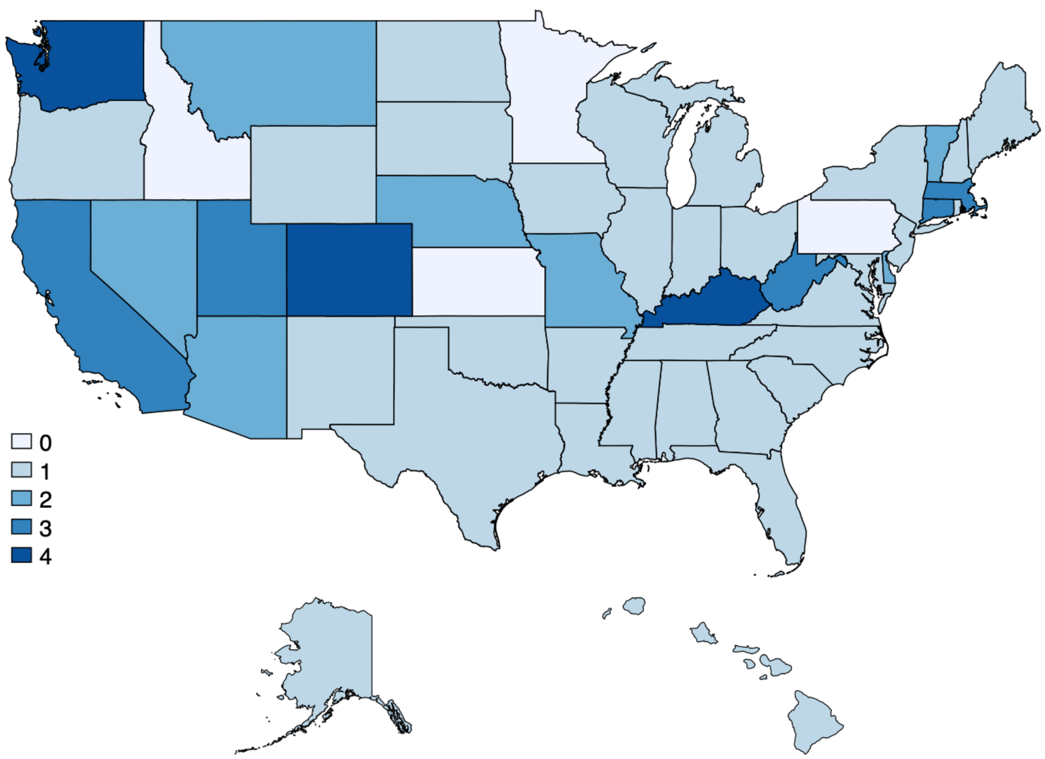
Notes: This figure shows the process for identifying costly mandates. I include all mandates that are estimated by the Council of Affordable Health Insurance (CAHI) to increase premiums by at least 1%. For the remaining mandates, I look for any estimate in three state reports (Connecticut in 2009, Massachussets in 2013, and Rhode Island in 2014) that the mandate will cost more than $50 per person per year.

Figure 1.7: Mandates Over Time



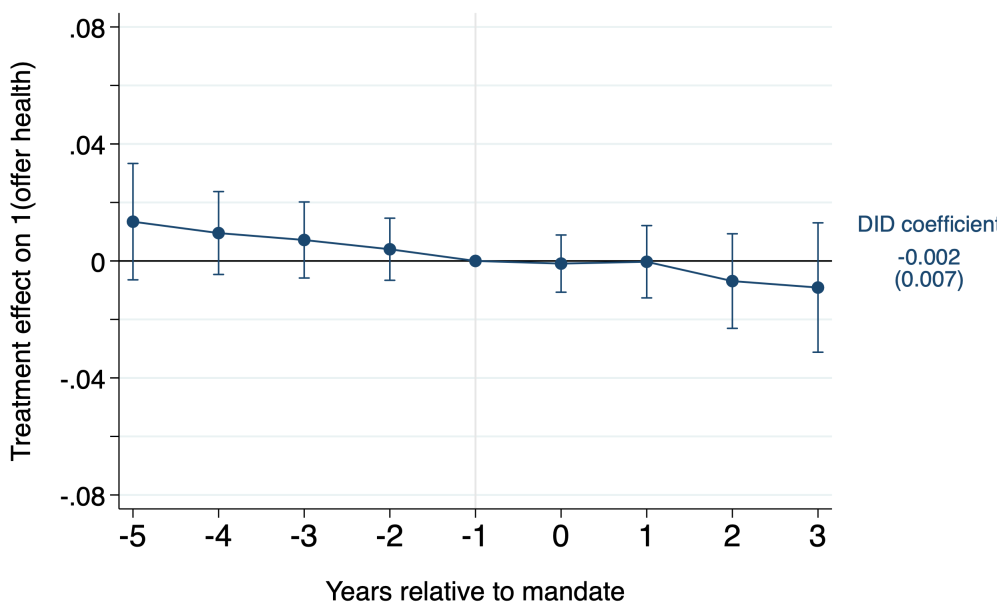
Notes: This figure shows the distribution of mandates across states over time. In each year, the mean number of mandates across states is shown. The median, 25th percentile, and 75th percentile are also shown. Only the costly mandates used in the analysis are included.

Figure 1.8: Variation in Number of Mandates Passed in Treatment Year



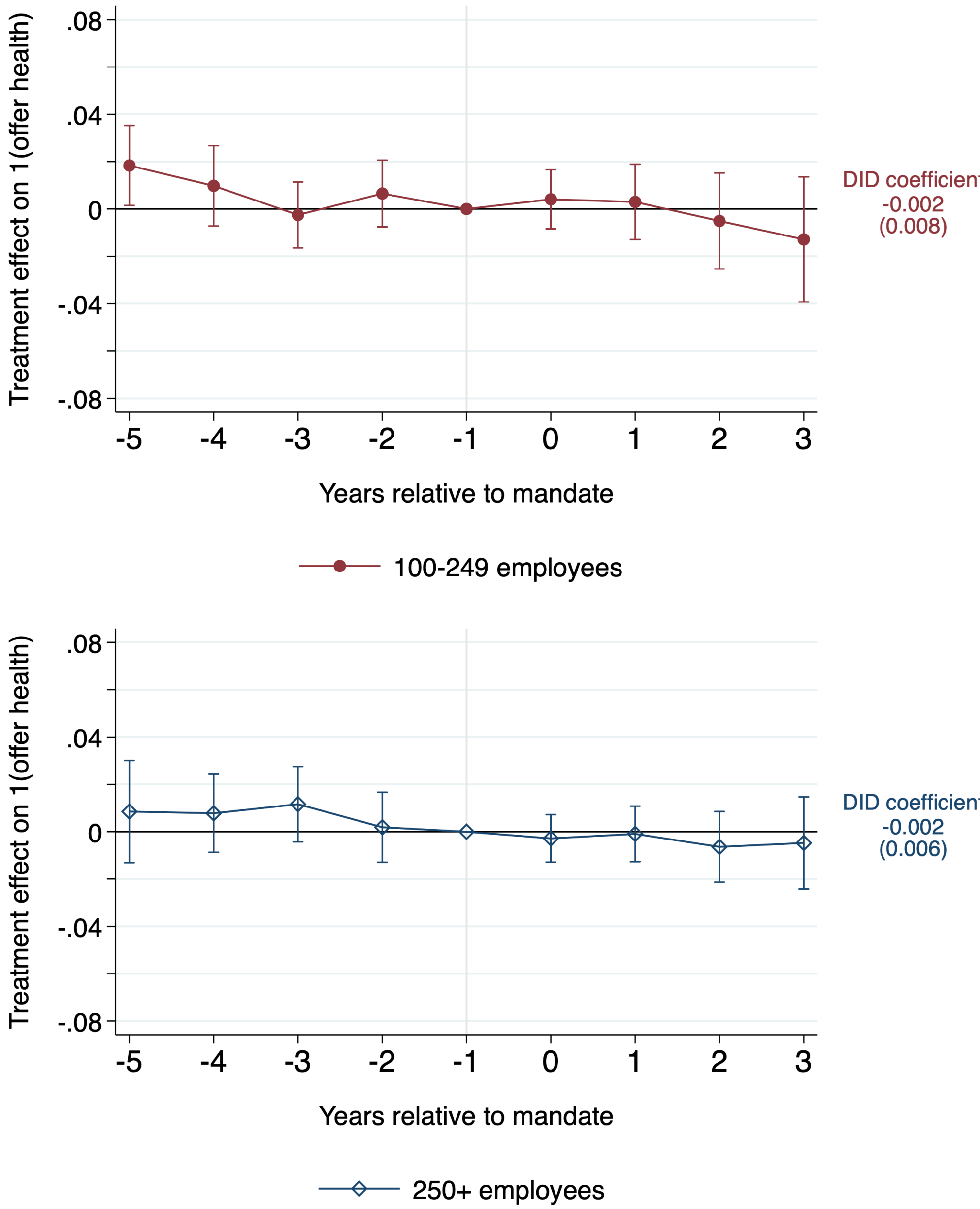
Notes: This figure shows the variation by state in the number of mandates passed in the treatment year. 31 states pass only one mandate in the treatment year; 8 states including DC pass two mandates in the same year; 5 states pass three mandates; and 3 states pass four mandates. There are 4 states that do not pass any mandates in the study period (never treated).

Figure 1.9: Effect of Mandates on Offering Health Coverage



Notes: This figure shows the estimated effect of mandated benefits on whether firms offer any health coverage. The sample includes all firms that report offering any welfare benefits through the Form 5500. Event study and difference-in-differences estimates are from a regression that includes state and year fixed effects, and standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

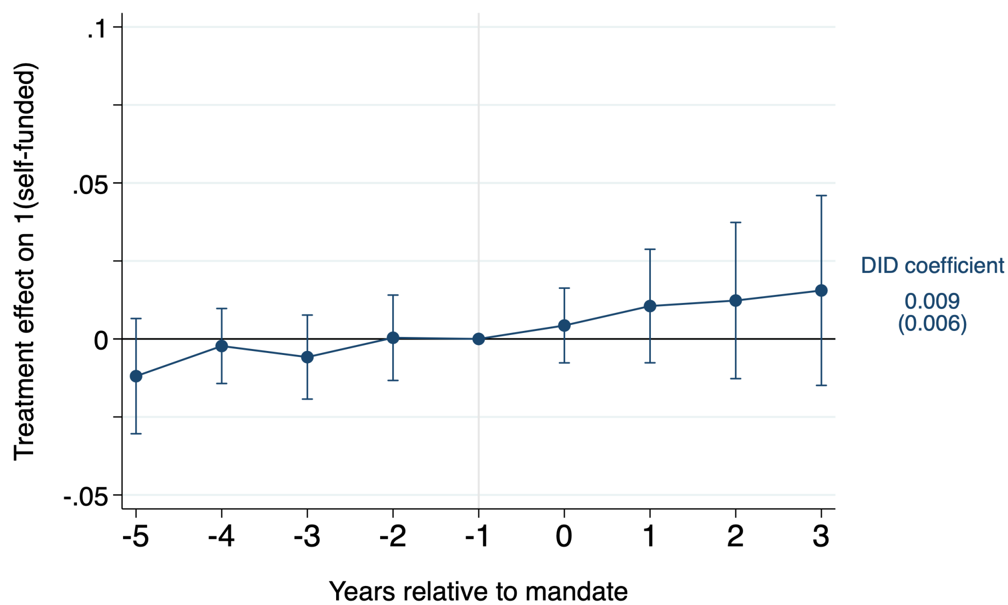
Figure 1.10: Effect of Mandates on Offering Health Coverage by Firm Size



Notes: This figure shows the estimated effect of mandated benefits on whether firms offer any health coverage, separately for smaller and larger firms. The sample includes all firms that report offering any welfare benefits through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.

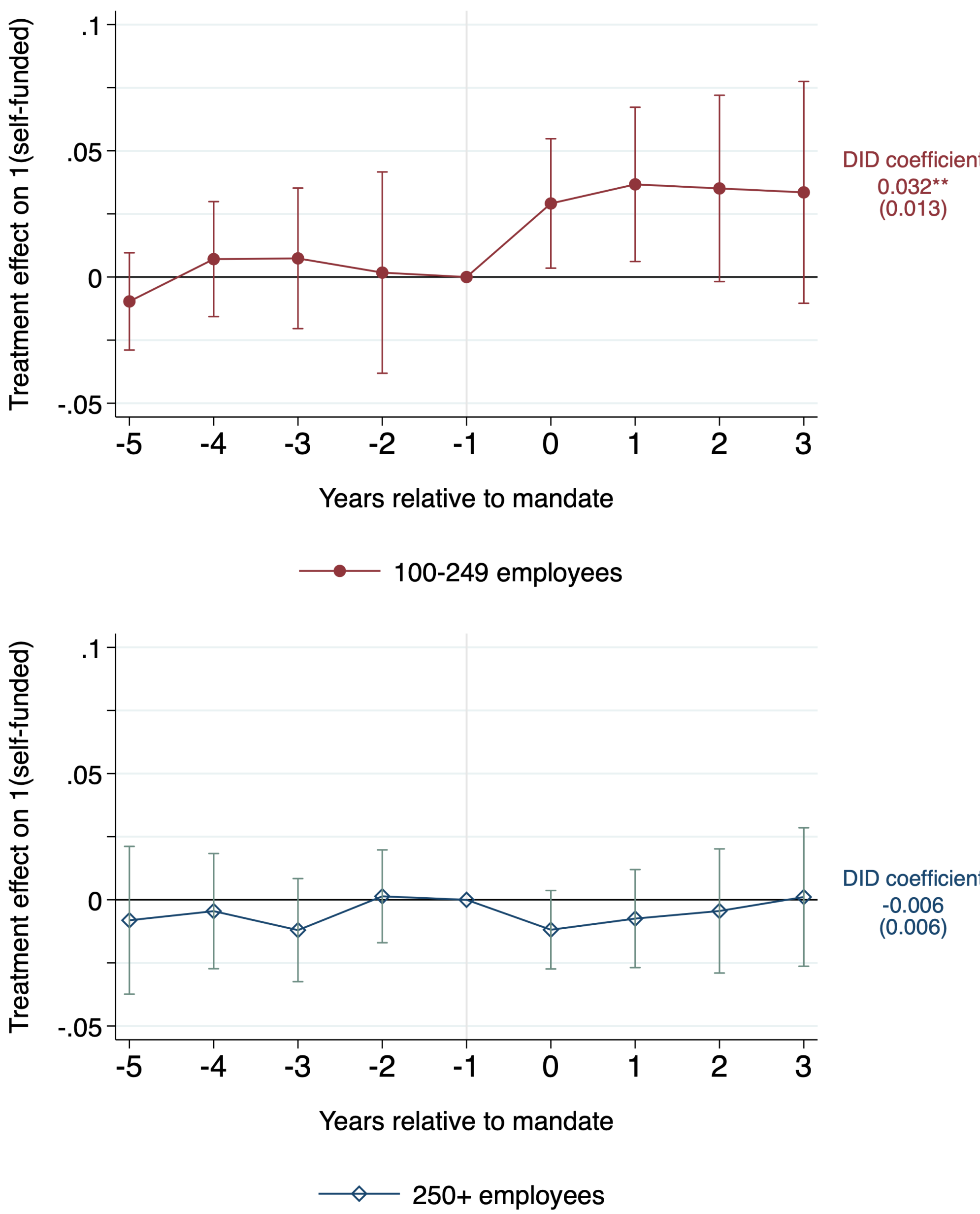
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.11: Effect of Mandates on Self-Funding



Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage. The sample includes all firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that includes state and year fixed effects, and standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

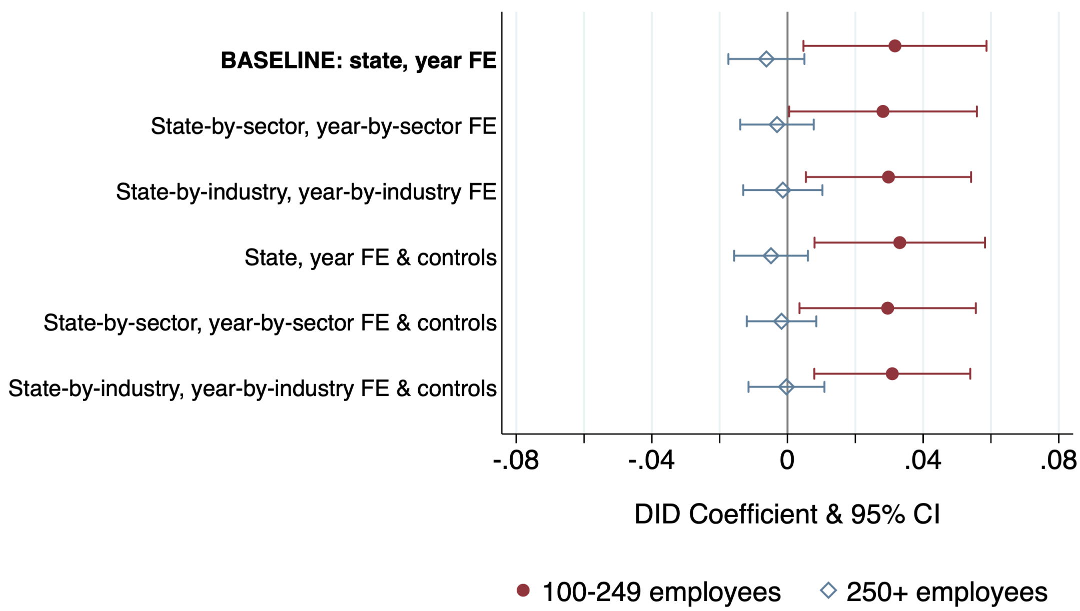
Figure 1.12: Effect of Mandates on Self-Funding by Firm Size



Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, separately for smaller and larger firms. The sample includes all firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.

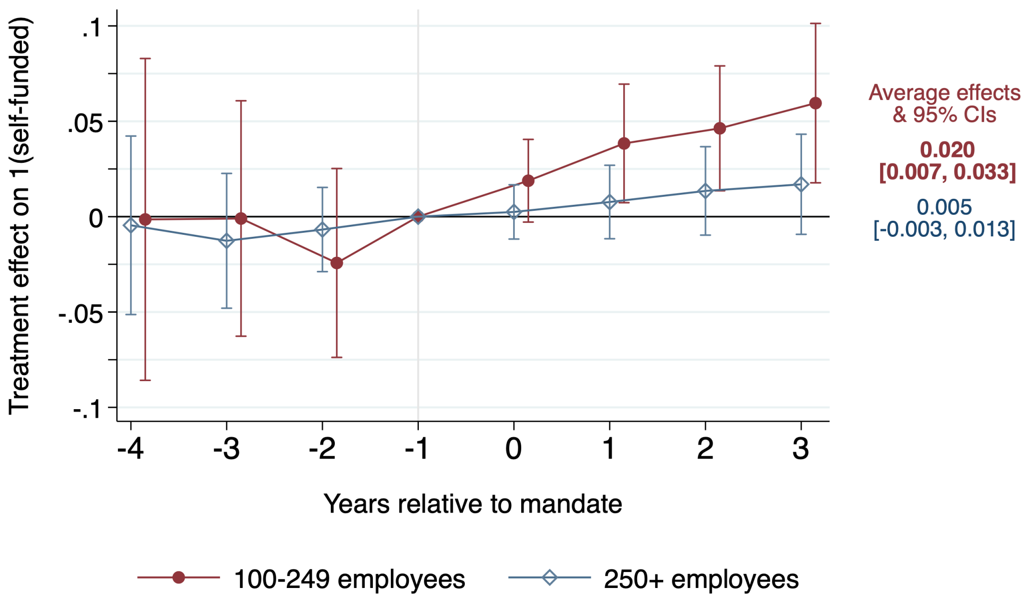
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.13: Effect of Mandates on Self-Fundingwith Additional Fixed Effects and Controls



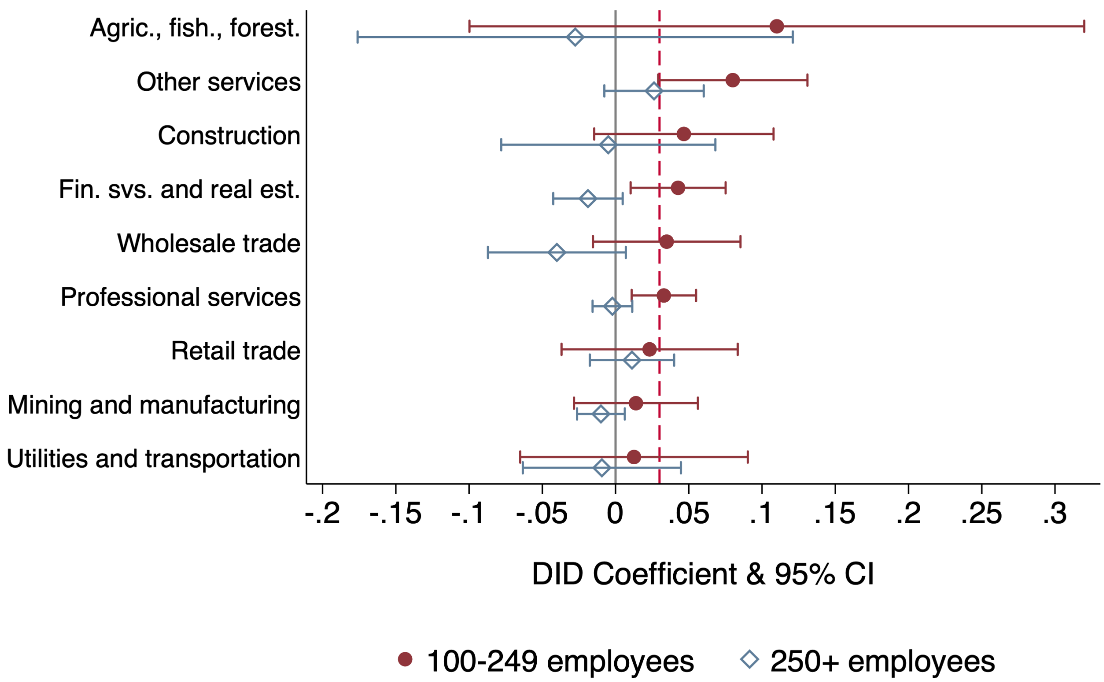
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage. For each specification, difference-in-differences estimates and 95% confidence intervals are shown for smaller and larger firms. All estimates are from a regression that interacts treatment with firm size category and controls for size category, and standard errors are clustered at the state level. Sector is defined by the 3-digit NAICS code, and industry is defined by the 6-digit NAICS code. Controls refer to the number of (contemporaneous) employees at the firm as well as the number of negligible cost (excluded) mandates in each state.

Figure 1.14: Robust DID Estimation



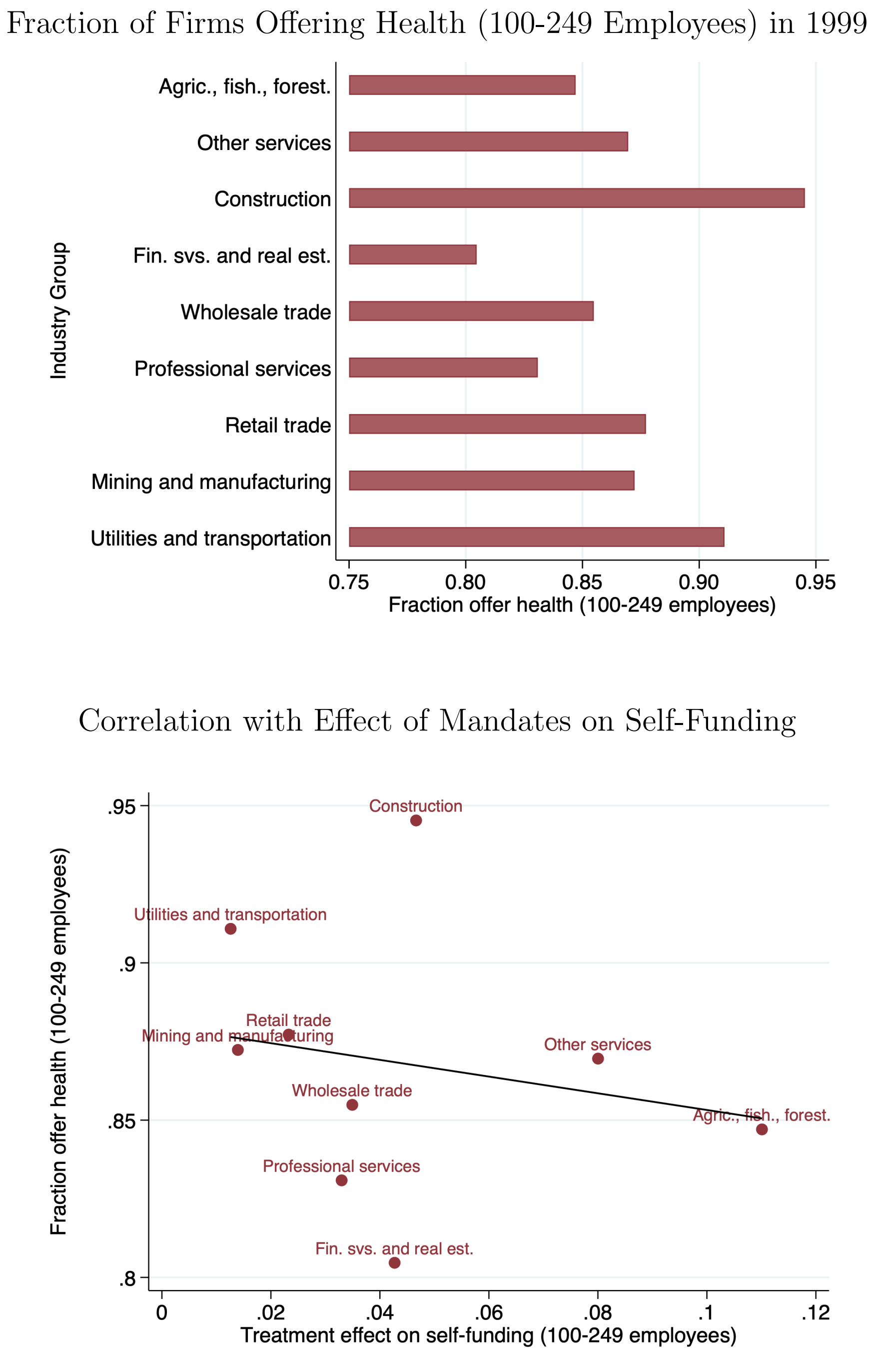
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, using the robust estimator from de Chaisemartin and D'Haultfœuille (2022). Results are estimated separately for small firms and for large firms, among firms that report offering health coverage through the Form 5500. The treatment variable is the number of costly mandates in a year, beyond the number existing in 1998. Standard errors are clustered at the state level, and estimated using 300 bootstrap replications.

Figure 1.15: Effect of Mandates on Self-Funding by Industry



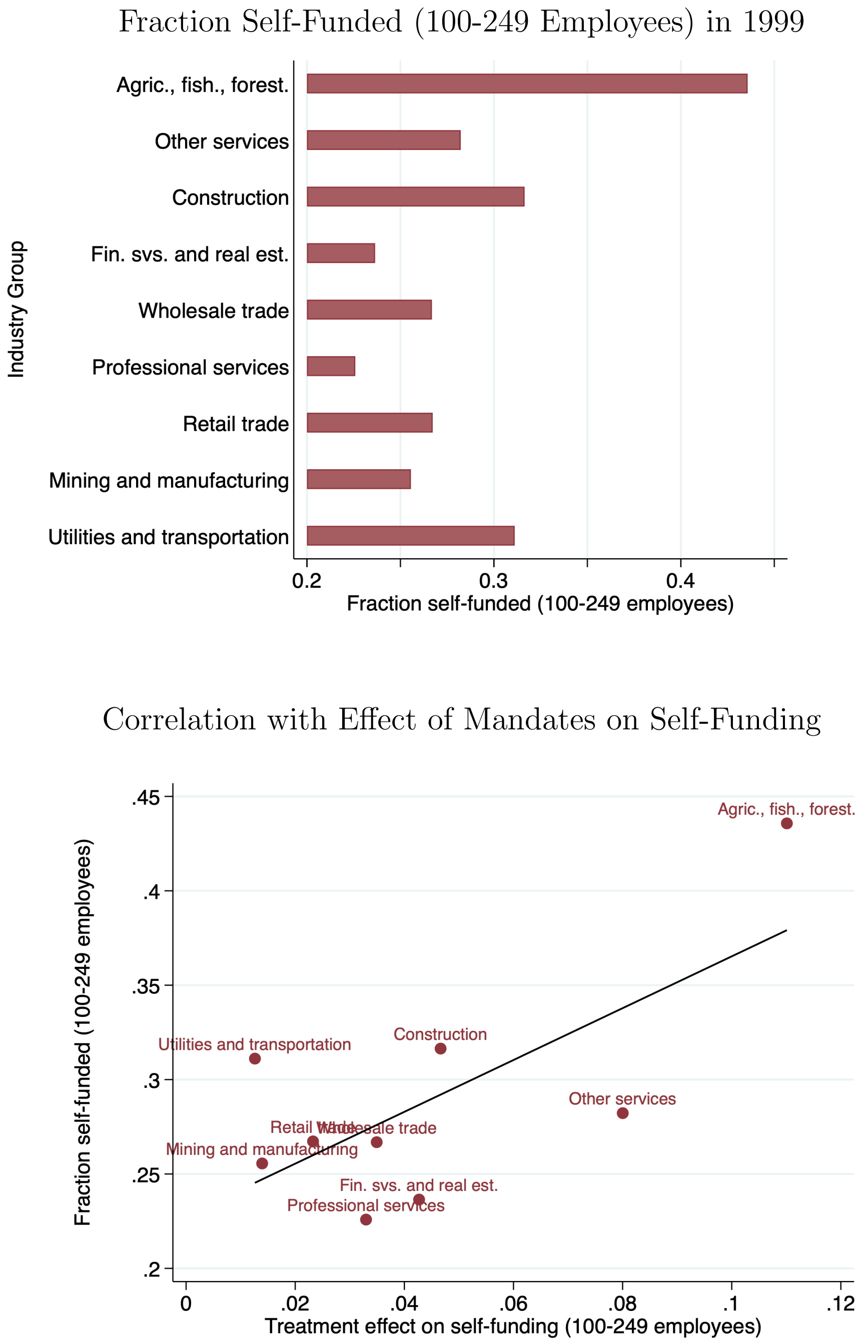
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage. For each specification, difference-in-differences estimates and 95% confidence intervals are shown for smaller and larger firms. All estimates are from a regression that interacts treatment with firm size category and industry, and controls for size category and industry. Standard errors are clustered at the state level. The red dotted line marks the difference-in-differences estimate for small firms overall.

Figure 1.16: Industry Heterogeneity in Offering Health



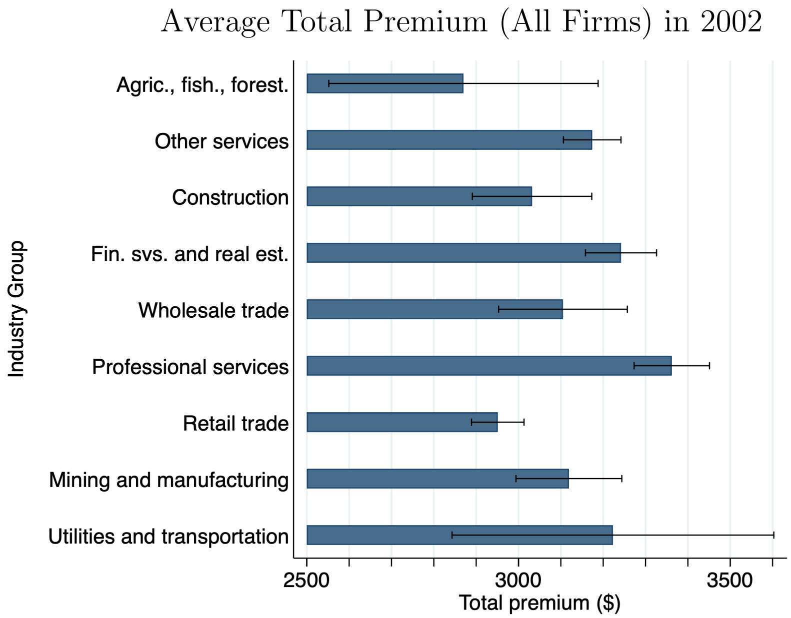
Notes: This figure shows (a) the fraction of firms (100-249 employees) that offer health by industry in 1999, and (b) the relationship with the estimated effect of mandates on self-funding for firms with 100-249 employees.

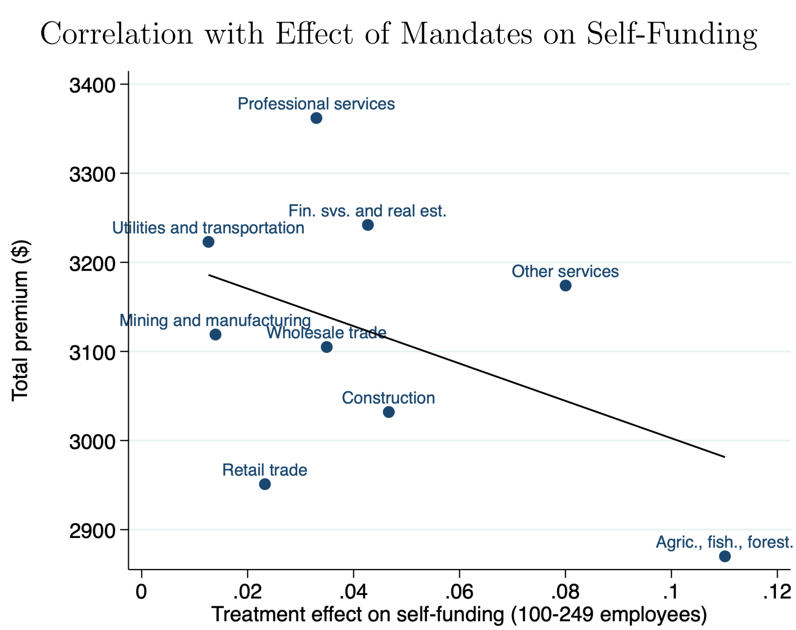
Figure 1.17: Industry Heterogeneity in Self-Funding



Notes: This figure shows (a) the fraction of firms (100-249 employees) that are self-funded by industry in 1999, and (b) the relationship with the estimated effect of mandates on self-funding for firms with 100-249 employees.

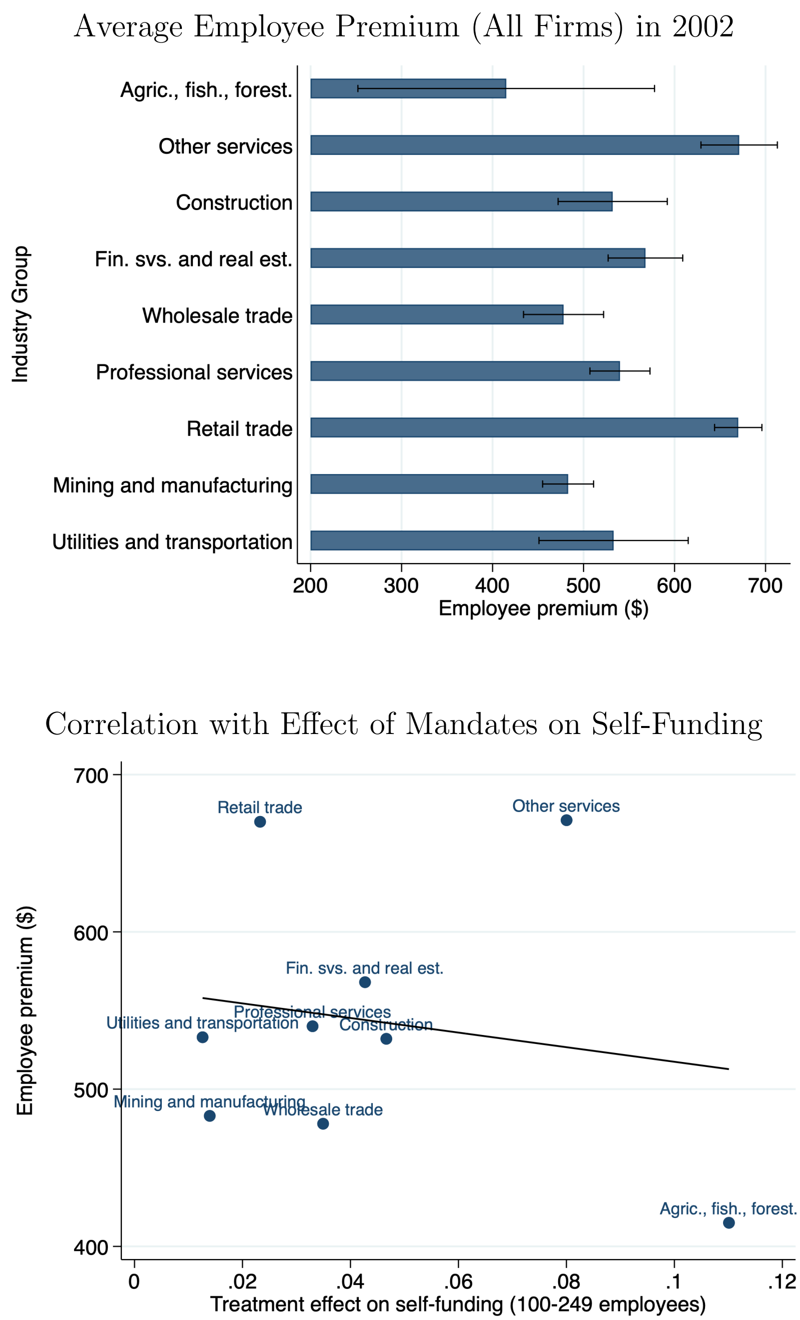
Figure 1.18: Industry Heterogeneity in Total Premiums





Notes: This figure shows (a) the average total premium for all firms by industry in 2002, and (b) the relationship with the estimated effect of mandates on self-funding for firms with 100-249 employees. Average total premiums and 95% confidence interval are from the Medical Expenditure Panel Survey – Insurance Component (AHRQ, 2021).

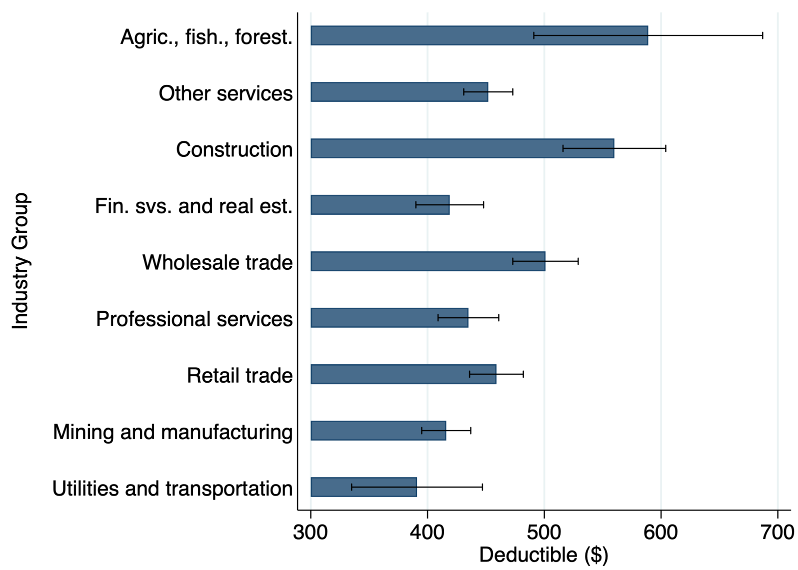
Figure 1.19: Industry Heterogeneity in Employee Premiums



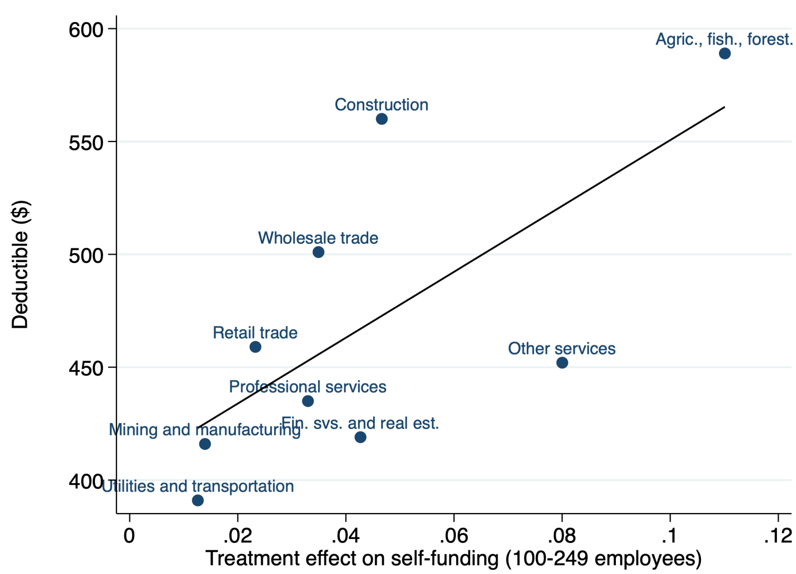
Notes: This figure shows (a) the average employee premium contribution for all firms by industry in 2002, and (b) the relationship with the estimated effect of mandates on self-funding for firms with 100-249 employees. Average employee premiums and 95% confidence interval are from the Medical Expenditure Panel Survey – Insurance Component (AHRQ, 2021).

Figure 1.20: Industry Heterogeneity in Employee Premiums

Average Deductible (All Firms) in 2002

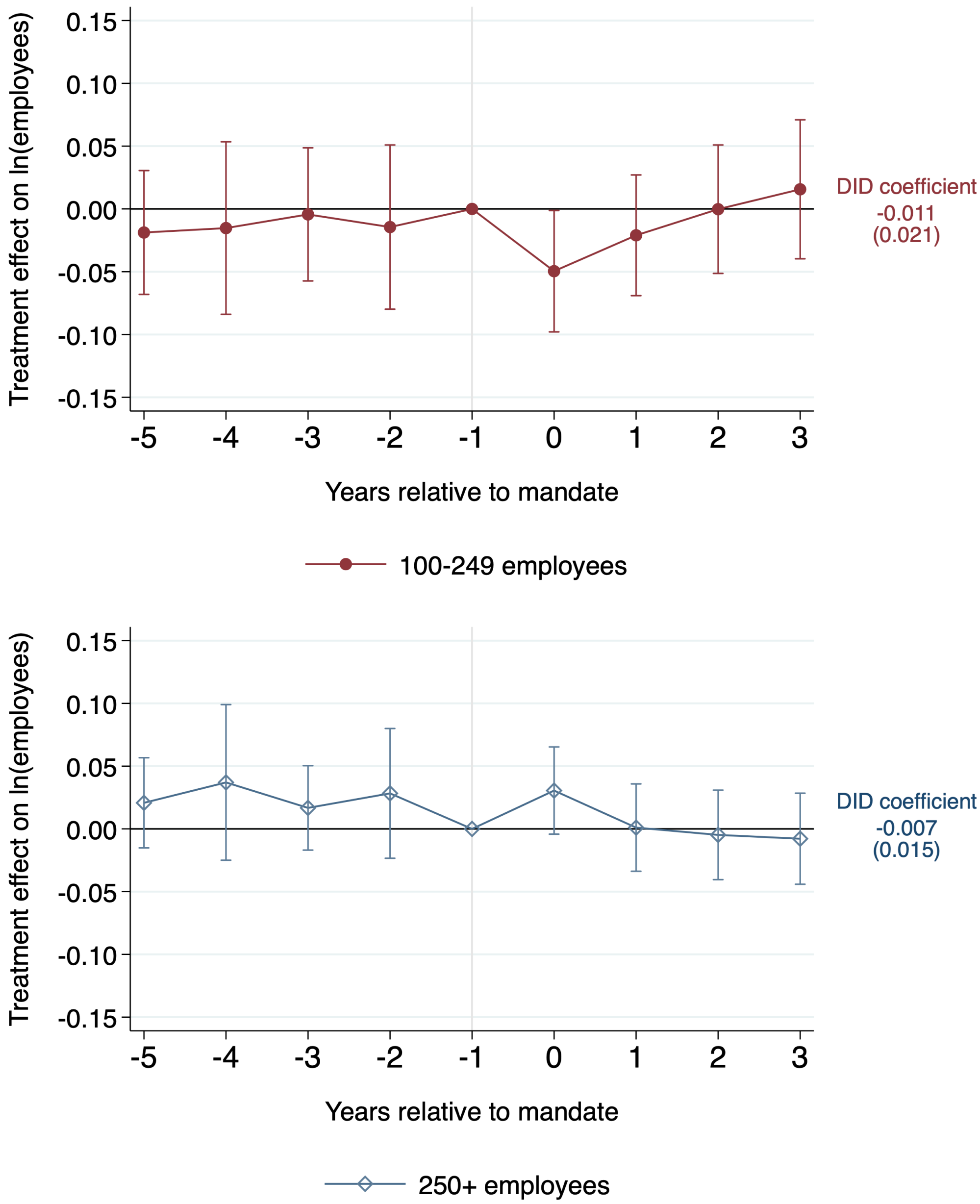


Correlation with Effect of Mandates on Self-Funding



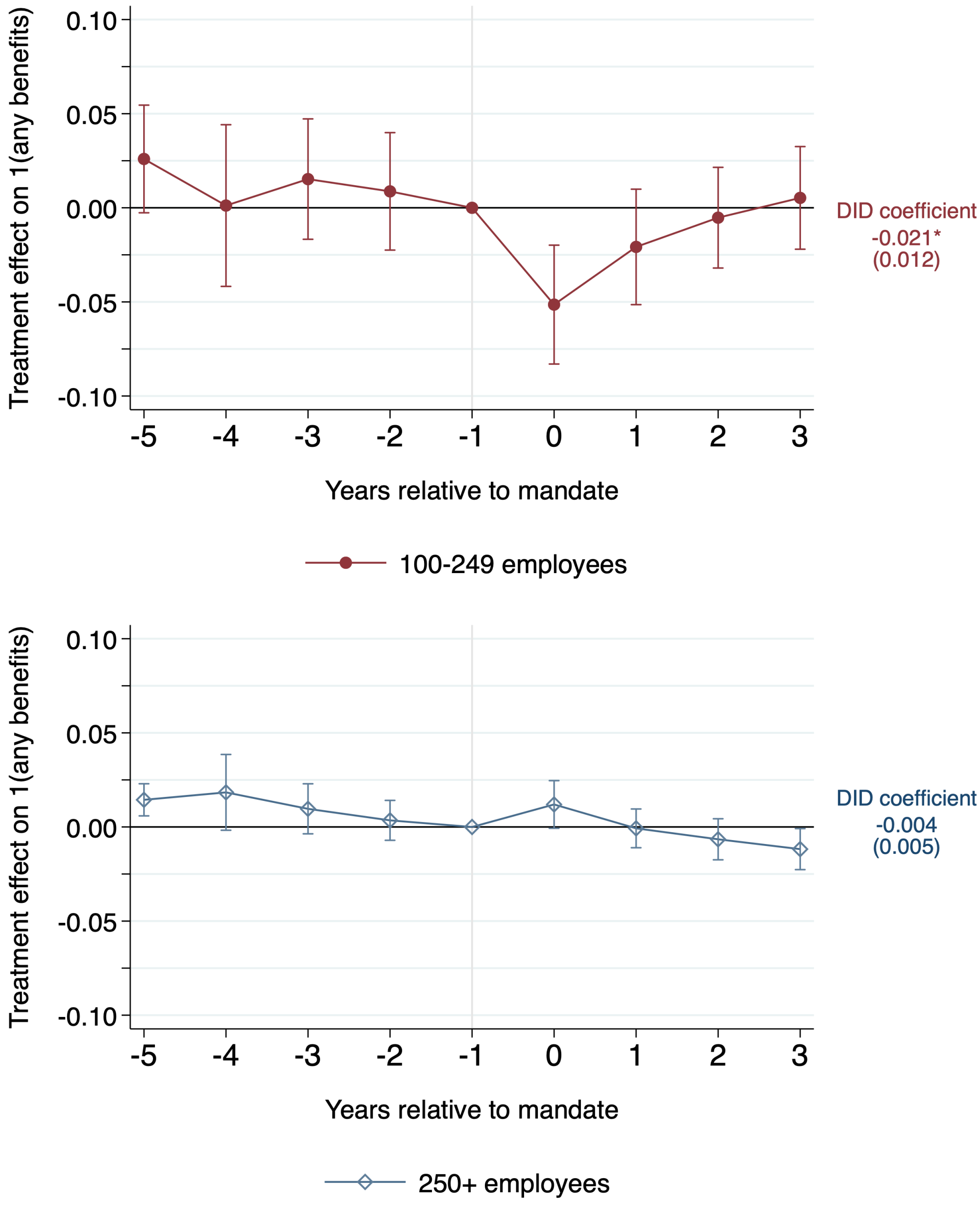
Notes: This figure shows (a) the average deductible for all firms by industry in 2002, and (b) the relationship with the estimated effect of mandates on self-funding for firms with 100-249 employees. Average deductible and 95% confidence interval are from the Medical Expenditure Panel Survey – Insurance Component (AHRQ, 2021).

Figure 1.21: Effect of Mandates on Employment by Firm Size



Notes: This figure shows the estimated effect of mandated benefits on (log) employment, separately for smaller and larger firms. The sample includes all firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1.22: Effect of Mandates on Any Benefits by Firm Size



Notes: This figure shows the estimated effect of mandated benefits on whether firms offer any benefits, separately for smaller and larger firms. The sample includes all firms that ever report welfare benefits through the Form 5500. Firms not reporting benefits could either not offer benefits, or not exist. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 2

What Drives Tax Policy? Political, Institutional and Economic Determinants of State Tax Policy

Taxation is at the heart of redistribution and can be a powerful tool for correcting market failures and smoothing business cycles. However, increasing political polarization and legislative gridlock in the U.S. has made it difficult to achieve these goals using fiscal policy.[1] While these issues have been studied on a national level, much less is known about the extent to which they pervade tax policy at the state level, despite its great importance – the U.S. states raise large tax revenues (over $1T or 5% of U.S. GDP each year) and provide a wide range of services and welfare benefits to their residents.

The goal of this paper is to provide a comprehensive analysis of the plausible determinants of U.S. state tax policy, focusing both on long-term trends and the actual timing of policy changes. We start by evaluating the direction of tax trends over the past 70

[1]For evidence on political polarization see McCarty, Poole and Rosenthal (2016). For empirical evidence on policy uncertainty and legislative gridlock, see e.g. Binder (2004), Baker et al. (2014), Mian, Sufi and Trebbi (2014), Aizenman et al. (2021).

years, including the degree of convergence or divergence across states, and compare these outcomes to the long-term predictions of fiscal federalism models. We further examine the frequency of policy changes, the degree to which changes in one tax overlap with changes in another, and the persistence of rates over time.

Next, we use permutation analysis, variance decomposition, and machine learning techniques to evaluate to what extent the *timing and magnitude* of tax changes are driven by incentives prominently featured in economic models. Broadly speaking, public economists highlight the importance of taxes for redistribution, revenue collection, and addressing externalities, while macroeconomists employ taxes as a policy stabilization tool. Political economists call attention to the importance of voter preferences, as well as political and institutional frameworks in tax setting processes. Finally, fiscal federalism models spanning all fields stress the importance of state competition on policy outcomes. These theoretical factors map well into policy motivations of the legislators: Romer and Romer (2010) show that at the federal level, most tax changes have a "clearly identifiable motivation that falls into one of four broad categories: offsetting a change in government spending; offsetting some factor other than spending likely to affect output in the near future; dealing with an inherited budget deficit; and achieving some long-run goal, such as higher normal growth, increased fairness, or a smaller role for government." Motivated by this, we consider *economic influences*, such as competition and changing revenue requirements due to economic downturns or federal mandates; *political influences*, such as election cycles, composition and changes of political powers within the state; *institutional features*, such as the size of state legislatures, term limit requirements, balanced budget and voter initiative rules; *demographic influences*, such as population measures, labor participation rates, poverty measures and demographic compositions; and the relationship between federal and state tax policies.

The comprehensive nature of our approach and the flexibility of the machine learning algorithms we employ thus allow us to evaluate the extent to which the factors most frequently featured in economic models of tax policy, taken together, can explain the tax-setting processes. We formally show that in a broad set of policy setting models and as long as policymakers are not evenly split, tax policy should be highly predictable even if policymakers' preferences are somewhat idiosyncratic. Empirically, this implies that a flexible machine learning algorithm incorporating the relevant explanatory variables should have a high predictive power. Therefore, a low explanatory power – as we find – implies that either tax policy has a large idiosyncratic component, or that relevant explanatory factors have been omitted from the model. Our results do not imply that the factors we consider are not important, merely that other factors may have even a larger influence on tax policy, suggesting a need for future work.

For our analysis, we have collected detailed information on state personal income, corporate, sales, cigarette, gasoline and alcohol taxes, from 1950 until 2020. We focus on these taxes because they are primarily controlled by state, rather than local, governments, and combined represent approximately 80% of state tax revenues. Since tax policies are multi-dimensional, in our analysis we focus on six key parameters – the top personal income tax rate, top corporate tax rate, standard sales tax rate, cigarette tax per pack, gasoline tax per gallon, and spirit tax per gallon. By focusing on (top) statutory rates, our analysis centers on tax features that are important for inequality considerations and that are likely to be most salient to voters. To ensure our results are as comprehensive as possible, we also collect information on other features of tax policy: the first-bracket personal income tax rate, married exemption, bottom and top income brackets, state EITC rates, deductibility of federal income taxes, sales tax inclusion rules, and various corporate tax features: e.g., number of years allowed for loss carry back and carry forward, apportionment weights (payroll, property and sales), and minimum corporate tax rate. We treat each state as an individual decision maker and for this reason do not weigh results by population.

Our analysis generates three key insights. First, focusing on the long-term trends, we show that tax rates exhibited a period of rapid convergence in 1950-1980s, which was primarily fueled by the adoption of new taxes by states. In the most recent 30 year period, however, all six tax rates have exhibited stable levels of variance, and have neither been converging nor diverging over time.[2] Our results are consistent with and complementary to the findings of Rhode and Strumpf (2003) who document a substantial convergence in state policies (mainly tax expenditures) over the 20th century but show a similar level of policy heterogeneity during the last 30 years of the century. As shown by Rhode and Strumpf (2003), the observed trends thus do not lend support to Tiebout-sorting models (which predict divergence of tax rates in the presence of lower mobility costs) or of race-to-the-bottom competition models (which typically predict convergence). Overall, long-term trends suggest that competition forces – while important – are unlikely to be the primary drivers of tax policies.

Second, despite the relative stability of average rates, states implemented many tax rate changes during the studied period. In an average year, 20 states changed at least one tax rate. Furthermore, states frequently change more than one tax rate at a time: 36% of state tax changes involve changes of two or more tax rates (13% of changes involve three or more rates). We show that states vary dramatically in how frequently they change tax rates, with more frequent changers favoring smaller tax changes. While sales and excise taxes followed a well-defined trend in nearly all states over time, income and corporate tax trends vary, with states frequently exhibiting fluctuating patterns. We see some persistency in tax rate levels, but overall conclude that the magnitude of tax changes appears to be rather unpredictable. With the exception of personal income

[2] Our results are robust to using various measures of convergence, e.g. coefficient of variation (CV) defined as the ratio of the standard deviation to the mean or simple standard deviation.

tax, the correlation between the rate in 1950 and the rate in 2020 is less than 20%. While Democratic-leaning states tend to have higher tax rates on average, the rates of Democratic- and Republican-leaning states largely overlap.

Third, we show that the timing and magnitude of tax changes are difficult to predict, suggesting that either taxes are not legislated "in response to" economic and political events as assumed in economic models, or that the response is often untimely, perhaps, due to legislative gridlock. We start by using permutation techniques to investigate what share of tax changes follow an event of interest highlighted in economic models: a recession (as macroeconomic stabilization tool), the introduction of an unfunded federal mandate (externally driven increase in spending), a neighboring state's tax change (competition), or a change of majority party (a change in voter preferences or political environment). We compare observed co-occurrences to a simulated benchmark that assumes the timing of tax changes is random. Our analysis shows that the rates of co-occurrences are not dramatically different from the simulated benchmark, suggesting that these events have a limited influence on the timing of tax changes (despite being prominently featured in models), or that their influence is untimely.

We continue this analysis by turning to a variance decomposition approach and machine learning techniques. Overall, we find that federal changes, economic needs, neighborly competition, institutional features, political factors, and demographics explain less than 20% of variation in the timing and magnitude of tax changes, even when employing machine learning techniques that allow for various interactions and flexible functional forms. Interestingly, variance decomposition suggests that tax increases and tax decreases may be influenced by different factors. For example, tax increases are substantially more influenced by federal tax policy than tax decreases. Similarly, economic factors (recessions and mandates), neighbors' tax rates levels, and own other tax rate levels are more important for tax increases, while balanced budget provisions are more important for decreases. Our main analysis focuses on *yearly* tax rate changes, but we also show that exploring *decade* changes results in similar conclusions.

While low explanatory power is rarely of concern in economics because of researchers' focus on identifying causal relationships, it is of great interest in the setting of state tax policies. For this reason, tax policy choice process has been the focus of a large number of empirical and theoretical studies, discussed below. Our inability to explain a large share of tax fluctuations thus suggests that a wide range of existing models, even when combined, do not explain the observed policy outcomes. One possibility is that our analysis omitted important drivers of tax policy, for example, lobbying and political contributions, that play a substantially more important role than the economic, political, and institutional influences the literature has focused on. However, political literature so far has found little support for such *quid pro quo* links in general (Ansolabehere, De Figueiredo and Snyder Jr, 2003), and for tax policies specifically. For example, Slattery, Tazhitdinova and Robinson (2023) show that state tax policies did not change in response to independent corporate expenditure increases as a result of the *Citizens United* ruling. Alternatively, the legislative process for tax policy may be so complex that idiosyncratic factors create substantial randomness in the timing of policy response.[3] This implies that tax policy may be unnecessarily volatile, resulting in excess state tax revenue volatility, business cycle volatility, and policy uncertainty that can have detrimental effects on growth and the welfare of state residents (see e.g., Seegert, 2016).

Our findings are relevant for empirical researchers who rely on tax variation as a source of identification. While our results do not imply that tax changes are outright "exogenous," they do suggest that the bias from omitting institutional, political, and economic factors is likely to be small in studies that exploit sharp variation in tax changes

[3]For example, Mian, Sufi and Trebbi (2014), provide evidence of delayed government interventions in response to financial crises due to increasing polarization and resulting weakening of the ruling coalition.

and focus on short-run outcomes. Simply put, the tax setting process appears to be sufficiently complex so that the exact timing of tax changes is sufficiently random. However, longer-term estimates need to be interpreted with caution as some tax rates appear to follow a trend. Finally, researchers should be careful when attributing estimated effects to a specific tax change since many tax changes are implemented as part of a package i.e., at the same time as changes in other tax rates.

A caveat to our analysis is that, while we try to paint a comprehensive picture of state policies, these policies are very complex and hard to summarize. Because of this, we focus on explaining changes in specific features of tax policy instead of comparing effective tax burdens across states. We choose to focus on tax rates because these are most salient to voters, subject to extensive media coverage, and are changed frequently. In contrast, isolated tax base rules are changed infrequently, even though altogether they are key to understanding the amount of tax revenue a given tax generates (Suarez Serrato and Zidar (2018)). Relatedly, we only explore changes to state tax rules but ignore changes at the local level. Our empirical analysis, however, includes year or decade fixed effects, and allows for differential predictions over time through interactions with decade dummies. Thus while we are not able to measure the influence of local policies on state policies, we account for these via fixed effects. Our analysis also does not account for differences in cost of living across states. This is of particular concern for property taxes, since the property tax burden is heavily influenced by its tax base. For this reason, we do not include property taxes in our analysis.[4] On the other hand, the excise taxes that we consider – gasoline, cigarette and alcohol taxes – have a uniform tax base and are robust to this issue.

This paper is related to several lines of prior work. Our paper builds on the vast

[4]Moreover, property taxes vary across localities within states, to a much larger extent than other types of taxes. To mitigate the importance of this exclusion, we include in our predictive analysis shares of 1995 tax revenues attributed to each tax type, thus controlling for states' tax structures.

literatures that study the policy choices of the federal and local governments. This wide range of work explores fiscal competition (e.g. Besley and Rosen (1998); Rork (2003); Devereux, Lockwood and Redoano (2007)); preference-based sorting (e.g. Tiebout (1956); Rhode and Strumpf (2003); Boadway and Tremblay (2012)), the importance of political cycles and structures (e.g. Alesina, Roubini and Cohen (1997); Nelson (2000) and Alt and Lowry (1994); Bernecker (2016)), federal mandates (Baicker, Clemens and Singhal (2012)), and various institutional features, such as balanced budget provisions (Poterba (1994)), size of legislatures (Gilligan and Matsusaka (2001)), term limits (Besley and Case (1995*a*); Erler (2007)), and legislative initiative rules (Matsusaka (1995); Matsusaka (2000); Asatryan, Baskaran and Heinemann (2017)). Our work builds on these studies but differs in four dimensions: we focus on overall explanatory power instead of causal relationships, we take a comprehensive approach by considering numerous influences together instead of emphasizing a specific channel, we use machine learning techniques to allow for flexible modeling, and we focus on the timing of tax changes rather than tax levels in general.[5] Our focus on predictive power allows us to evaluate to what extent these models are able to explain the observed behavior. Consistent with previous work, we confirm that competitive, political, and institutional forces highlighted in economic models matter, but show that they explain a relatively small share of the overall tax policy fluctuations.

Furthermore, this paper builds upon a small number of studies that document basic facts about state and local tax policies. The closest study, Baker, Janas and Kueng (2020), document how state and local taxes have changed over time, while Suarez Serrato and Zidar (2018) and Slattery and Zidar (2020) provide a comprehensive overview of state business tax policies. We extend the previous work by collecting extensive data on state

[5]Our work is thus related to Ferede, Dahlby and Adjei (2015), Kakpo (2019) and Gupta and Jalles (2020), but is more comprehensive both in our approach and in scope.

tax policies, as well as on political and institutional factors.

2.1 Should Tax Policy be Predictable?

In this paper we evaluate the extent to which the timing and magnitude of tax changes are driven by economic, political, and institutional factors. To do so, we measure the share of observed variation in tax changes that can be explained by the variables that the previous literature identified as explanatory. Our approach thus raises a natural question: should tax policy be predictable, and if it is not, what does that imply? In this section, we argue that in a broad set of policy setting models, tax policy should be highly predictable even if the individual behavior of policymakers is not. As a consequence, if a sufficiently flexible econometric model has limited predictive power, then the explanatory variables included in the model are unlikely to be drivers of the policies we analyze. This implies that either other factors are at play, or the policy setting process is truly idiosyncratic.

Consider two broad categories of policy setting models. In the first set of models, tax policies represent implementations of "optimal" policies as defined by the optimal tax literatures. In this case, tax changes should be fully determined by changes in economic fundamentals, such as elasticities, population shares, and other relevant parameters. To the extent that these fundamentals (or their proxies – e.g. demographic and economic indicators) are observable to policymakers, they should also be observable to researchers, making tax policy highly predictable.

The second set of models treats policy makers as potentially self-interested utility-maximizing agents who may or may not take voter preferences into account. The most well-known of these models, the median voter framework, is fully deterministic – thus as long as the median voter's preferences are observed to policymakers, they should be observable to researchers. This intuition, however, can be extended to settings with idiosyncratic shocks, where policymakers (or voters) "tremble" when making choices. In these frameworks, tax policy should still be highly predictable, as long as the appropriate measures of aggregate policymakers' preferences and relevant decision-making factors can be observed. To build intuition, consider the outcome of a 70-30 weighted coin flip. If we were to predict the outcome of an individual flip, we would fail approximately 30% of the time. However, if our goal is to predict whether 100 coin flips will result in a majority heads outcome, we are likely to succeed with nearly a 100% probability. Note, however, that the majority heads outcome becomes harder to predict as the coin gets closer to the 50-50 unweighted case.

Turning back to policy setting, suppose a policymaker's decision to vote yes on a given policy at time t is driven by a time-varying individual preference α_{it}, a vector of observable factors X_{it}, and a random shock ε_{it}. Furthermore, assume policymaker i votes yes if the policy results in positive utility, and no otherwise:

$$Vote_{it} = \begin{cases} 1 & \text{if } U(\alpha_{it} + \beta X_{it} + \varepsilon_{it}) \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

If we were to predict individual policymakers votes using observable factors X_{it}, then the explanatory power would depend on the variance of the idiosyncratic factors ε_{it}. For example, some policy votes are easily predictable because they strictly follow party lines, while others appear to be driven by unobservable factors.

However, if policy adoption is determined by majority rule, as is common in U.S. state legislatures, then predicting policy outcomes is equivalent to predicting whether the share of yes votes, $\frac{1}{n}\sum_i Vote_{it}$, exceeds 0.5 (or another cutoff in case of supermajority rules). By the law of large numbers, for a sufficiently large number of voters n, the share of yes votes is approximated by the expected value:

$$\frac{1}{n}\sum_{i} Vote_{it} \rightarrow \mathbb{E}[Vote_{it}] = Prob[U(\alpha_{it} + \beta X_{it} + \varepsilon_{it}) \geq 0].$$

Therefore, in contrast to individual policymakers' votes that are idiosyncratic to some degree, policy decisions are effectively deterministic and are driven by the joint distribution of policymakers' preferences, observable factors and idiosyncratic shocks. One exception to this rule is circumstances where $\mathbb{E}[Vote_{it}]$ is close to 0.5. In these situations, the policymakers are evenly split thus making policy outcomes potentially as difficult to predict as individual votes.

Two practical considerations are worthy of a discussion. First, state legislatures are not very large, ranging from 20 to 67 members in the upper chamber and from 40 and 400 members in the lower chamber, with the averages of 40 and 110 members, respectively. Due to legislatures' size, the law of large numbers will not hold perfectly, resulting in some uncertainty. This uncertainty should be smaller for larger state legislatures and when the variance of $Vote_{it}$ is small. Second, the joint distribution of policymakers' preferences, observable factors and idiosyncratic shocks is not known. For this reason, an econometric model employed to predict policy outcomes ought to be sufficiently flexible, in order to allow for unknown relationships and functional forms. Since machine learning techniques such as LASSO and Random Forest allow for such flexibility, lack of predictive power in such models would imply that either relevant explanatory factors have been omitted, individual idiosyncratic shocks dominate policymakers' preferences and decision-making factors, or policymakers are evenly split in their preferences.

2.2 Data

2.2.1 Tax Rate Data

We collect data on top personal income, top corporate income, sales, cigarette per pack, gasoline per gallon, and alcohol spirit per gallon taxes from the Council of State Governments Book of the States from 1949 until 2020. Whenever possible, we cross-validate tax data with other sources, such as Tax Foundation, Tax Policy Center, OTPR's World Tax Database, CDC, and the Federation of Tax Administrators. We complement this information with corresponding federal tax rates.

In addition, we collect information on tax base features: minimum personal and corporate tax rates, income thresholds for minimum and top tax rates, personal income tax exemptions, whether federal tax liabilities are deductible, state EITC rates, and the inclusion of food and prescriptions in the sales tax base. We also utilize policy measures of the corporate tax base from Suarez Serrato and Zidar (2018). This data covers the following details: investment tax credit rate, number of years for loss carryback and carryforward, whether the federal income tax base is the state tax base, whether state has franchise tax, whether the state follows federal accelerated depreciation, whether the state follows accelerated cost recovery system depreciation, whether the state follows federal bonus depreciation, and state tax apportionment weights (payroll, sales, and property). The Suarez Serrato and Zidar (2018) data ends in 2010-2015, so we extend the series for apportionment weights and loss carrybacks/forwards to 2020.

Since we are interested in understanding the timing of tax changes, we record the new tax rate in the year it becomes effective even if the change occurs at the end of the calendar year. When studying tax changes, we disregard tax changes that are smaller than 0.1 percentage points for personal, corporate income tax and sales taxes. For excise taxes, we disregard tax changes that are smaller than $0.005. The latter restriction allows us to disregard the frequent but small changes of gasoline taxes that arise from automatic adjustment rules implemented in some states. We consider all tax changes as independent observations, even when these changes were legislated as a set of reforms. We do so because legislative decisions are frequently overturned: temporary tax changes often do not expire as scheduled and instead turn into permanent changes, while scheduled tax changes are often cancelled and/or changed in magnitude. Finally, we inflation-adjust nominal rates of cigarette, gasoline, and alcohol excise taxes using the BLS CPI series.

2.2.2 Political, Institutional, and Demographic Data

We follow the previous literature, summarized in Appendix Table A.5, to identify economic, political, institutional and demographic features that are likely to have important effects on tax policy. Our choice of factors has been motivated both by political economy and fiscal federalism studies that directly explore tax setting processes, as well as by economic models in general. While we are not able to include all plausibly relevant factors, we consider a wide range of explanatory variables. In this section we briefly summarize the nature of our data, details are available in Appendix Section A.2.1. The complete list is available in Table 2.1.

We consider 11 groups of explanatory variables. First, we account for a linear time trend to account for trends that may affect all states equally. Second, we consider variables related to federal tax policy: federal top income, top corporate, cigarette, and gasoline tax rates, both in levels and as changes. These variables are state-invariant and thus account for policy changes that occur across the states simultaneously. Third, we account for economic influences: federal and state-level recessions, federal mandates, unemployment rates, inflation and the prices of natural resources, and state's outstanding debt. We account for contemporaneous, lagged, and lead values.

The fourth, fifth, and sixth groups consider state institutional features which cover both time-invariant rules such as size of legislatures, balanced budget provisions, as well as time-variant rules such as the existence of rainy day funds, term limits, whether states require supermajorities and type for tax increases, and whether the state is a right-to-work state. Our seventh group accounts for political influences: party of legislatures' majorities and governorship and their strength (including lead values), number of party switches, whether this is the first year of new party in charge, state and federal government shutdowns, outcomes of presidential elections, and DW scores of state representatives/senators.

Eighth, we include variables that measure neighboring states' (top income, top corporate, sales, cigarette, gasoline and alcohol) tax policies – average tax rates of the neighbor and indicators of tax changes. Ninth, we control for other tax rates in the state, including lagged values. Tenth, we include state demographics: population measures (total, labor force, employment to population, density), poverty rates, demographic composition of the state (share of black and non-white/non-black residents, age composition), and median household income, again both in levels and changes. Our last group of explanatory variables includes values of top income, top corporate, sales, cigarette, alcohol and gasoline tax rates in 1995 as well as revenue shares of these six types of taxes in 1995. Including these variables allows us to control for the structure of tax system in the state, e.g. the importance of each tax type or lack of such.[6]

Finally, for completeness we also measure how the adjusted R^2 increases when state and year fixed effects are included, to account for remaining time-invariant state characteristics and state-invariant time effects.

[6]Year 1995 was chosen as the first round-year after which no major tax types were adopted.

2.3 How Have Tax Rates Changed Over Time?

2.3.1 Long-Term Trends

Figures 2.1 (a) and (b) show the unweighted average tax rates across 50 states and, when applicable, corresponding federal rates. Averages weighted by population are available in Appendix A.13 and are very similar. Two observations stand out. First, the six tax rates considered do not show similar patterns: while the sales tax rate steadily increased over the 70 year period, corporate and personal income tax rates both increased and decreased, while gasoline and alcohol taxes generally decreased. Cigarette taxes showed the most dramatic growth, tripling between 2000 and 2020. Second, with the exception of cigarette taxes, the most dramatic changes to tax rates happened during 1950-1990. Since approximately 1990, however, average tax rates have remained substantially more stable. The large increases in average rates were both due to adoptions of tax rates by various states and due to actual rate increases – Figures 2.1 (e) and (f) show similar patterns despite including only states with nonzero tax rates. These figures also show the tax levels of the newly adopted taxes and years when they were introduced. Most adoptions happened before 1970, and in most cases – though not always – taxes are first adopted at rates lower than the prevailing average at the time.

The average tax rates mask substantial heterogeneity in rates across states. Figures 2.1 (c) and (d) plot the coefficient of variation (CV) – the ratio of the standard deviation to the mean for all 50 states.[7] Figures 2.1 (c) and (d) show two distinct patterns. For income and sales taxes, we see a dramatic decrease in variation during 1950-1990 and little convergence in rates since then. In contrast, for excise taxes, the coefficient of variation remains relatively stable. Among the six tax rates, alcohol spirit taxes exhibit the largest heterogeneity, followed by personal income and cigarette taxes, then corporate

[7] Results are robust to using other measures of convergence, e.g., the standard deviation.

income and sales taxes. Gasoline taxes are the most homogenous. The large decrease in heterogeneity of income and sales taxes could either be due to adoptions of these taxes by the states or due to changes of existing rates. Figures 2.1 (g) and (h) plot the coefficient of variation (CV) for states with nonzero rates only, thus shutting down the extensive margin effect due to adoptions. Figures 2.1 (g) and (h) show that the 1950-1990 convergence was primarily due to a large number of new tax adoptions rather than convergence of rates. In fact, personal income taxes exhibited increasing heterogeneity through the 1970s. For excise taxes, adoptions played a smaller role.

Our results are consistent and complementary to findings of Rhode and Strumpf (2003) who document a substantial convergence in state policies over the 20th century, but find similar levels of heterogeneity during the 1970-90s. The lack of substantial convergence or divergence in presence of reduced mobility costs, as argued by Rhode and Strumpf (2003), is inconsistent both with Tiebout-sorting and race-to-the-bottom competition model predictions, suggesting that these are not the primary drivers of tax policy changes.

Additional results are available in the appendix: Figure A.14 shows how the long-term trends vary by region. Overall, all regions follow a similar pattern but the changes are more pronounced among Northeast and Midwest states, and lowest among South states. Figure A.15 shows long-term trends for other tax rules, such as minimum tax rates, tax brackets, and corporate tax rules. Similarly to results in Figures 2.1, we do not see much of a convergence or divergence among most tax rules.

2.3.2 Timing of Tax Changes

Figure 2.2 shows the percent of states that increase (resp. decrease) a given tax rate in each year.[8] The grey vertical lines highlight changes in corresponding federal tax rates.

[8]The percent of states that change the tax rate is conditional on already having the tax.

Alcohol taxes are adjusted the least frequently, by 5% of states on average each year. Gasoline taxes are changed the most frequently, by 17% of states in an average year. Across all tax rates, each year saw an average of 20 states changing at least one tax rate, ranging from 4 states in 1952 to 35 states in 1983.

With the exception of top personal and corporate tax rates, most tax rates have been increasing over time. Income taxes (both personal and corporate), on the other hand, saw a large number of tax increases prior to 1980, but since then have mostly decreased. Importantly, we see that while tax changes are numerous, they do not appear to follow a well-defined pattern. For example, we do not see a consistent clustering of tax increases or decreases around federal tax changes, nor do we see clustering of tax changes in general, as would happen in the case of fierce state competition. Finally, tax increases and decreases often occur in the same year.

Next, we explore whether different tax types are changed in the same year, and if yes, whether states tend to increase or decrease all tax rates across the board, or instead, shift tax structures by increasing some rates while decreasing others. In Figure 2.3, among the increases (or decreases) in each tax on the x-axis, the vertical bars specify the share that coincides with an increase (or decrease) in another tax type in the same state and year. For example, Figure 2.3 (a) shows that among all of the times that states increased the personal income tax, 10% occured alongside a decrease in the alcohol spirit tax in the same state and year. The results are striking: a large number of tax changes occur simultaneously! Overall, 36% of state tax changes involve changes of two or more tax rates, and 13% involve three or more rates.

This pattern is particularly true for tax increases, and for personal, corporate, and sales tax rates. We see that 46% of top income tax rate increases coincided with a corporate rate increase, and 23% coincided with a sales tax rate increase. Meanwhile, personal income tax decreases coincided with corporate tax decreases in 26% of cases. Corporate tax increases and decreases also show a high overlap with both personal and sales taxes. However, Figure 2.3(d) provides strong evidence against tax substitutions: when states increase their tax rates, they rarely cut other tax types to compensate. Instead, we find many instances of multi-tax increases or decreases. A possible explanation for the observed patterns is that certain combinations of tax changes are more politically feasible than others (Bierbrauer et al., 2021).

Figure 2.3 highlights the importance of paying attention to other tax changes when using cross-state tax variation in empirical studies. This is particularly important for researchers that employ variation in personal, corporate and sales taxes, as well as for studies of tax increases in general, as these are most likely to occur as a bundle. Empirical researchers must be mindful of such co-occurencies when attributing their estimated effects to a particular tax change.

Finally, Appendix Figure A.16 shows similar evidence but focusing on the minimum and top income tax rates among states with progressive tax schedules. Once again we see a a large degree of co-occurences among increases and decreases, however, the rates differ: top income tax rates increase in 61% of the cases when the minimum rate increases, but the minimum rate is raised in 35% cases of top rate increases, with similar pattern for corporate rates. Put simply: top rates are changed more frequently than minimum rates.

2.3.3 Heterogeneity in the Frequency of Changes

Figure 2.4 explores the extent to which states differ in how often they change tax rates and how. Figure 2.4(a) orders states by the total number of personal, corporate, sales, cigarette, gasoline, and spirit tax changes. The number of changes varies dramatically across states: over the 70 year period studied, the four least active states – AK, AL, VA, and WY – changed the six tax rates fewer than 20 times. On the other hand, the most active states – CT, NE, and NY– changed their taxes more than 80 times, i.e., more than once per year on average. Overall, states that do not have certain taxes – in particular personal income taxes (AK, FL, NH, NV, SD, TN, TX, WA, WY), sales taxes (AK, DE, MT, NH, OR) or corporate taxes (NV, OH, SD, TX, WA, WY) – appear to be less likely to change tax rates than states that have all six types of taxes.

Figure 2.4(b)-(g) explore whether states that change their tax rates frequently tend to make smaller changes when compared to states that change their taxes infrequently. This may happen if some states prefer to adjust their rates gradually instead of making large occasional adjustments. For all tax rates we see a weakly negative relationship between the size of tax changes and frequency of changes, with this relationship being most pronounced for sales, cigarette and gasoline taxes.

2.3.4 Tax Rate Level Persistence And Political Leanings

Figure 2.5 shows how tax rates have varied over time within each state. For each tax rate, we show the tax rate in 1950 (or the year that tax was adopted), the tax rate in 2020, as well as the average, mininimum, and maximum over the time period. Furthermore, we color each state in blue, red, or grey depending on their political leanings in most recent years. Specifically, we break down states into three groups based on states' pledges in recent presidential elections. We consider a state a "safe" Republican ("safe" Democratic) state if the state voted for a Republican (Democratic) presidential candidate in every election since 2000 (see Table A.6). All other states are considered swing states. Figure 2.5 thus shows how much tax rates deviated from the mean during the studied period, and whether state tax changes generally moved in the same direction or saw a large number of fluctuations around the mean.

We see that for personal and corporate income taxes, most states exhibit a fluctuating pattern: for many states, 1950 tax rates are at or near the minimum, yet, 2020 rates are often below the maximum, and in many cases below the mean. However, consistently with Figure 2.4, states vary dramatically in their tax ranges. For some states, e.g. PA, IN, AL, VA, KY, LA, we see minimal changes of the top income tax rate. For other states, we see significant swings: DE's top income tax rate ranged from 3pp to 19.8pp, despite the fact that the rates in 1950 and 2020 were very similar (6.25pp and 6.6pp respectively).

In contrast to income taxes, sales and excise taxes show a one-directional pattern. We see that almost all states increased their sales tax and cigarette tax over time and that in many states, the 2020 sales tax rate is at the highest level sales tax has ever been. Low cigarette taxes generally reflect lack of inflation adjustments rather than active tax changes. The opposite pattern is seen for gasoline and alcohol spirit taxes: current tax rates are at their lowest point in the past 70 year period for most states.

Figure 2.5 also shows that there is limited persistence in tax rates over time. Some states increased their rates dramatically, others less so, and the magnitude of change is not well correlated with the starting or ending rates. While the correlation between 1950 and 2020 rates for the personal income tax is 32%, it is substantially smaller for the other tax rates: -6% for corporate income, -4% for sales, 16% for cigarette, -11% for gasoline, and 15% for alcohol.[9]

Finally, we see that there are more Democratic-leaning states at the higher end of the tax rate distribution and more Republican-leaning states at the lower end. Yet, the differences are rather small, and there is substantial overlap. Therefore, political leanings affect tax policies but do not provide an exhaustive explanation of tax rate levels.

[9] Correlations between each state's *rank* in 1950 and 2020 are very similar. Correlations across tax rates types are available in Table A.7.

2.4 Do Tax Rates Respond to Economic, Political and Institutional Influences?

In this section we explore to what extent the substantial heterogeneity in tax rates and the numerous tax changes over time can be explained by economic and political causes or is driven by institutional rules discussed in the previous literature. We consider three types of influences on tax changes: economic needs, such as interstate tax competition, economic downturns and federal mandates; political incentives, such as election cycles, and changes of governing parties; and institutional rules, such as balanced budget provisions, terms limits, legislature size, session duration, and voter initiative rules. In this section, we omit alcohol spirit taxes from our analysis because tax changes are very infrequent.

We ask how much of the tax policy variation can be explained by these factors using three complementary approaches. First, we consider each potential cause individually and conduct a permutation analysis where we calculate the share of tax changes that occur after an event, relative to the share that would occur if tax changes are randomly timed. Second, we combine all factors together in a simple linear model, and investigate how much of the variation in tax changes can be explained as well as the explanatory power of individual factors. Third, we consider a more flexible set of models that allow for interactive terms, using LASSO and random forest techniques.

2.4.1 Permutation Analysis

To understand whether taxes respond to economic needs, we explore the extent to which tax changes occur simultaneously or following economic changes. Of course, such co-occurrences need not be causal in nature, and may occur by pure chance, especially, if tax changes are numerous as is the case for top personal income taxes. For this reason, we supplement the observed coincidence rates with simulated ones, which are calculated as follows: we keep the number of tax changes fixed but randomize their timing. We then calculate the number of random matches. We repeat this procedure 100 times and then show the average number of simulated coincidences, as well as the 5th and 95th percentiles.

The above exercise does not prove the existence of causal responses when the observed co-occurrences greatly exceed simulated rates. However, it provides evidence against such causal relationship in cases where the observed co-occurrence matches the simulated rate, which is what we find in many cases. We now describe how we measure co-occurrences in the data.

Tax Competition. Tax competition has long been seen as a likely force behind state tax changes. While tax competition could in principle be responsible for both tax increases and tax decreases, it is typically predicted to drive tax rates down. To investigate whether states change their tax rates in response to competition, we identify tax changes in the neighboring states. For excise taxes, we consider geographical neighbors, since competition is likely to be driven by cross-border shopping. For all other taxes, we define neighbors based on migration flows, following Baicker (2005). For each state, we identify five "neighbor" states that accept the largest number of migrants from that state, and use those states' tax changes in our analysis. Tax changes that were motivated by tax competition are likely to *follow* neighbors' tax changes. However, because the legislative process can be slow yet observable by other states, we focus on tax changes that occur simultaneously and/or follow neighbors' tax changes; or occur within a set number of years of neighbors' tax changes. We find that our results are qualitatively robust to the choice and type of time-window studied and the measure of neighborliness.

Our approach thus differs from the previous literature that generally focused on identifying a causal relationship between neighboring states' tax rate levels (e.g. Devereux, Lockwood and Redoano (2007)). Instead, we focus on the timing of tax changes, as we believe this presents a stronger test of competition-driven responses, since similarity in tax rates levels may represent similarity in preferences in nearby jurisdictions.

Recessions. Economic downturns may force states to increase or decrease taxes in order to collect more revenue or to stimulate their economy. To the extent that states are generally required to balance their budgets on a yearly basis, tax rate increases are more likely. The extent of responses, however, is likely to depend on the nature of the balanced budget rules of a given state. An average state recession episode lasts 2.2 years. Since revenue needs and stimulus incentives are time-sensitive, we expect economic-downturn-driven tax changes to occur during the recession years. As a further robustness check, we also allow tax changes to occur during or 1 year after the recession.

Federal Mandates. Unfunded federal mandates may impose significant revenue burdens, requiring states to raise more tax revenue – and thus increase their tax rates – in order to finance mandate-related expenditures. We consider federal mandates summarized in Table A.8. Most mandates became effective within two years of their enactment. For this reason, we focus on tax changes that occur in the year of enactment or in the year of becoming effective, as well as on tax changes that occur during the enacted-effective window for mandates that became effective within three years of enactment.

Figure 2.6 shows the percent of all tax changes that occur (a) following a neighbor's tax change, (b) during a state recession, and (c) upon implementation of a federal mandate. In each figure and for each tax type, the top bar shows the actual percent of tax changes that coincide with the studied event, while the bottom (gray) bar shows the simulated mean. Appendix Figure A.18 shows that our results are robust to the choice of window, while Figure A.19 shows that results are similar when focusing on largest 50% of tax changes.

Figure 2.6(a) shows some support to the notion that competition may affect tax policy – for a number of tax types, we see that taxes are more likely to be implemented following a change in neighbors' taxes. For sales as well as gasoline and cigarette taxes, we see that tax changes are more common after a neighbor's tax change than a pure coincidence would predict. However, the changes in personal and corporate income taxes appear to be largely coincidental. One possibility is that purchases of goods are perceived by state legislatures to be more responsive, due to temporary travel across borders, than the location of personal or corporate income.

Figure 2.6(b) explores what share of tax changes occur during recessions: between 10% and 22% of tax changes occur during the years of recessions. Nonetheless, most of these occurrences appear to be coincidental: the observed shares are very similar in magnitude to simulated shares. While Figure 2.6(b) tells us what share of tax changes could in principle be explained by recessions, it does not provide us a clear answer as to whether recessions necessitate tax changes, since the observed occurrences depend on the frequency of recessions. Figure A.17 explores this question further by showing the share of recession episodes that lead to a tax change, separately for episodes of state-specific recessions and federal recessions. Personal income tax rates change in 22% of state recessions, corporate taxes are changed in 24% of cases, while sales taxes are changed in 22% of recessions. Taxes are changed significantly less frequently during federal recessions: only in 6-9% of cases. Overall, Figure 2.6(b) and Figure A.17 provide suggestive evidence that most tax changes are unlikely to be driven by ongoing recessions.

Finally, Figure 2.6(c) explores what share of tax changes occur in response to federal mandates. Again, we see no difference between the observed co-occurrence rates and the simulated, suggesting that the federal mandates are unlikely to result in timely tax changes. To the extent that federal mandates are frequent (a new mandate was introduced or became effective in 40% of years), they are likely to influence tax policy but not in an urgent way.

Figures 2.6 explores the frequency of tax changes but not their direction. Figure 2.7 explores whether the tax changes that coincide with neighbors' tax changes, recessions and federal mandates are tax increases or decreases. As a point of comparison, Figure (a) shows the composition of tax changes in all years. Several key observations stand out: neighboring states' changes are generally followed with tax changes in the same direction, but not always. During recessions, states are more likely to increase personal and corporate taxes than decrease them. But overall, the relative share of decreases/increases approximately matches the averages in the top panel.

Party Control Changes. Next we explore to what extent tax changes appear to be driven by political incentives. Previous research has documented that governments can be more or less successful at passing reforms when having full versus partial control of the legislative chambers and governorship (Roubini and Sachs (1989), McCubbins (1991), Alt and Lowry (1994), Castanheira, Nicodème and Profeta (2012), Bernecker (2016)). We start by exploring whether tax changes primarily occur after majority party switches, and whether tax changes are more likely to happen when one party holds a majority in both chambers of the legislature and holds the governorship. The top row of Figure 2.8(a) shows the breakdown of party affiliations of the House majority, Senate majority and Governor during the 70 year period we study. In 53% of observations, a given party holds majority in all three offices, and roughly one fifth of these (11%) represent first term years after one of the majorities was switched. In 28% of observations, the House and Senate majorities coincide but differ from governor's party affiliation. Finally, 18% of observations represent years with divided House and Senate majorities.

The overall shares of the top row can be compared to shares of political structures when tax changes occur. Since the shares in all rows of Figure 2.8(a) are quite similar, this suggests that tax changes are not disproportionately likely to occur when party controls change. A small exception to this rule are changes of sales tax rates: these are less likely to occur during periods of divided governments but the differences are relatively small. This finding is perhaps not surprising in light of the fact that Republicans or Democrats hold the majority of both legislative chambers in 82% of years, providing them with ample opportunities for changes. The results are similar, when looking separately at safe Democratic and Republican states (Figure 2.8(c) and (d)) and even swing states (Figure (b)), or when focusing on the 50% largest tax changes (Appendix Figure A.20). Appendix Figure A.21 suggests, however, that there is some heterogeneity across Republican and Democratic states when considering tax increases and decreases separately.

Next, Figure 2.9 explores to what extent presidential elections affect states' tax policies. Specifically, we break states into four categories based on whether the state is "happy" or "upset" about the election outcome (i.e., whether the winning presidential candidate won in the state or lost), and whether the winning candidate matches the majority party of the state's legislatures (both lower and upper chambers). The top row summarizes the share of years a given outcome occurs, which then can be compared to shares when given tax changes occur.[10] Figure 2.9 shows two notable patterns: states that vote for a Republican candidate that loses are significantly less likely to pass a tax increase of any tax type. Interestingly, this happens irrespective of whether the Republicans hold a majority in the state's legislature or not. We see the opposite pattern for states that vote for Democratic candidates: they are more likely to pass tax increases when their preferred candidate loses. The observed pattern is thus consistent with polarization in tax policy and may represent a response to *anticipated* federal tax policies.

[10]For example, for state-year observations that vote for a Democratic nominee, 56% result in that candidate winning and 44% losing. In 62% of states voting for a Democratic candidate, states' legislative majority was Democratic.

2.4.2 Simple Linear Model and Variance Decomposition

Next, we combine all of the explanatory variables together in a simple linear regression model. This approach allows us to quantify the extent to which these factors can jointly explain the observed variation in tax policy, as well as the relative importance of each factor when controlling for the others. Because our explanatory variables are not orthogonal, most covariates contribute to the explanatory power in a non-unique way. For this reason, we use a Shapley decomposition method to assign each group of variable's contribution to the overall explanatory power, measured by the R^2.[11]

One may worry about endogeneity issues in our specifications since many explanatory variables are likely to be chosen simultaneously with our outcome variables. We allow for such endogeneity because we are interested in measuring the predictive power rather than causal estimates. Since the resulting predictive powers are low for all specifications we have considered, we err on the side of being too generous when choosing which variables to include.

We include a number of lagged and lead variables in our analysis – mainly those related to economic conditions or political changes. We include these terms to account for possibly dynamic relationships. At the same time, we choose to not include lags beyond 2 years because some events necessitate a speedy response – e.g., enacting tax reforms three years after recession is unlikely to be useful.

Our analysis does not account for tax changes at the local level, including changes in property taxes. Our empirical analysis, however, includes year or decade fixed effects.

[11] Another potential approach would be to use a Kitagawa-Oaxaca-Blinder decomposition method to understand the extent to which various factors can explain a gap in tax rates or changes between two groups. For example, Seegert (2016) documents a clear break in tax revenue volatility pre- and post-2000 and uses this approach to explain the gap between the two time periods. However, we do not observe a clear break in tax rates or changes across time or across states. As a result, we prefer a decomposition approach that seeks to explain the variation overall, without needing to separate similar states and years into two arbitrary groups.

The machine learning analysis discussed in the next section allows for differential predictions over time through interactions with decade dummies. Thus while we are not able to measure the influence of local policies (including property taxes) on state policies, we account for these via fixed effects and interaction terms. Decade interactions also allow our explanatory variables to influence state tax policies differently over time.

Variance decomposition results are summarized in Figures 2.10-2.11, which show the shares of total explained variation attributed to the above-mentioned groups of explanatory variables. Since the number of observations varies across tax rates, we show the adjusted R^2. Figure 2.10(a) summarizes decomposition of tax rate levels (in percentage points or in \$2020), while 2.10(b) and (c) show tax changes in p.p. or dollars (all changes or the largest 50% of tax changes respectively). Figure 2.11 focuses on the timing of tax changes, and thus performs decomposition of indicators of tax rate increases and decreases respectively, looking at all tax changes (figures (a) and (b)) or the largest 50% of changes (figures (c) and (d)).

We find that nearly all of the *tax rate level* variation can be explained with our chosen variables. However, most of explanatory power comes from lagged own tax rate and past (1995) tax rates. Put simply, past tax rates do well at predicting future tax rates (especially combined with a linear trend and federal tax rates), because tax rates are somewhat persistent (see Section 2.3.4). Since this decomposition does not distinguish between within-state variation and across-state variation, it exaggerates our ability to predict taxes. For this reason, we next turn to explain the magnitude and timing of tax changes.

Our ability to explain the *magnitude of tax changes* and the *timing of tax changes* is significantly weaker – the explanatory power decreases to under 20%. Unsurprisingly, when focusing on the magnitude of tax changes, past tax rates play a less important role. Instead, federal tax rates, political and demographic factors increase in relative importance. We see some variation in the relative importance of factors for different tax rates, but the overall ranking is generally consistent across tax types.

Our ability to explain the timing of tax changes is equally weak – under 30%. Interestingly, the tax increases and decreases appear to be influenced by different factors. For example, tax increases are more consistently influenced by federal tax policy than tax decreases. Similarly, economic factors (recessions and mandates), neighboring tax rates, and other tax rate levels are more important for tax increases. Political factors are important for both and yet account for less than one quarter of overall explanatory power. In general, tax rate increases are easier to predict than tax decreases. A likely explanation for this is that tax increases are likely to be more driven by economic needs while tax decreases are likely to be ideologically motivated. Since the former are more time-sensitive than the latter, timing of increases should be more predictable.

Figures 2.10-2.11 also show that focusing on the largest 50% of tax changes does not improve our predictive powers. A possible explanation is that the magnitude of tax changes is more idiosyncratic than whether the tax change is legislated in the first place. This explanation is consistent with the fact that explanatory powers are generally higher when focusing on the timing only, and when focusing on all changes rather than the largest ones.

Finally, one may worry that while predicting yearly changes may be difficult, predicting long-term changes may be easy. We test this possibility by conducting an equivalent analysis but using decade changes in Figure 2.12. One caveat to this comparison is that the adjusted R^2 does not perfectly account for changes in the relative number of observations and explanatory variables, making comparison of Figures 2.11 and 2.12 imperfect. With this caveat in mind, we see that the predictive power indeed increases. Importantly, the increase is not driven by one particular group of covariates – most groups explain a larger share of variation with the exception of own other tax rates. But the overall conclusion remains: tax changes are hard to predict, even in the long run.

There are two plausible explanations for why decade changes are more predictable. One, is that tax changes are costly to implement and therefore not every change in economic or political conditions results in action, e.g., similarly to investment decisions of firms. However, if this were the case, large tax changes should arguably be more predictable, since these are driven by stronger needs. This explanation is not consistent with evidence in Figures 2.10-2.11. Alternatively, our preferred explanation is that the increased explanatory power across all groups is consistent with the possibility of a long-term gridlock: economic, political, and demographic influences matter, but the timing of tax changes is volatile because of gridlock. Over time, however, changes in these factors do lead to changes in tax policy, making decade changes more predictable.

Results are similar when looking at other tax rules, though the predictive power is slightly higher (see Appendix Figure A.22). Most of this explanatory power, however, does not come from economic or political factors. Instead, federal tax rates and other tax rates matter. A likely explanation is the fact that tax base rules often change in conjunction with tax rates.

Overall, we conclude that a simple OLS model does a poor job of predicting the timing and magnitude of tax changes, especially year-to-year fluctuations.

2.4.3 Enriched Models Using LASSO and Random Forest

The above model allowed only for the simplest relationships between the explanatory variables and tax policy. It is possible and likely, however, that the relationship between economic, institutional and political factors and state tax policies is more nuanced than this simple linear model would permit. For this reason, we then turn to a richer set of models, LASSO and random forest, which allow for nonlinear and interactive terms.

Because it is neither possible nor desirable to include all possible nonlinear and interactive terms in a regression analysis, we employ LASSO techniques to select a subset of variables in a data-driven way. The LASSO approach selects a model that minimizes the prediction error while keeping the model not too complex by including a penalty parameter that increases in model complexity. The practical implementation of the LASSO method varies in penalty functional form approaches to determining the optimal model. In our setting, we found that LASSO and elastic net approaches work equally well, and the best results are achieved when the model is selected by cross-validation or using an adaptive approach; linear, probit and logit models yield similar qualitative results.

Random forest is a machine learning technique that allows for more flexible modeling. To make predictions, the algorithm builds multiple decision trees using a different random subset of the variables provided and a different bootstrapped sample of the data. The final predictions are then obtained by averaging individual predictions from the randomly built trees. The randomness of the sample variables and the dataset used to build a given tree ensure that individual trees are not correlated. This gives random forest its high predictive power and partially shields it from overfitting.

Table 2.2 summarizes our results. As our baseline comparison we take the models from Section 2.4.2 which included 166 "core" variables as well as state and year fixed effects. Next, we use LASSO to select the best model using these variables plus decade fixed effects and the full set of interactions – a total of 33,211 variables. Finally, the random forest algorithm uses all 166 core variables plus decade fixed effects and quadratic and cubic terms – a total of 379 variables. (Note that the random forest algorithm implicitly allows for additional fixed effects and variable interactions via its "tree" structure). The results summarized in Table 2.2 are based on 100 random splits of the data into a training sample (80%) and a test sample. Both algorithms search for the best model using the training sample, and that model is then used to make predictions on the test sample. We must note that while both LASSO and random forest algorithms require a number of choices made by a researcher, Table 2.2 presents the results from the "most promising" specifications. While the quantitative results vary depending on specification, the qualitative results do not.

Table 2.2 shows that while machine learning algorithms improve the predictions, the improvement is modest. The random forest algorithm does an outstanding job making predictions in the training sample but out-of-sample predictions are still poor and typically fall well below 20%. Moreover, many predictions are negative, suggesting that the models selected by machine learning perform worse than a simple sample mean. Note that these models are not selected to maximize the predictive power in the 20% out-of-sample. Instead, model selection is performed using the K-fold approach, whereby the training data is divided into K folds and the model is trained/tested K times, each time using a different fold as a test sample and the rest as training data. Thus the final predictions shown in Table 2.2 present a true and independent test of the model's performance, by evaluating prediction on the 20% sample of data that has never been used to select a model.

Overall, we find that tax policy is not well explained by the economic, institutional and political factors that we accounted for in this study. It is unlikely that the low predictive power is due to misspecification, as we consider both interaction terms and nonlinear specifications. Instead, it suggests that other factors – not considered by us – may drive tax policy or that tax policy is truly idiosyncratic.

2.5 Discussion and Conclusion

In this paper we explore determinants of state tax policy in the past 70 years. We document that while tax policy shows a fair amount of persistence over time, it also shows a tremendous amount of variation, both across states and within states over time. We consider numerous explanations for observed variation – economic, political and institutional influences – but conclude that most tax changes are difficult to predict. Overall, our best attempts explain less than 20% of observed tax variation, suggesting that more work needs to be done to understand the drivers behind state tax policy.

What are the possible explanations for the low predictive power? Our analysis may have omitted potentially important drivers of tax policy, for example, lobbying and political contributions. Whatever these omitted factors are, they appear to play a more important role than the economic, political, and institutional influences the literature has largely focused on. Alternatively, policymakers may be evenly split in their preferences, making policy decisions highly unpredictable, as our conceptual framework demonstrated. Since most states have strong Democratic or Republican majorities in the state legislatures, this would imply that policymakers are not as split on taxes as their speeches would suggest (Gentzkow, Shapiro and Taddy, 2019). Finally, it is also possible that the legislative process for tax policy may be so complex that idiosyncratic factors create substantial randomness in the timing and nature of policy response. If the variance of idiosyncratic factors is very large relative to the variance of other decision-making factors, policy decisions would be hard to predict. More work is needed to explore the nature of omitted explanatory factors and the source of idiosyncratic shocks. Since tax policy has direct consequences on state tax revenue and business cycle volatility, and can lead to policy uncertainty, excess tax volatility can have detrimental effects on growth and the welfare of state residents.

2.6 Tables & Figures

Table 2.1: 166 Core Explanatory Variables

Group	Type (N of var)	Variables included
1	Linear trend (1)	year
2	Federal rates (10)	rates and changes from previous year of top federal income tax rate, top federal corporate rate, and federal cigarette, gasoline, and spirit taxes
3	Recessions and mandates (22)	indicators: federal recession and one year lag and lead, state recession and one/two year lags and leads, 3 indicators for federal mandates: welfare-program-related, minimum wage change, and other, inflation and one year lag and lead, prices of crude oil, natural gas and coal, unemployment rate and one/two year lags and leads, long-term debt, change in long-term debt
4	State legislatures (11)	number of seats in the lower chamber, number of seats in the upper chamber, average legislative session duration in calendar days, average salary (in 2019/20), average per diem expenses (in 2019/20), indicator of right-to-work state, 5 indicators of whether the state adopted personal income, corporate, sales, cigarette and alcohol tax by 1949
5	Balanced budget rules (3)	indicators: whether budget deficits are allowed, whether capital expenditures are part of the budget, whether rainy day fund exists
6	Term limits (9)	indicators: whether there is governorship term limit, whether there is legislature term limit, whether this is a year in governor's last term, whether such a governor is Republican or a Democrat, whether voter initiatives are allowed, and 3 indicators for the supermajority requirements (60%, 67%, 75%)
7	Political factors (38)	number of times governor party switched, number of times majority in house, in senate or both switched, share of Republicans/Democrats in the senate/house; indicators: majority-Republican legislature and lead, majority-Democratic legislature and lead, governor Republican and lead, governor Democratic and lead, Southern Democratic governor, Southern Democratic legislature majority, divided government (party of house, senate and governor is not the same), first term after governor party change, first term after senate party change, first term after house party change, federal government shutdown that year, state government shutdown that year, Democratic president, state's preferred presidential candidate lost, legislature majority matches the party of the winning presidential candidate in the state, indicators for each year in the presidential election cycle, indicators for each year in the gubernatorial election cycle, interaction term of divided government and deficit not allowed, 4 DW nominate scores: for house/senate representatives and for each of two dimensions
8	Neighbors' taxes (22)	average tax rates in neighboring states this year and previous year; indicators of tax rate increases and tax rate decreases in neighboring states this or previous year; all separately for top income, top corporate, sales, cigarette, spirit and gasoline tax rates (decrease indicators omitted for cigarette and spirit taxes)
9	Own other taxes (11-18)	level and tax change regressions: level/change of other tax rates in the state top income, top corporate, sales, cigarette, spirit and gasoline tax rates; similarly in indicator regressions: indicators of tax rate increases and tax rate decreases in other rates; as well as lags of all 6 tax rates (decrease indicators omitted for cigarette and spirit taxes; own tax variables always omitted)
10	Demographics (20)	population, population density, labor force participation rate, employment to population ratio, poverty rate, percent of black residents, percent of non-white and non-black residents, percent of children (0-17 years old), percent senior residents (65+ years old), median household income; as well as changes in these variables
11	1995 tax rates and revenue shares (12)	tax rate (top income, top corporate, sales, cigarette, spirit and gasoline) levels in 1995; 1995 tax revenue shares of income, corporate, sales, cigarette, spirit and gasoline taxes

Notes: This table summarizes the 166 variables used in Sections 2.4.2 and 2.4.3. The simple linear analysis also includes state and year fixed effects (total of 285 variables). The LASSO analysis also includes decade fixed effects and the full set of interactions (total of 33,211 variables). The random forest algorithm also includes decade fixed effects and quadratic and cubic terms (total of 379 variables).

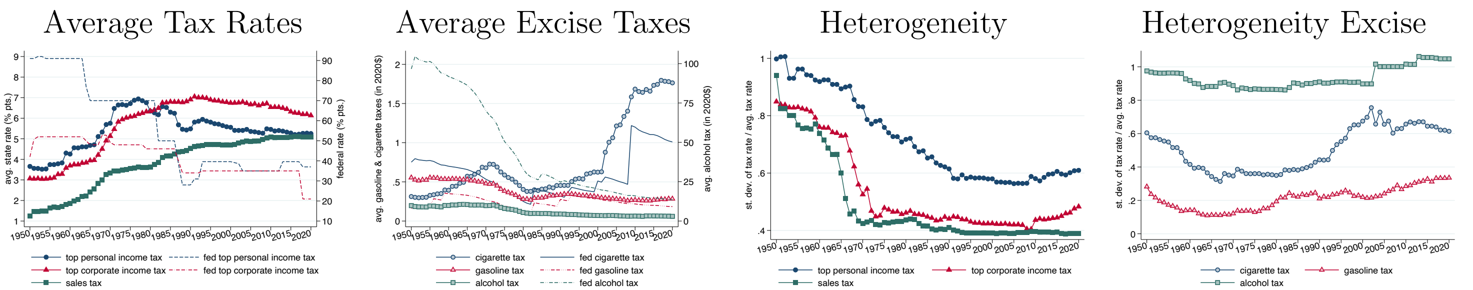
Table 2.2: Machine Learning Results

Outcome	OLS Training R^2	OLS Out-of-Sample R^2	LASSO Training R^2	LASSO Out-of-Sample R^2	Random Forest Training R^2	Random Forest Out-of-Sample R^2
Income tax change (pp)	0.19	0.02	0.19	-0.1	0.63	0.01
Corporate tax change (pp)	0.21	-1.25	0.11	-0.04	0.63	-0.01
Sales tax change (pp)	0.19	-9.55	0.2	-0.11	0.63	-0.01
Cigarette change ($)	0.14	-1.33	0.18	-0.02	0.64	0.01
Gasoline change ($)	0.18	-0.09	0.13	-0.23	0.65	-0.01
Alcohol spirit change ($)	0.13	-0.23	0.00	-0.01	0.56	-0.08
Income tax decrease	0.24	0.11	0.29	0.11	0.69	0.1
Corporate tax decrease	0.26	-0.71	0.33	0.09	0.7	0.17
Sales tax decrease	0.11	-5.62	0.01	-0.02	0.6	-0.03
Gasoline decrease	0.16	-0.32	0.1	-0.01	0.63	0.06
Income tax increase	0.31	0.19	0.37	0.18	0.67	0.08
Corporate tax increase	0.31	-0.27	0.35	0.13	0.66	0.05
Sales tax increase	0.22	-24.83	0.21	0.04	0.65	0.00
Cigarette increase	0.22	-1.63	0.27	0.03	0.68	0.04
Gasoline increase	0.20	0.07	0.26	0.01	0.69	0.08
Alcohol spirit increase	0.20	-0.03	0.22	-0.06	0.63	-0.02

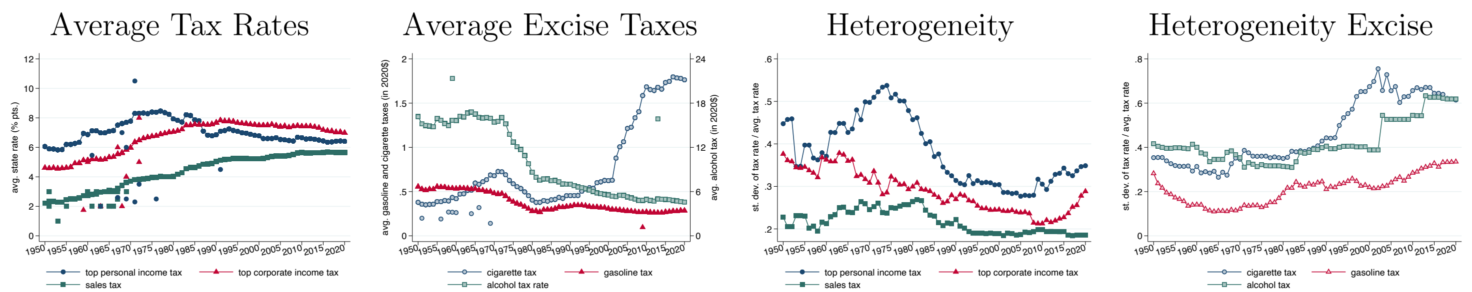
Notes: This table compares the results of linear regression models with LASSO selection models and random forest algorithms. The table reports the average R^2 obtained when estimating the model on the training sample (80% of the data) and when making predictions on the remaining 20% test sample. The average is calculated over 100 random splits of the data. The linear regression is estimated on the explanatory variables summarized in Table 2.1 plus state and year fixed effects (total of 285 variables). The LASSO model is a linear regression estimated on a subset of variables selected to minimize prediction error. The pool of variables includes those in Table 2.1 plus decade fixed effects and the full set of interactions (total of 33,211 variables). The random forest algorithm randomly selects subsets of variables to build decision trees, and then averages over the predictions from many trees. The pool of variables includes those in Table 2.1 plus decade fixed effects and quadratic and cubic terms (total of 379 variables), and the methodology implicitly allows for more fixed effects and interaction terms. Throughout, cigarette and alcohol tax decreases are omitted due to lack of events.

Figure 2.1: State Tax Rates Over Years

Panel A: All 50 States

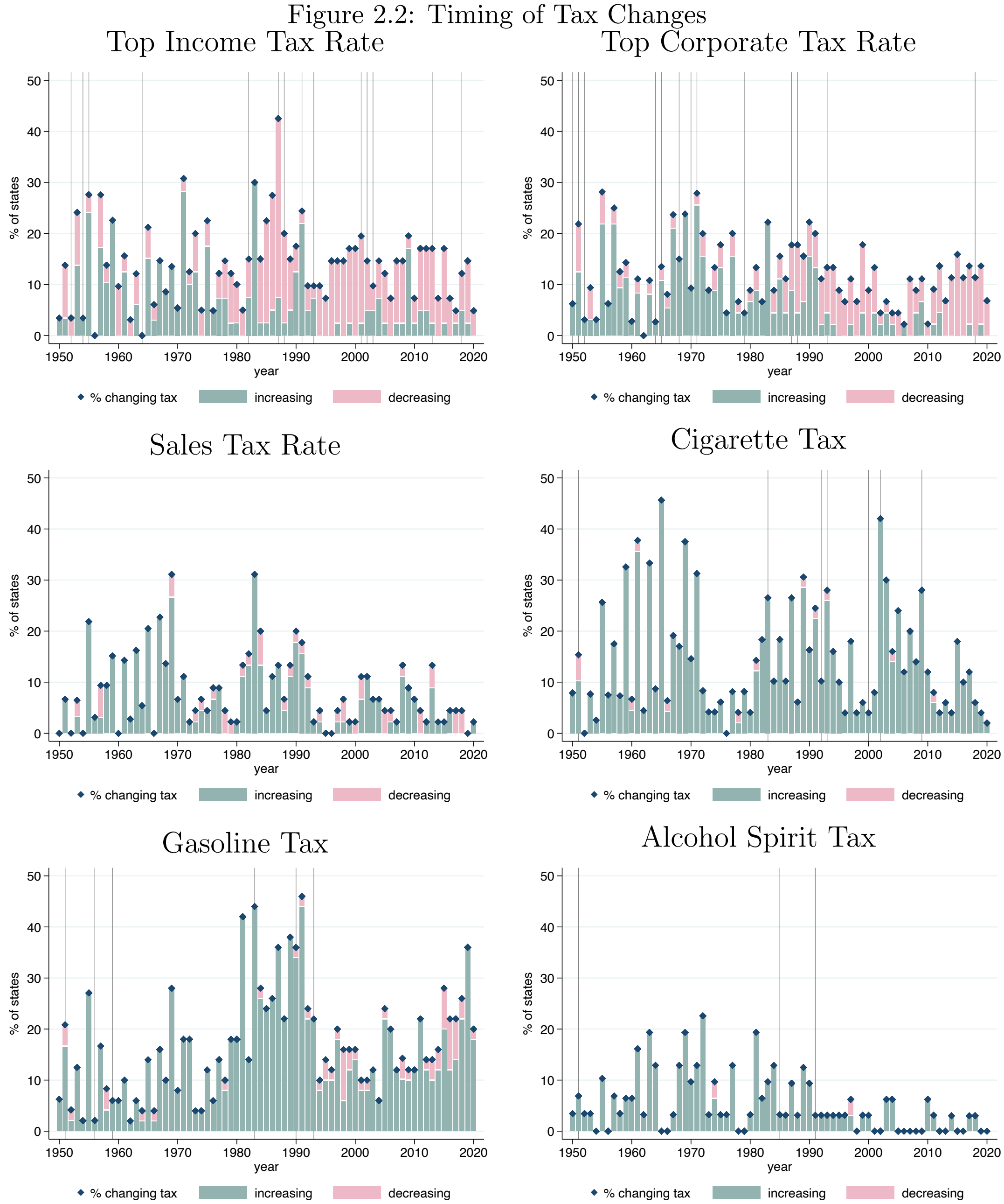


Panel B: States with Nonzero Rates



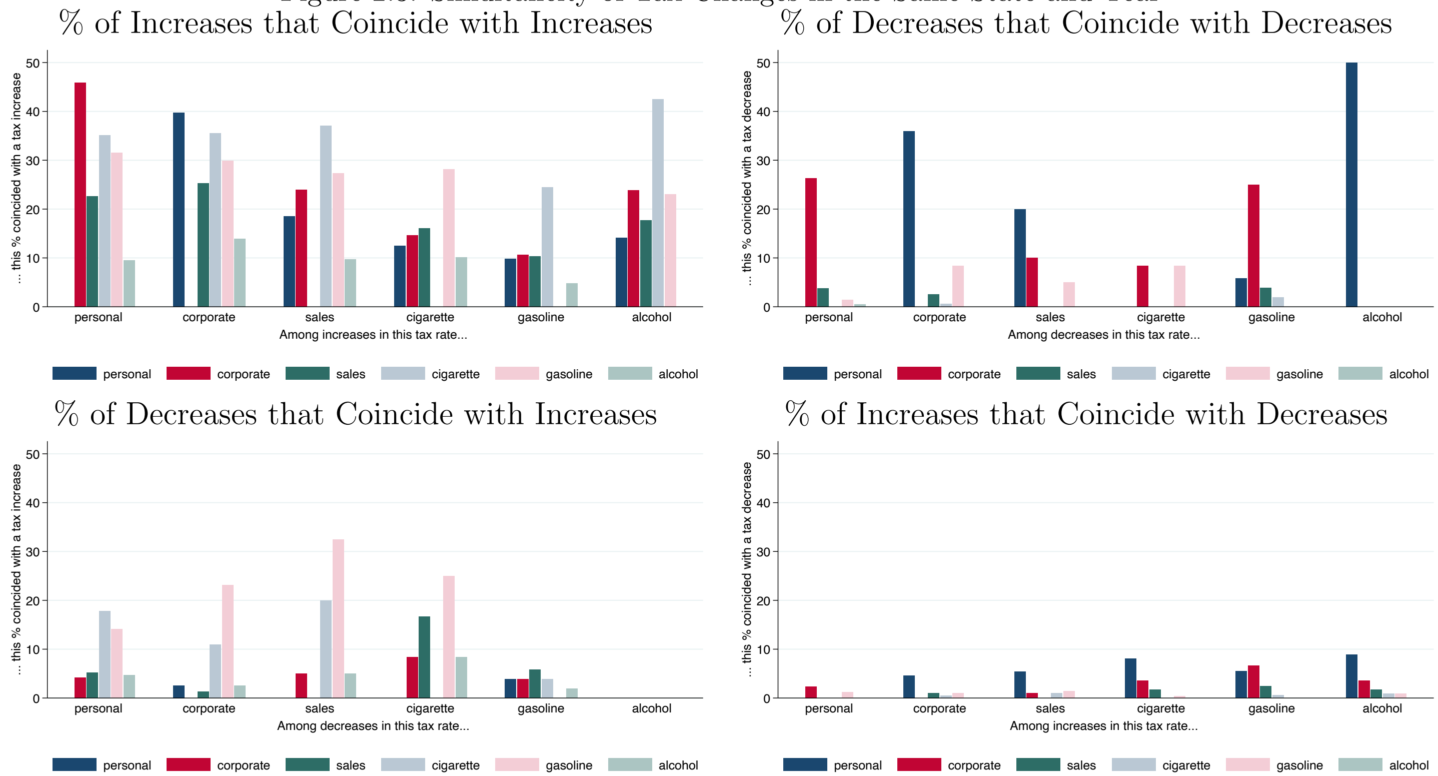
Notes: Figures (a) and (b) show average top personal income and corporate tax rates, sales tax rates, and average cigarette, alcohol (spirit) and gasoline tax rates, as well as corresponding federal tax rates. Figures (c) and (d) show the standard deviation of the state taxes divided by average tax rate (coefficient of variation). All states included, including those with zero rates. Figures (e)-(h) repeat the above but only for states with nonzero rates. Figures (e) and (f) in addition show new tax adoptions: tax rates levels and year of adoption. Population-weighted averages available in Appendix A.13.

Figure 2.2: Timing of Tax Changes



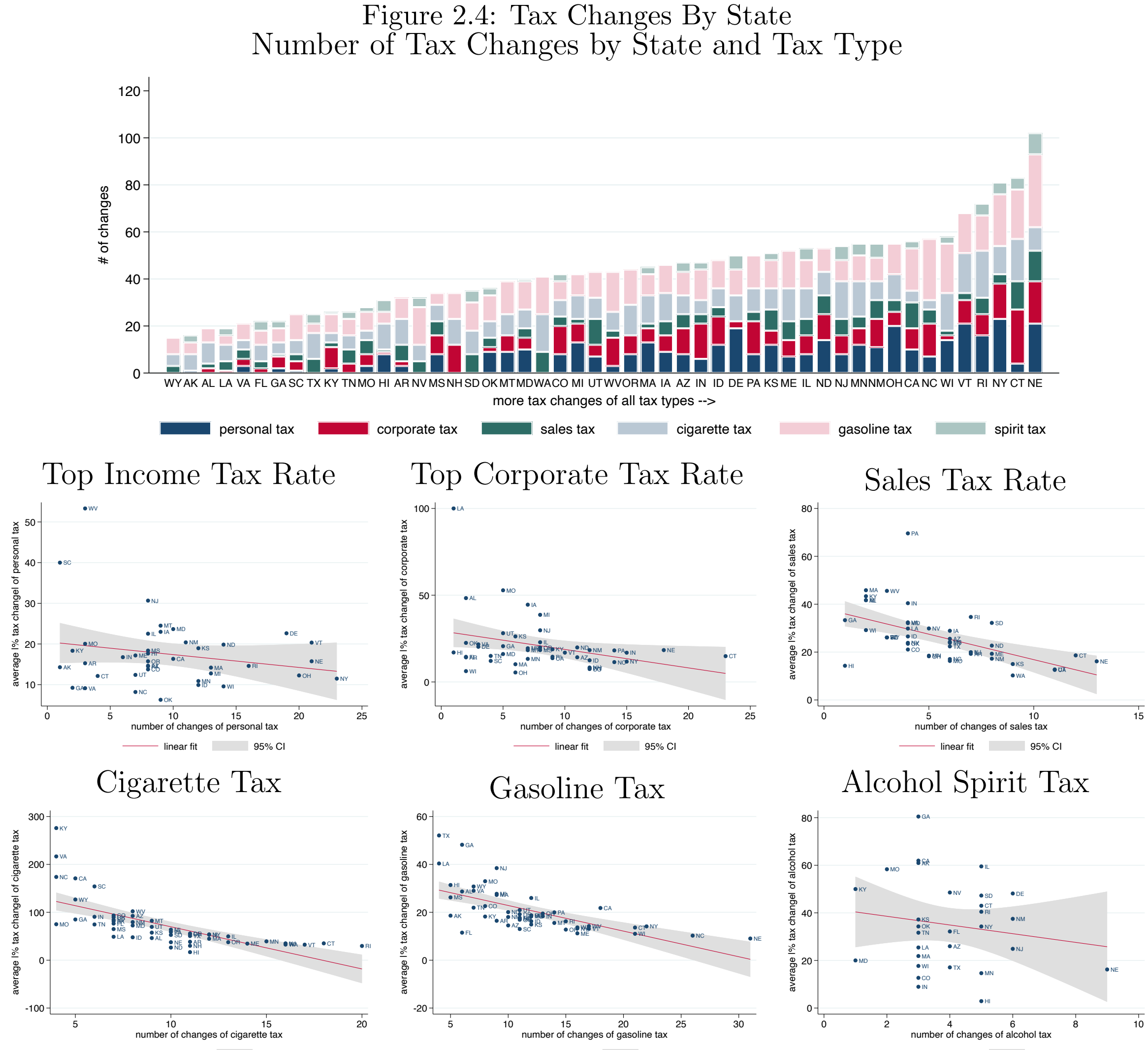
Notes: These figures show the percent of states that change a given tax rate in a given year (scatter points), increase it (green bars) or decrease it (pink bars). These statistics are shown for (a) top income tax rates, (b) top corporate tax rates, and (c) standard sales tax rates, (d) cigarette excise tax rates, (e) gasoline excise and (f) spirit excise tax rates. Gray lines identify changes in corresponding federal tax rates.

Figure 2.3: Simultaneity of Tax Changes in the Same State and Year



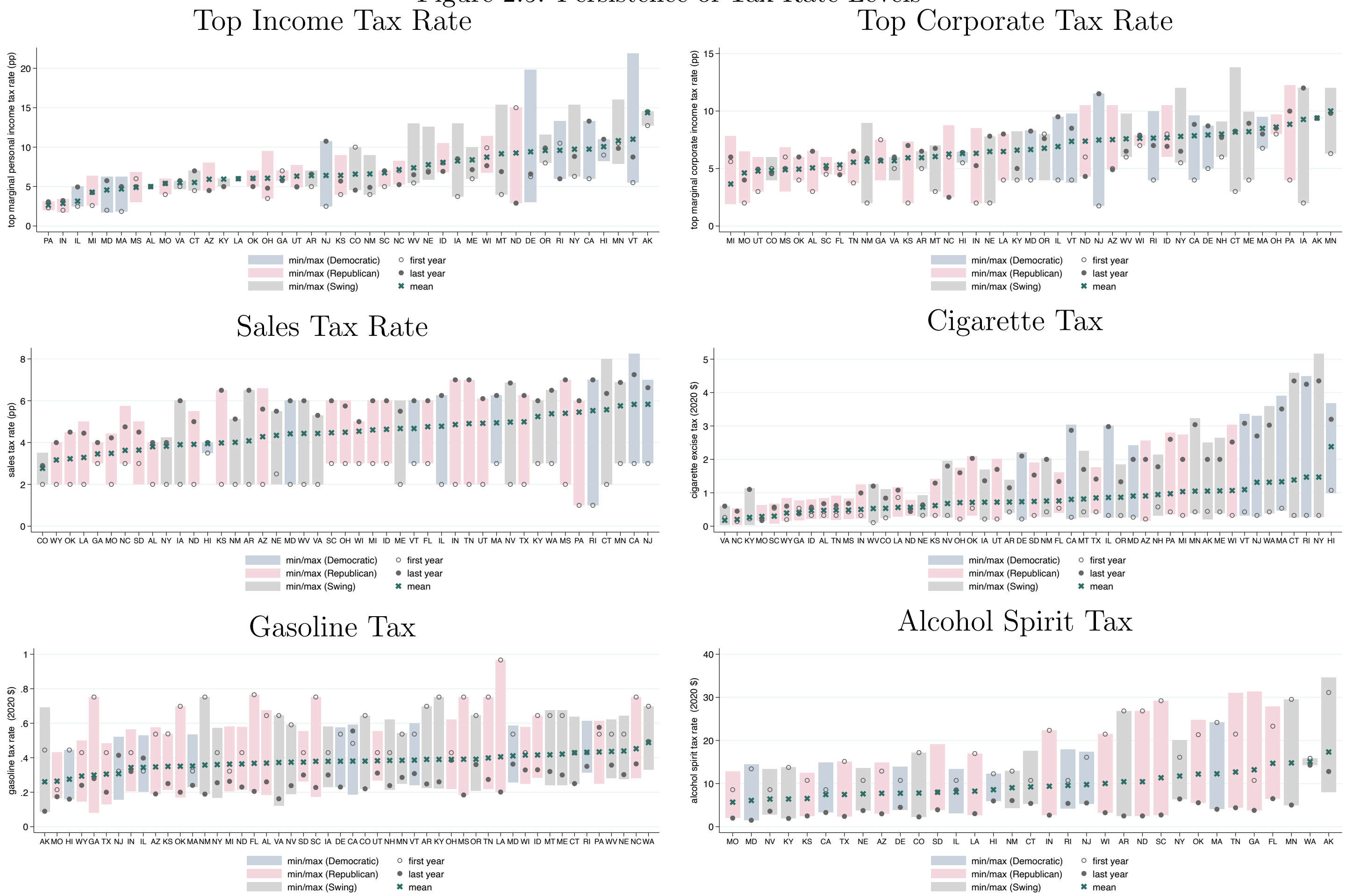
Notes: These figures explore the extent to which states change one tax rate while simultaneously changing another tax type (i.e., in the same year). Among the increases (or decreases) in each tax on the x-axis, the vertical bars specify the share that coincides with an increase (or decrease) in another tax type in the same state and year. These other tax types are identified by the color of the bar (top income tax rates, top corporate tax rates, standard sales tax rates, cigarette excise tax rates, gasoline excise, or spirit excise tax). For example, Figure (c) shows that among all of the *decreases* in top corporate income tax rates, 11% occured in the same year as an *increase* in the the cigarette tax rate in the same state.

Figure 2.4: Tax Changes By State



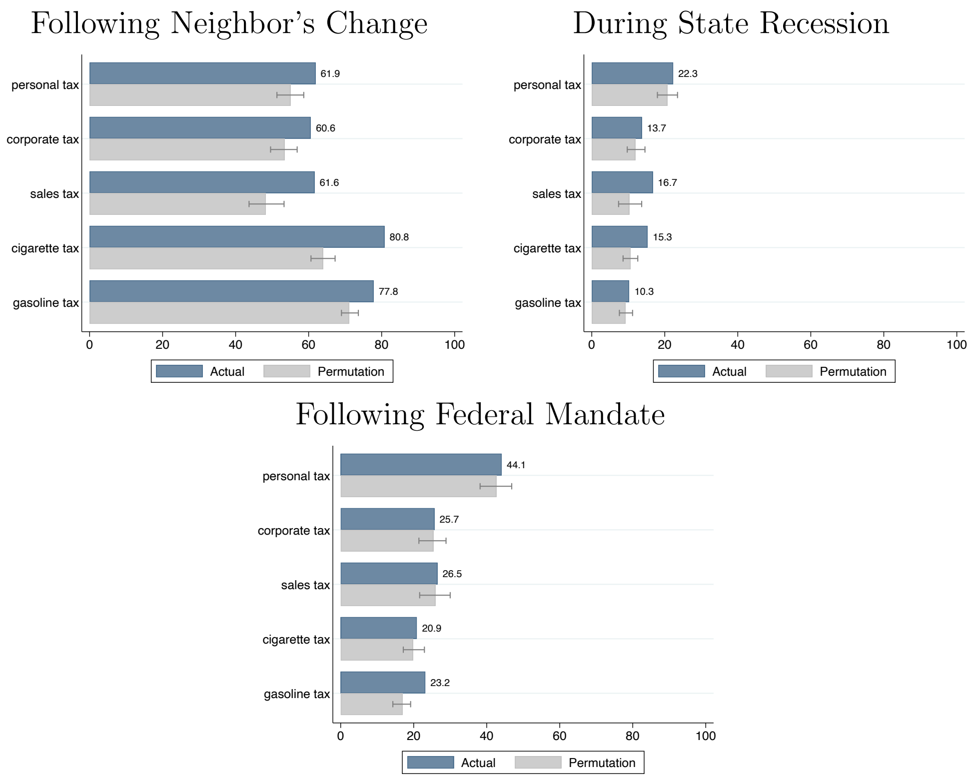
Notes: Figure (a) shows the number of tax changes in each state for six tax rates (top income tax rates, top corporate tax rates, standard sales tax rates, cigarette excise tax rates, gasoline excise tax, and spirit excise tax). Figures (b)-(g) show, for a given tax rate, the relationship between the number of tax changes and their magnitude (the average percent change in absolute value). Additionally displayed is the linear fit for this relationship, as well as the 95% confidence interval reflecting the uncertainty in both the slope and the intercept.

Figure 2.5: Persistence of Tax Rate Levels



Notes: This figure shows the average tax rate, tax rate in 1950 or in the year of tax adoption, tax rate in 2020, as well as the min and max rates in 1950-2020. States are ordered by average tax rate, and only non-zero values are included. These statistics are shown for (a) top income tax rates, (b) top corporate tax rates, and (c) standard sales tax rates, (d) cigarette excise tax rates, (e) gasoline excise and (f) spirit excise tax rates.

Figure 2.6: Percent of Tax Changes that Occur in Response to Economic Causes

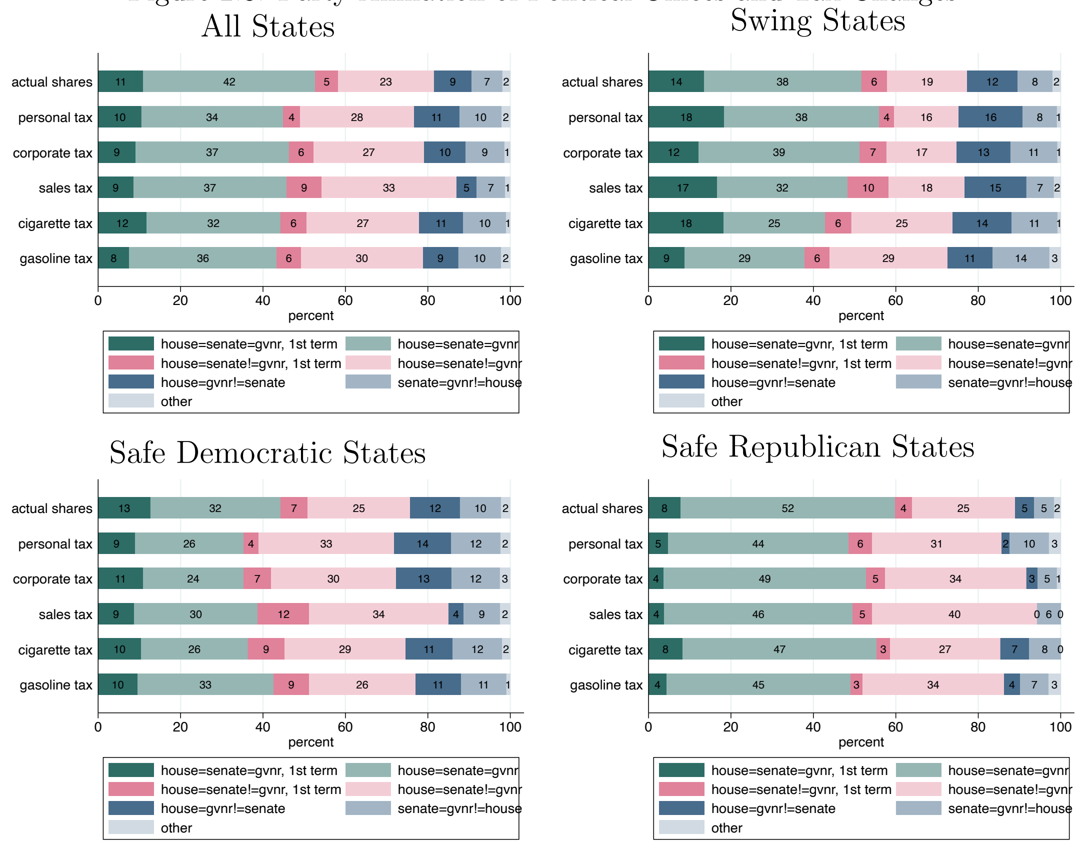


Notes: This figure shows the percent of tax changes that occur (a) in the same year or 1 year after neighboring state changes its tax rate; (b) during a state recession, or (c) in the years the federal mandate becomes enacted and/or effective. In all figures, the top blue bars show actual observed percentages, while the bottom grey bars show the simulated average, calculated by randomizing the timing of tax changes 100 times. The thin interval bars show the 5th and 95th percentiles of the simulated percentages.

Figure 2.7: How Do Taxes Change?

Notes: This figure shows the percent of tax changes that are increases or decreases and that occur (a)-(b) in all years, (c)-(d) in the same year or 1 year after neighboring state changes its tax rate; (e)-(f) during a state recession, or (g)-(h) in the years the federal mandate becomes enacted and/or effective.

Figure 2.8: Party Affiliation of Political Offices and Tax Changes



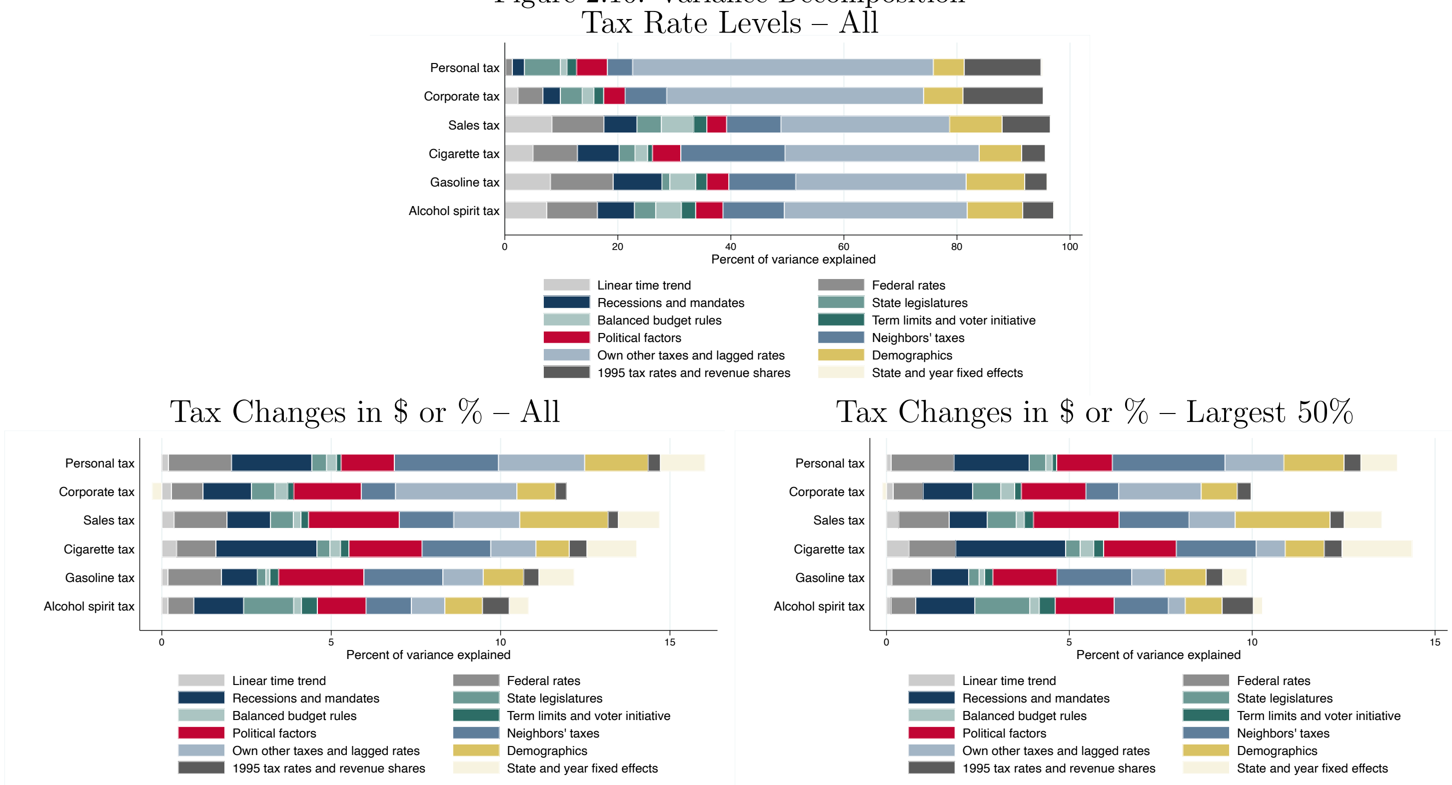
Notes: The top row of each figure shows the percent of yearly observations in which (i) the majority party of the House is the same as that of the Senate and of the Governor, and one of these three bodies switched party control; (ii) same as (i) but no party control change; (iii) House and Senate majorities are the same party, but Governor of a different party, and the joint majorities in House and Senate were obtained this term; (iv) same as (iii) but no party control change; (v) House majority matches Governor's affiliation but not Senate majority's; (vi) Senate majority matches Governor's affiliation but not House majority's; (vii) all other options (i.e. non-Democratic/Republican affiliations or lack of majorities). The next five rows show party affiliations in years when respective tax changes occur. Figures (b), (c) and provide these statistics separately for states that have only voted for a Democratic (Republican) presidential candidate since 2000 elections (see Table A.6), and swing states.

Figure 2.9: Presidential Election Outcomes and Tax Changes

Left: Vote Democratic   Right: Vote Republican

Notes: The top row of each figure shows the percent of yearly observations in which the state votes for a Democratic (left panel) or for a Republican (right panel) presidential candidate and that candidate wins ("Happy") or loses ("Upset"), while the state's House and Senate majorities match the preferred presidential candidate ("Match") or do not ("Not Match"). The other rows show similar break downs when tax increases or tax decreases of a given tax type occur.

Figure 2.10: Variance Decomposition

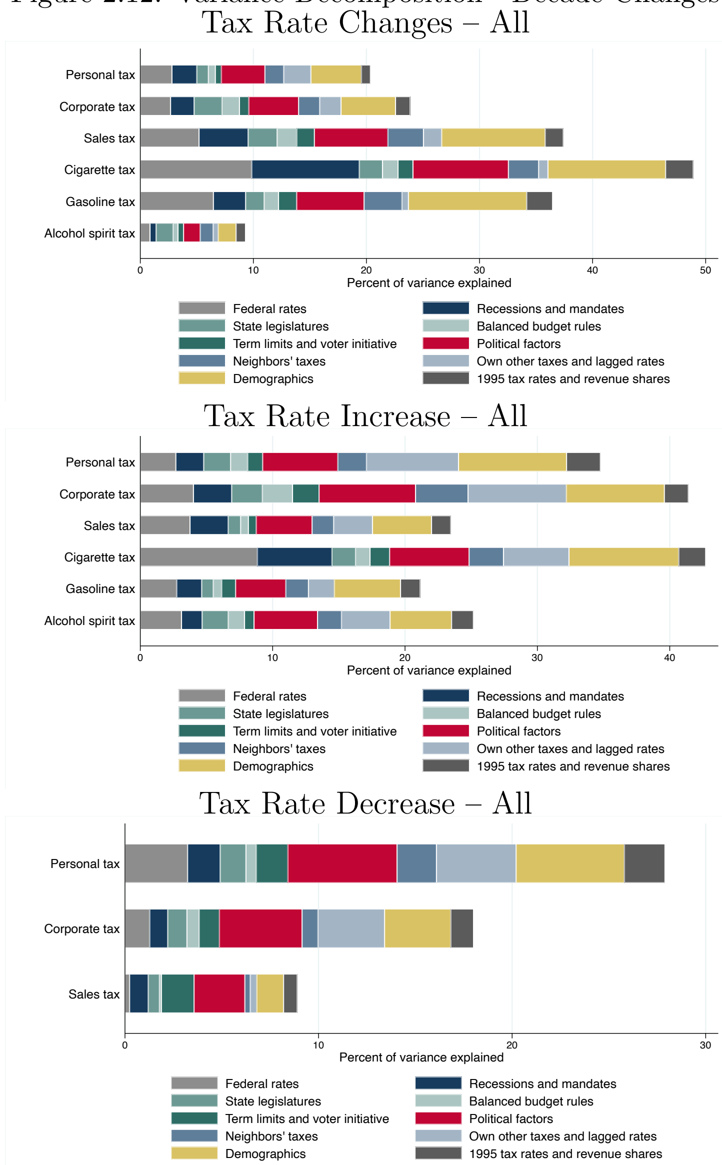


Notes: This figure shows the Shapley variance decomposition of adjusted R^2 for (a) all tax rates in percentage points or in 2020 dollars; (b) all tax changes (i.e., differences between a given year's tax rate and the previous year's tax rate) in p.p. or in \$2020; (c) same as (b) but only including 50% largest tax changes. All decompositions use the 166 variables summarized in Table 2.1 plus state and year fixed effects.

Figure 2.11: Variance Decomposition

Notes: This figure shows the Shapley variance decomposition of adjusted R^2 for (a) all tax rate increases (indicators for years when a tax increase occurs) and (b) all tax decreases (indicators for years when a tax decrease occurs); (c) and (d) – 50% largest tax increases and decreases, respectively. All decompositions use the 166 variables summarized in Table 2.1 plus state and year fixed effects.

Figure 2.12: Variance Decomposition - Decade Changes



Notes: This figure shows the Shapley variance decomposition of adjusted R^2 for (a) all tax rate changes from one decade to the next in percentage points or \$2020, (b) all tax rate increases (indicators that tax rate increased over the decade), (c) all tax rate decreases (indicators tax rate decreased over the decade). All decompositions use the 166 variables summarized in Table 2.1. For decreases, excise taxes are omitted because they are very infrequent.

Chapter 3

Geographic Variation in C-Sections in the United States: Trends, Correlates, and Other Interesting Facts

3.1 Introduction

C-section is the most common surgical procedure performed in the United States. The decision to proceed with a C-section is subject to considerable provider discretion, as the conditions precipitating the use of C-section are multi-faceted, and the consequences of unnecessary C-sections are highly-debated (e.g., Hyde et al. (2012); Shah (2017); Keag, Norman and Stock (2018)).[1] The complexity of this decision process makes it challeng-

[1] Conditions precipitating the use of C-section include stalled labor, distressed infant, breech birth, multiple birth pregnancy, placenta previa, prolapsed umbilical cord, and a previous pregnancy delivery via C-section. Placenta previa is a condition whereby the placenta covers the cervix, and a prolapsed umbilical cord occurs when the umbilical cord precedes the infant's head during delivery. See `https://www.mayoclinic.org/tests-procedures/c-section/about/pac-20393655`.

ing for providers to reach towards an appropriate use of C-section. As a consequence, providers sometimes rely on simple heuristics to guide their decisions (Singh, 2021) which often lead to worse outcomes. Given the ambiguity in whether or not a C-section is appropriate, it is not surprising that rates of C-section vary widely across geographies. For example, in 2017, the 95$^{\text{th}}$ percentile of county C-section rates was 39.4 percent, whereas the 5$^{\text{th}}$ percentile of county C-section rates was 24.4 percent.

Geographic variation in health care inputs is present in numerous other domains, from tonsillectomies in children (Glover, 1938), to catheterization in heart attack patients (e.g. Molitor, 2018), and for medical spending more generally (e.g. Wennberg and Gittelsohn, 1973; Finkelstein, Gentzkow and Williams, 2016). Frequently, areas using health care more intensively have no better outcomes than areas with less intensive use. This phenomenon has given rise to the natural question: is the usage of health care inefficient in high usage areas such that reductions in care would increase welfare? In the context of childbirth, this type of logic has fueled a push to reduce the purported overuse of C-sections. For example, the Centers for Disease Control & Prevention's public health goals, *Healthy People 2030*, include a goal of decreasing rates of C-section for low-risk first births to 23.6 percent.[2] In California, the California Health Care Foundation in cooperation with the Robert Wood Johnson Foundation funded an initiative to reduce unnecessary C-sections through the provision of transparent hospital performance metrics and decision aids to providers.[3]

We expand upon existing work documenting geographic variation in C-section (e.g., Baicker, Buckles and Chandra (2006); Epstein and Nicholson (2009)). Our analysis examines variation in C-section usage and health outcomes across over a span of nearly 3 decades (1989-2017). We characterize this variation using insights from Currie and

[2] For 2019, the actual rate was 25.6 percent. Source: Authors' calculation.

[3] See `https://www.chcf.org/project/reducing-unnecessary-c-sections/`.

MacLeod (2017) and Chandra and Staiger (2007) who provide frameworks to evaluate appropriateness of health care usage (i.e., is the volume and type of care justified given the patient and provider environment?). The goal is to understand the role of place in the determination of C-sections. Our hope is to provide further evidence to inform the debate of whether C-section rates could be reduced without welfare consequences.

The area of maternal and infant health is a rich environment for the study of geographic variation in health care and health outcomes, in part due to the availability of data. Specifically, our birth and infant death data cover the universe of births and infant deaths. This is beneficial in that our analysis does not condition on health care usage or health insurance, two potential outcomes, and covers all geographies.[4] Our analytic sample is first-time mothers with a singleton birth. Focusing on this sample avoids the selection problems inherent in second- and higher-order births due to the strong impact of a previous C-section delivery on subsequent deliveries.

Our analysis is divided into two parts. In the first part, we document patterns in county-level C-section rates. We focus on three relationships: time series trends, geographic variations, and county-level associations between risk-adjusted C-section rates and infant and maternal outcomes. Much of the emphasis on reducing C-section rates focuses on low-risk births, for which C-sections are considered less appropriate and hence more prone to overuse. Thus, our analysis also considers these relationships separately across low- and high-risk births.

In the second part of the paper, we peer inside these county-level aggregates to detail the relationship between C-section need[5] (or equivalently, appropriateness) and C-section delivery across counties categorized by their overall C-section rates. We also

[4]Much of the literature on geographic variation in health care has focused on the Medicare setting, which shares the advantage of universally available data, but may be limited in its ability to speak to other domains of interest (due to the distinct patient population and insurance environment).

[5]We estimate C-section need from a model of medical and demographic covariates.

relate this measure of C-section need to health outcomes —allowing us to provide insights on whether higher rates of C-section lead to worse outcomes, separately for mothers most appropriate for C-section and mothers least appropriate for C-sections (whom are often a target of policy intervention). Given the documented disparities in care across racial groups, we also investigate racial differences in these associations.

Our analysis yields several key findings. First, while C-section rates have risen over time, the time trend in C-sections in the last decade has flattened. However, C-sections among high-risk, first-birth, singleton mothers continue to increase, albeit at a much slower rate than that experienced in the 2000s. On the other hand, among low-risk singleton first births, C-section rates have recently fallen. Second, geographic variation in C-section is persistent. A county with a 10 percentage point higher C-section rate at the beginning of our study period (1989-1991) has, on average, a 4.4 percentage point higher C-section rate for high-risk births and 4.7 percentage points higher rate for low-risk births at the end of our study period (2015-2017). Third, counties performing more C-sections for high-risk births also perform more C-sections for low-risk births—a 10 percentage point increase in high-risk C-section rates is associated with a 7-8 percentage point increase in low-risk C-section rates. Fourth, counties with higher C-section rates do not experience worse outcomes in regards to maternal morbidity, infant morbidity and neonatal mortality. Fifth, high C-section and low C-section counties respond comparably to measures of C-section need, although high C-section counties have higher underlying rates of C-section across the C-section need distribution. Despite the higher proclivity for surgical intervention, the incidence of poor outcomes in high C-section counties is not appreciably higher than that in low C-section counties across the distribution of C-section need. Sixth, non-Hispanic Black mothers with the highest measured C-section need receive C-sections *less* frequently than similarly-needy non-Hispanic white mothers, yet have *higher* rates of neonatal mortality and infant and maternal morbidity, potentially suggesting misallocation among Black mothers.

Collectively, our correlational results are consistent with geographic variations in C-section driven by place-based differences in surgical skill (skill in performing a C-section) rather than differences in diagnostic skill (skill in targeting interventions to the most appropriate patients; Currie and MacLeod (2017)). Under a model of surgical skill, reductions in C-section rates among low-risk births in high C-section places could unintentionally lead to reductions in C-sections among mothers and infants who would actually benefit from the procedure. Our results by racial group highlight possible welfare-reducing disparities in surgical intervention for Black mothers, especially those we estimate to be most in need of C-section delivery.

One challenge to isolating place effects is the endogenous sorting of providers and patients across geographies. While our paper does not provide a fully satisfactory solution to this fundamental issue, we control for a wealth of observable characteristics in an effort to address selection. Thus, all of our analysis should be rightfully interpreted as conditional correlations rather than causal effects. Nevertheless, the hope is that the compendium of descriptive statistics in this paper expands the narrative in regards to C-section usage across the United States.

Our paper fits into a broader and voluminious literature on geographic disparities in health care access and health outcomes.[6] However, the bulk of this literature focuses on the Medicare population. The nature of Medicare is advantageous for examining geographic differences; it covers the entire United States and provides relevant and objective health measures (e.g., mortality). Place-based differences are dramatic for infant & maternal health, too, albeit less studied.[7] Recent work on maternity ward closures

[6] Cites include but are not limited to Wennberg and Gittelsohn (1973); Baicker, Buckles and Chandra (2006); Chandra and Staiger (2007); Skinner (2011); Finkelstein, Gentzkow and Williams (2016); Molitor (2018); Deryugina and Molitor (2020, 2021).

[7] An exception includes Baicker, Buckles and Chandra (2006).

(Battaglia, 2022; Fischer, Royer and White, 2023) hint at a role of place-based factors in C-section usage. The study of infant health shares some of the benefits of studying Medicare —full coverage of all geographies in the United States and a focus on an at-risk population. While geographic disparities in the domain of childbirth are important on their own, these gaps also hold significance for later life outcomes due to the long-run effects of early life health shocks (Almond and Currie, 2011) and the sizable impact of improvements in infant health on overall mortality (Cutler, Deaton and Lleras-Muney, 2006).

3.2 Geographic Variation in Health Care and the Appropriateness of C-Section

Our work is connected to and informed by three related but distinct sets of studies. Baicker, Buckles and Chandra (2006), using 1995-1998 natality data, study geographic variation in C-section use to understand whether higher C-section use is medically appropriate. Currie and MacLeod (2017) probe further to discern C-section appropriateness via the development of a two-dimensional model of physician decision-making with both surgical and diagnostic skill. Lastly, outside of the domain of childbirth, Chandra and Staiger (2007) use a Roy model with productivity spillovers to explain a number of patterns underlying geographic variation in heart attack treatment rates. We discuss these studies below to provide a collection of frameworks and mechanisms that will help guide the interpretation of our set of facts.

Baicker, Buckles and Chandra (2006) examine the use of C-section among normal and low birthweight infants in the 198 most populous counties in United States from 1995 to 1998. The data they utilize are the same as our data but cover a shorter time period (our sample covers 29 years) and fewer counties (our sample includes 2,346 counties). They find the strongest correlates with geographic patterns to be health system characteristics such as provider density, the capacity of the medical system, and medical malpractice liability, as opposed to patient characteristics. They conclude that higher C-section usage areas perform C-sections on less appropriate patients and do not see better outcomes. Their results are consistent with a model of physician decision-making whereby physicians rank patients by C-section appropriateness and perform C-sections until reaching some threshold defined by outside influences (i.e., non-focal-patient characteristics), such as the availability of health care services and malpractice policies, rather than the decision being dictated by allocative efficiency. Our study complements Baicker, Buckles and Chandra (2006) by examining geographic differences in the use of C-sections across the distribution of patient appropriateness, by considering more health outcomes, and by expanding the timespan and geographic coverage of the analysis.

The work of Currie and MacLeod (2017) leads to further insights about C-section appropriateness. They use detailed birth certificate data from New Jersey between 1997 and 2006 to model physician treatment decisions as both a function of surgical and diagnostic skill. Diagnostic skill measures the degree to which the physician makes the correct decision for delivery, whereas surgical skill reflects how well the physician carries out that decision. A physician with higher diagnostic skill responds more to the underlying characteristics of the patient, essentially improving the targeting of intensive treatments. Currie and MacLeod (2017) conclude that an increase in diagnostic skill leads to a decline in C-sections among low-risk mothers but a rise in C-sections with better outcomes for high-risk mothers, while an increase in surgical skill raises C-section rates for all. In our analysis, we see that high C-section areas tend to raise C-section rates across the distribution of patient appropriateness, consistent with differences in procedural skill as the main driver. On the other hand, our examination of patient race suggests that diagnostic skills are also at play, as we find the treatment of Black mothers to be less sensitive to their need.

Finally, moving outside of the healthcare of childbirth, Chandra and Staiger (2007) provide an alternative explanation to the frequent "flat of the curve" observation in the healthcare literature (i.e., areas that deliver more intensive medical care have the same or worse health outcomes). They adopt a Roy model whereby there are two treatment options —invasive and non-invasive —and the physician maximizes utility for each patient, which is a function of cost and survival. Survival from a particular procedure is a function of the fraction of individuals in the area who undergo that procedure. This particular modeling assumption leads to productivity spillovers that are harmful to patients better suited for non-invasive care. Such spillovers could occur because of physician learning, a phenomenon backed by several studies including Coleman, Katz and Menzel (1957). Both Currie and MacLeod (2017) and Chandra and Staiger (2007) share the feature that a uniform decrease in the rate of intervention cannot be Pareto improving, as appropriate cases will either receive interventions less often or, when they do receive the intervention, their provider will be less skilled in performing it. These models highlight the importance of examining treatments and outcomes across the distribution of patient appropriateness in understanding variation across providers, across space, or across populations.[8]

[8] Chandra and Staiger (2020) expand on their earlier paper by studying who is treated intensively for heart attack within hospitals, focusing on the distinction between allocative inefficiency and comparative advantage. Allocative inefficiency in that framework incorporates the notion that geographic variation arises when some areas perform too much intensive care while other areas perform too little. On the other hand, comparative advantage could lead to geographic variation if some areas are better at performing more intensive care. Their main conclusion is that much of hospitals' treatment overuse is due to incorrect beliefs about the benefits of treatment. Related work in this vein includes Chan, Gentzkow and Yu (2022), Abaluck et al. (2016), Silver (2021) and Mullainathan and Obermeyer (2022), all of which provide evidence of inefficiencies in the allocation of medical care underlying standard patterns of variation in treatment and testing rates across places and providers.

3.3 Data

3.3.1 Detailed Natality Files

Our primary data source is the Detailed Natality Files for the United States for the period 1989 to 2017.[9] These data cover the universe of births in the United States. They are derived from birth certificates and contain information on demographic characteristics of the parents, infant conditions (e.g., birthweight, gestational length), and procedure use (e.g., C-section, induction, ventilator, neonatal intensive care unit usage). We use the restricted version of these data with county of occurrence and county of residence information, which is the most granular geographic information available on the national data files. We tabulate our county-level statistics by county of residence to limit endogeneity concerns stemming from mothers travelling to other counties for care they prefer.[10]

We make several sample restrictions. First, since our focus is on C-section use and there is considerable path dependency in C-section use (i.e., many hospitals require that mothers deliver via C-section for higher-order births if they had a C-section for an earlier birth), we narrow our sample to first-birth mothers for our main analyses. Second, we include only singleton births in the main analysis sample, as the risk factors associated with multiple births are distinct. Third, very small counties experience very few births each year, and thus their rates of C-section are quite variable from year to year —largely due to noise. We exclude counties that ever have fewer than 100 births in a year.[11]

[9] We exclude more recent years because at the start of our project, the most recent infant mortality data was for 2017.

[10] For 75 percent of our sample, the county of occurrence and the county of residence are identical.

[11] Alternatively, one might prefer a selection criteria based on population since we study births. If we consider all counties with a population exceeding 9,091 persons for all years (i.e., a threshold of 100 births with a birth rate of 11 births per 1,000 population equates to a cutoff of 9,091 persons), our sample changes only slightly. One percent of counties (0.13 percent of all births) are included in our current analysis but would be excluded from the population-based selection rule. On the other hand, 2.7 percent of counties and 0.28 percent of births would be included under a population-based selection

Since our analyses use data over time, it is important to characterize the significant changes in practice recommendations regarding C-sections occurring during this period. One major shift occured starting in 1998, when the American College of Obstetricians and Gynecologists' (ACOG) bulletin yielded caution to administering a vaginal birth after a previous C-section. While this change is not directly related to our analysis of singleton first births, a second change occured in 2001, when ACOG's bulletin began to recommend C-sections for a breech birth (Oster, 2018), following on the results of the Term Breech Trial (Hannah et al., 2000).

Defining Low- vs High-Risk for a Poor Infant Health Outcome & C-Section Need

Our analysis necessitates distinguishing births by their underlying risk. We characterize a risk type in two different but related ways. First, we categorize mothers as either high- or low-risk of a poor infant health outcome. A birth is high-risk if it exhibits any of the following characteristics: preterm (less than 37 weeks gestation), maternal age under 18 or over 35, 19 or more prenatal visits, pregnancy-associated hypertension, maternal diabetes, eclampsia, or breech. We define a birth as low-risk if it exhibits none of these high-risk characteristics, generally following Card, Fenizia and Silver (2023).[12]

Second, in later analyses, we model the probability of a C-section birth. Our predicted probabilities, which we deem "C-section need" or "C-section appropriateness" come from a model with maternal age, gestational age, prenatal visits, growth restrictions, breech, eclampsia, pre-eclampsia, and diabetes, in addition to several demographic and birth characteristics which have strong predictive power for C-section rates among singleton first births. This set of characteristics (excluding the demographic and birth characteris-

rule but are excluded under our current selection criteria.

[12]Departing from Card, Fenizia and Silver (2023), we also include diabetes as a risk factor and exclude body mass index from our main analysis because it is only available for more recent years.

tics) combines wherever possible those used in Card, Fenizia and Silver (2023) with those used in Currie and MacLeod (2017). However, unlike Currie and MacLeod (2017), we only study first-birth singleton births, so previous birth history variables (i.e., previous C-section and parity) are excluded.

Our departure from Card, Fenizia and Silver (2023) and Currie and MacLeod (2017) in our inclusion of various birth and demographic characteristics is motivated by our goal of understanding the role of place in the determination of C-section. Our ideal thought experiment to identify these place-based effects would involve random assignment of otherwise identical women to different counties within the United States. However, there is selective migration of individuals across counties in the United States, making the ideal thought experiment hard to emulate. To remove the part of the variation in place effects due to this selection and more closely resemble the thought experiment involving contrasts of otherwise identical mothers, we control for several covariates highly correlated with C-section.

Table 3.1 further details these distinctions. The column labelled "High-risk (vs. low-risk) births" lists the characteristics determining the high-risk/low-risk designation. The remaining four columns list the variables for four different models of C-section need we estimate. When we control for the set of covariates listed in the "All covariates" column, we call this an "adjustment for all covariates," whereas an adjustment for the medical risks listed in the "Medical only" column we call an "adjustment for medical covariates." We contrast our selected characteristics directly in Appendix Table A.9 with those used by Card, Fenizia and Silver (2023) and Currie and MacLeod (2017). As our sample period covers an extended period, we are unable to include characteristics available only in older data (e.g., cord prolapse). The R^2's of the four C-section need prediction models are quite similar, all between 12 and 15 percent —implying that the additional variables beyond the medical risk variables provide little extra explanatory power for C-sections.

However, as we will see, these additional variables spread out our distribution of predicted C-section need.

To assess how these two classifications compare quantitatively, Table 3.2 displays the cross-tabulations for the 1989-1991 and the 2015-2017 periods separately. In light of the fact that C-section need is a continuous measure, we create a dichotomous measure above and below 0.6. Virtually all births with a high predicted need of C-section fall into our high-risk birth classification. About two-thirds of births with predicted C-section need below 0.6 we deem as low-risk births. These patterns are similar if we use the predicted C-section need model with only medical covariates (see Appendix Table A.10).

3.3.2 Other Data

We use U.S. county population estimates from the National Institutes of Health (NIH) Surveillance, Epidemiology, and End Results Program (SEER). In addition, we use data on county demographics as well as the supply of healthcare facilities and practitioners from the Area Health Resource File (Griffith et al., 2021). Finally, we use data on medical malpractice liability payments at the state level from the National Practitioner Data Bank.

3.3.3 Descriptive Statistics

Table 3.3 presents summary statistics for our sample of natality-data counties. For the natality data, we calculate these statistics separately for two periods of time (1989-1991 and 2015-2017) bookending the sample period. Much of our analysis will focus on contrasts between these two periods, with the earlier period characterized by relatively low and stable c-section rates, and the latter period characterized by relatively high and stable c-section rates, at the national level.

Our sample from the natality data covers a balanced panel of 2,346 counties once we drop counties experiencing few births. This sample covers 75 percent of all counties in the United States and 98.6 percent of births. Counties vary widely in their population size and number of births, and thus in our analysis we weight counties by population size as appropriate. The 75$^{\text{th}}$ percentile of the number of births is over 4 times the number of births for the 25$^{\text{th}}$ percentile. The share of births that are singleton first births (the primary focus of our analysis sample) is 32.7 percent (1989-1991) and 29.4 percent (2015-2017) with roughly one-third of these births categorized as high-risk and nearly all of the rest as low-risk.[13] Overall, the difference in C-section rates between the 25$^{\text{th}}$ percentile and the 75$^{\text{th}}$ percentile is nearly 8 percentage points for both periods (between a 32 to 39 percent difference). In percentage-point terms, the variation is larger amongst high-risk singleton first births than amongst low-risk singleton first births.

Looking across the two time periods and all births (including higher order births), one of the most stark patterns is the change in the C-section rate over time —rising from 0.24 in the earlier period to 0.32 in the later period. Interestingly, however, the standard deviation of the C-section rate across counties has stayed stable. Contrary to what one would have predicted given the increase in C-sections, the fraction of births that are high-risk singleton first births has dropped slightly in the most recent years.[14]

[13]A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either high or low-risk.

[14]Another pattern worth mentioning is the fall in infant morbidity. However, as the underlying measures for this composite measure change over time, it is not possible to know whether infant morbidity is truly falling.

3.4 Results

3.4.1 Time Trends in C-Section Usage and Geographic Variation

Figure 3.1 documents the trend in county-level C-section rates over time, including the mean, median, and the interquartile range. Overall C-section rates for all births in the late 1980s and early 1990s hovered at 22 percent, rose rapidly between 2000 and 2010, and then flattened to their current level of 32 percent. The variation in C-section rates across counties has not changed appreciably over time as seen earlier in Table 3.3.

A naive but incorrect interpretation of the patterns in Figure 3.1 is that the rise in C-sections is due to a decline in vaginal births after a previous C-section (the practice recommendation came out in 1998). However, the time trend for singleton first births, the focal sample of our later analysis, mirrors that of the overall trend. Contrasting high-risk and low-risk singleton first births in Figure 3.2, the time trends share the same common shape (i.e., falling slightly during the 1990s, a significant rise during the 2000s, and a leveling during the 2010s) with a few nuances. First, C-section rates among the low-risk group are roughly fifty percent smaller than that of the high-risk group. Second, C-section rates for low-risk mothers have been declining during the past decade, whereas for high-risk mothers they continue to climb albeit at a slower rate than that experienced during the 2000s. Lastly, as with all births combined, cross-county variation for both high- and low-risk singleton first births has remained relatively stable in magnitude.

3.4.2 Place Effects in C-Section Usage

Digging into the cross-sectional variation in these C-section rates, there are significant geographic differences in C-section for high-risk singleton first births as seen in Figure

3.3. The change in the shading of the map is quite dramatic moving from the 1989-1991 period to the 2015-2017 period, with rates of C-section among high-risk mothers rising significantly.

Looking at Figure 3.3, in the 1989-1991 period, the South exhibited the highest rates of C-section while the lowest rates were in the Mountain Census region. At the end of the analysis period, the variation looks quite similar —indicative of persistence in these rates over time albeit with higher levels. Geographic patterns for high-risk singleton first births for 1989-1991 deviate from those for low-risk singleton first births (Figure 3.4) in that high rates of C-section are scattered throughout counties from North to Southeast of the Mississippi rather than almost exclusively in the South.

A question arising from Figures 3.3 and 3.4 is the degree to which the geographic patterns are stable over time. Our data span nearly 3 decades, facilitating a longer-term examination of the persistence of geographic practice variation than is feasible in most settings. We display the graphical evidence of persistence in Figure 3.5, a binned scatter plot of "all covariate-adjusted" C-section rates across time periods. Recall that we refer to "all covariates" model as including the controls listed in Table 3.1. We compare the county's C-section rate over a three-year period at the beginning of the sample period (1989-1991) to the rate at the end of the sample period (2015-2017). Averaging across three years and weighting by the number of births in each county mitigates some of the concern about small county-level samples dampening the across-time correlation. Despite substantial changes in medical technologies and obstetrics practices over our study period, persistence is present for both low-risk and high-risk singleton first births. The slope of the relationship is slightly stronger for low-risk singleton first births but qualitatively the slopes are similar across the risk types.[15] For high-risk singleton first births, counties

[15]The figures look similar if we adjust for medical covariates only. The slopes of the relationships are slightly larger. See Appendix Figure A.25.

with a 10 percentage point higher C-section rate in the 1989-1991 period experienced a 4.4 percentage point higher C-section rate in the 2015-2017 period.

Another dimension of interest with these place effects is whether the place effects are correlated across the two risk types. Are places that perform more C-sections for high-risk mothers where a C-section may be more appropriate also conducting more C-sections for low-risk mothers? Using the same time periods as in Figure 3.5, Figure 3.6 displays a binned scatterplot of the relationship between C-section rates for high-risk first births and that for low-risk first births. The C-section rates are highly correlated across the two risk types; the slope is 0.85 for the 1989-1991 period and smaller at 0.68 for the 2015-2017 period.[16] High C-section counties for high-risk women are high C-section counties for low-risk women and vice versa, a pattern we delve into further in subsequent analyses.

If diagnostic skills were the prevailing determinant of treatment choices (Currie and MacLeod, 2017), then places with better diagnosticians would arguably have high rates of intervention for the high-risk populations and lower rates for the low-risk populations. While our patterns do not at all rule out diagnostic skill as an important factor, the findings from Figure 3.6 suggest a less dominant role for diagnostic skill. Rather, the data are more consistent with variation in C-section arising from a common force that leads to higher intervention rates across the distribution of patient types. For example, providers in different areas could differ in their overall inclination for intervention, with providers in high C-section counties more inclined to use medical interventions, including C-section, regardless of risk type. The malpractice environment, either in terms of the cost of malpractice insurance or the likelihood of a successful malpractice lawsuit, could lead providers to do more C-section deliveries for both low- and high-risk mothers (Baicker,

[16] An adjustment for medical covariates only leads to slightly-altered slopes (0.88 and 0.72 respectively). See Appendix Figure A.26.

Buckles and Chandra, 2006). In Currie and MacLeod (2017), by performing a high volume of C-section deliveries, providers may gain a comparative advantage in C-section deliveries relative to vaginal deliveries —resulting in higher rates of C-section across the continuum of C-section need.

3.4.3 Are Higher Rates of C-Section Correlated With Infant and Maternal Morbidity and Infant Mortality?

Without data on outcomes, it is difficult to assess welfare. To this aim, we explore how county C-section rates vary with infant and maternal morbidity and neonatal mortality.[17]

Neonatal Mortality

In Figure 3.7, we present the extent to which the cross-county patterns in C-section rates correlate with neonatal mortality for both analysis periods and both risk types.[18] Across both time spans for high-risk singleton first births, the slope of the relationship is positive (i.e., pslaces with higher C-section rates have higher rates of neonatal mortality). The slope is modest —for the 1989-1991 period, increasing the C-section rate among high-risk singleton first births by 1 standard deviation is associated with an additional 0.36 neonatal deaths per 1,000 births (a 3 percent increase). The relationship is flatter for the later period —falling by over 50 percent. In the bottom panel of Figure 3.7, we repeat

[17]We choose neonatal mortality, as opposed to infant mortality, as neonatal mortality is more closely connected to hospital interventions (Cutler and Meara, 2000).

[18]We have investigated this relationship, along with the relationship between C-section rates and infant and maternal morbidity, across three other subsamples, including births in counties offering continual hospital-based maternity ward services over the entirety of our study period, births to white mothers, and non-Medicaid births (only for the more recent period as Medicaid information was unavailable for the earlier time period). Qualitatively, the results look similar. In particular, higher rates of C-section are generally associated with the same or better health outcomes. However, for high-risk white mothers, the slope of the C-section vs. neonatal mortality relationship is twice the size of that for the analysis sample. Specifically, the slopes for high-risk white mothers in the 1989-1991 and 2015-2018 time periods are 8.1 and 3.9, respectively.

this exercise except with low-risk singleton first births. Similar to the top panel figure for 1989-1991, there is an upward-sloping association but the slope is substantially smaller —6.5 percent the size of that for high-risk births for 1989-1991. The slope flips signs for 2015-2017 and is still substantially smaller than the analogous slope for high-risk mothers.[19] None of the estimated slopes are statistically different from 0 at the 5 percent level. Collectively, these correlations imply a weak association between C-section rates and neonatal mortality, especially for low-risk births.

The correlations found in Baicker, Buckles and Chandra (2006) differ somewhat from those of Figure 3.7. In their study of the 198 most populous counties for the years 1995-1998, they estimate a negative relationship between C-sections and neonatal mortality. A one standard deviation increase in the C-section rate is correlated with a drop in neonatal mortality of 0.02 per 1,000 births. However, none of these correlations are statistically different from zero, so the general conclusion from both our analysis and that of Baicker, Buckles and Chandra (2006) is that there is a weak association between cross-sectional county variation in C-section use and county-level neonatal mortality rates.

Infant Morbidity

Figure 3.8 displays relationships between county c-section rates and our constructed measure of infant morbidity. Following Fischer, Royer and White (2023), this morbidity measure is defined as the presence of any of the following conditions: moderate or heavy meconium staining, birth injury, seizure, or assisted ventilation. Note that due to changes in the birth certificate, the first two conditions (meconium and injury) are only available during the 1989-1991 period and thus infant morbidity indices are not directly comparable over time. The relationships in this figure are consistently downward-sloping with higher

[19] In a model adjusted for medical covariates only (see Appendix Figure A.27), one slope changes sign (i.e., the slope for high-risk mothers in the 1989-1991 period is negative) but the general conclusions hold (i.e., weak associations between C-section rates and neonatal mortality).

C-section rate counties having improved infant morbidity.[20] The associations are stronger than those for neonatal mortality. A 10 percentage point increase in the C-section rate is associated with changes in infant morbidity that are 12 to 18 percent of the mean. For neonatal mortality, the estimated effect of an equivalent rise in the C-section rate is 2.0 to 3.6 percent of the mean.

For both low-risk and high-risk singleton births, the relationship is weaker for the later time period.[21] However, one should be cautious comparing these measures over time due to the mentioned changes over time in the index.

Maternal Morbidity

As a final set of health outcomes, we examine maternal morbidity in Figure 3.9. Following Fischer, Royer and White (2023), maternal morbidity is defined by the presence of any of the following conditions: febrile, excessive bleeding, seizure, transfusion, third or fourth degree perineal laceration, ruptured uterus, unplanned hysterectomy, or admission to ICU.[22] Due to changes in the birth certificate, the first three conditions (febrile, excessive bleeding, and seizure) are only available during the 1989-1991 period, and the remaining five conditions are only available during the 2015-2017 period; as a result, maternal morbidity indices are not comparable over time. Here we do not consider maternal mortality, due to issues of undercounting (MacDorman and Declercq, 2018) and its fortunate infrequency. However, if we expand our set of maternal health components to include maternal mortality, the shape of the relationship is consistent with those in

[20] Looking at the components of infant morbidity separately, the county-level relationship between C-section rates and each infant morbidity component is downward-sloping.

[21] An adjustment for medical covariates only, instead of the full set of covariates, has minimal impact on our conclusions. See Appendix Figure A.28.

[22] Note that these measures exclude many complications occurring during the post-partum period, as these morbidity measures are collected when the birth certificate is completed (typically in a one or two days after the birth). Thus, complications such as ruptured sutures, postpartum hemorrhage, or infection are excluded from our maternal morbidity measure.

Figure 3.9.[23]

The most apparent and consistent pattern across the four graphs in Figure 3.9 is the downward-sloping relationship.[24,25] This relationship is somewhat stronger than that for infant morbidity. A 10 percentage point increase in the C-section rate is associated with changes in maternal morbidity that are 14 to 32 percent of the mean. Taken together, the morbidity measures point to health-improving correlational returns to higher C-section rates, whereas the association of a county's C-section rate and neonatal mortality rates is weaker —exhibiting both positive and negative, albeit small, correlations.

3.4.4 Appropriateness of C-Section and Its Use Across Geographies

We next delve further into these county-level aggregate statistics by examining the relationship between C-sections and health outcomes through the lens of the appropriateness of care. Chandra and Staiger (2020) propose testing for allocative inefficiency by examining whether the probability of treatment varies across geographic space for

[23]However, this analysis treats each outcome equivalently, which may be inappropriate since a maternal death should arguably have more weight given its severity.

[24]The slopes are qualitatively similar if we control for the medical covariates rather than the full set of covariates. See Appendix Figure A.29. In an unreported analysis available upon request, we compare the magnitude of our results to those in Card, Fenizia and Silver (2023) for the one overlapping measure of maternal morbidity between our studies: perineal laceration. Our results are directionally similar, though substantially smaller in magnitude. Card, Fenizia and Silver (2023) find that delivery at a high C-section hospital reduces the incidence of laceration by 7 percentage points, while we find that delivering in a high C-section county is associated with a 0.5 percentage point lower rate of laceration. The sources of variation in each study are distinct though —the present study using cross-county comparisons and Card, Fenizia and Silver (2023) utilizing a causal framework leveraging relative distance to a high C-section vs. a low C-section hospital.

[25]If we examine the maternal morbidity components separately, most of the estimated slopes are negative with a few exceptions. This includes maternal seizure (both high- and low-risk births) during the 1989-1991, transfusion for high-risk births during the 2015-2017 period, and ruptured uterus for high-risk births during the 2015-2017 period. For first births, a ruptured uterus would likely precipitate a C-section (i.e., not be caused by a C-section). None of the estimated slopes from these positively-sloped relationships, however, are statistically significant. Collectively we feel that these additional analyses, with a few exceptions, point to similar conclusions based on our maternal morbidity index measure.

patients of similar predicted probabilities of treatment, while the model in Currie and MacLeod (2017) suggests that procedural skill could increase C-section rates across the distribution of patient appropriateness. In essence, we want to know how a particular mother's probability of a C-section varies across counties when the mother moves from a high C-section county to a low C-section county or vice versa. Of course, it is impossible to observe a mother delivering in two places at once. The feasible comparison is to contrast women with similar C-section propensities (based on their observable characteristics aside from their place of residence) and see how their probability of C-section varies across counties differing in their propensity for C-section deliveries.

To do this type of analysis, we create a model of C-section appropriateness. We first obtain covariate-adjusted C-section rates for year t by regressing whether a birth is delivered via C-section on county fixed effects and a comprehensive set of individual-level characteristics for year t. We provide details on these characteristics in Section 3.3.1. From that regression, the covariate-adjusted county C-section rate is the estimated county fixed effect. The individual predicted probability of C-section ("C-section need") is, for each individual birth, the prediction using all of the individual covariates but excluding the county fixed effects. The intent of this measure of appropriateness is to remove the contribution of the county and only consider individual-level circumstances. Thus, individuals with the same individual-level covariates will have the same appropriateness even if they live in counties that differ in their underlying rate of C-sections. However, because of possible selection on unobservables (e.g., physicians sorting across geographies based on surgical skill, heterogeneous patient preferences across counties), we are cautious about interpreting the correlations we examine. Additionally, the simultaneous partialling-out of county fixed effects means that the risk model is identified using only within-county variation in patient characteristics, addressing the potential bias that could arise from high-risk mothers sorting to high-C-section counties.

Appendix Figure A.23 displays the distribution of our predicted rates of C-section based on a prediction model using all covariates for 2015-2017. The distribution is bunched around 0.25. Nearly all of the distribution has a predicted probability of C-section below 0.4. There is little mass between 0.4 and 0.8. As such, our statistical inference will be much less precise for probabilities outside of the range of 0.1 to 0.4. One benefit of using the full set of covariates for adjustment is that the predicted probability of C-section is smoother (see Appendix Figure A.30 for the distribution of C-section need for the model with only medical covariates). In the medical covariate model, due to the few number of discrete medical covariates, mothers are bunched into fewer values.

This model of appropriateness and its usefulness for categorizing mothers based on their C-section probability rests on several assumptions worth discussing. First, our prediction model assumes the individual-level covariates are additively-separable risk factors for a C-section apart from geography.[26] Second, while the natality data cover many of the inherent risks of delivering vaginally, the set may be incomplete, especially if physicians have access to additional patient information that influences their decision making. Third, this model assumes that appropriateness can be constructed from treatment choices made in the aggregate. To the extent that this model "learns from the crowd," if the crowd is wrong, the model will be as well. These concerns notwithstanding, our C-section risk score is strongly related to the probability of C-section, as shown below, suggesting it provides a useful ranking of patients for analyzing the allocation of C-sections across counties.[27]

[26] A fully interacted model would relax this assumption, but the large number of covariates considered would raise concerns about the stability of estimates given small cell sizes.

[27] Related work by Currie, MacLeod and Van Parys (2016) has addressed this issue by estimating patient appropriateness using only the decisions of better-trained providers. We do not have details on individual-level providers, but we have explored how our appropriateness measure varies with the sample used to estimate the appropriateness model (i.e., we estimate the appropriateness model using only the top 25 percent of counties in terms of the lowest infant mortality rates). The correlation of our C-section need measure and this modified C-section need measure is 0.996 —indicating some insensitivity of the prediction model to the sample used to deem appropriateness.

C-Section Need and the Use of C-Section

Figure 3.10 plots the measure of individual-level C-section need measured in percentage points against the probability of C-section for three geographic groupings (low-rate C-section counties, medium-rate C-section counties, and high-rate C-section counties).[28,29,30] Moreover, the consistency of the shape of the curve across the three groups indicates that these three types of counties respond similarly to the underlying patient-level characteristics in their determination of whether to perform a C-section.

Another notable feature of this figure is that the curves do not overlap or cross —meaning that higher C-section areas perform more C-sections across the entire support of predicted C-section risk probabilities. If worse diagnostic skills in high C-section areas drive the observed C-section patterns, the curve for high C-section counties should be flatter than that for lower C-section counties.

That is, under a model of diagnostic skill delivering geographic variation in C-section usage, the gap in usage across counties arranged by their C-section rates should be larger for those least appropriate for C-section than the gap across counties for those most appropriate for C-section. Instead, the geographic variation is more consistent with what Currie and MacLeod (2017) term "surgical skill." Physicians in areas with elevated C-section rates may have a comparative advantage in the surgical procedure (or at least believe they have such an advantage, as discussed in Chandra and Staiger (2020)),

[28] The C-section rate is adjusted for all covariates and is measured for singleton first births.

[29] In honor of David Card, to preclude his frequent questioning of where the 45 degree line is, we include the 45 degree line. The raw C-section rate increases with the individual predicted C-section rate, indicating that rates of C-section are responsive to our underlying measure of C-section need. This finding is true for all three groups across both time periods (1989-1991 in the top panel and 2015-2017 in the bottom panel).

[30] Considering a model adjusting for medical covariates only, the relationship between predicted probability of a C-section and the rate of C-section is similarly upward sloping and tracks the 45 degree line well. However, not surprisingly due to the lumpiness of the data, the patterns are less smooth. Also, for high levels of C-section need, C-section rates consistently fall below the 45 degree line —indicating that our prediction model overpredicts C-section rates. See Appendix Figure A.31.

leading them to do C-sections more frequently across the C-section risk distribution. Alternatively, physicians in these places may have weaker risk tolerances or face stronger malpractice environments, possibly leading to higher rates of intervention without any skill differences, as suggested in Baicker, Buckles and Chandra (2006). One other caveat to consider is that high C-section counties may be more sensitive to factors outside of our risk models that are present across the distribution of observed risk, such as fetal heart rate anomalies. In this case, our observed C-section usage patterns may be less sensitive to underlying risk than the true response function and thus, lead us to discount the diagnostic role in explaining geographic differences. However, it would be surprising to see such large and uniform differences across our observed C-section need measure (a pattern not consistent with diagnostic-skill explanations) alongside strong patterns of diagnostic skill in terms of unobserved risks.

Also evident in Figure 3.10 is that the majority of the distribution of C-section need falls between 0.1 and 0.5, as each marker in the figure represents a percentile bin. This pattern has not changed much between the two time periods. Only 4 to 5 percent of the sample has very high rates of predicted C-section need (completely accounted for by the presence of a breech birth), a percent that has increased slightly over the two time periods. At these percentiles, the rate of C-section is very high —0.8 or above. As the rest of the paper relies on analysis by C-section need, it is important to keep in mind where the marginal C-sections lie in this C-section need distribution. The top 20 percent of mothers in this distribution have a predicted C-section need exceeding 0.33, whereas the top 30 percent of mothers have predicted C-section need above 0.29 (see Appendix Figure A.23). A useful benchmark is the *Healthy People 2030* target of a 23.6 percent C-section rate for low-risk first births. Thus, taking into consideration that a C-section target for high-risk first births is higher, the marginal births are likely to lie between C-section needs of 0.29 and 0.33.

C-Section Need and Health Outcomes

To see how these practice patterns translate into outcomes, Figure 3.11 plots unadjusted rates of neonatal mortality as a function of C-section need (the same x-axis as in Figure 3.10).

With the exception of the highest percentiles of C-section need and the lowest percentiles of C-section need for the earlier period, rates of neonatal mortality are generally increasing with respect to C-section need.[31] This finding means that patients for whom a C-section may be more appropriate generally experience higher risks of neonatal mortality. Across much of the support of C-section need (i.e., between 0.1 and 0.4), rates of neonatal mortality are low and not visually distinct across low, medium, and high C-section counties. At higher levels of C-section need, neonatal mortality rates move together across the three types of counties with none of the low, medium, or high C-section counties consistently experiencing lower rates of neonatal mortality. From a statistical perspective, distinctions between the three county types is difficult with neonatal mortality because of its low incidence. Only for levels of C-section need exceeding 0.5 do we see neonatal mortality rise appreciably as a function of C-section need. This result serves as a good motivation for examining other outcomes such as infant morbidity and maternal morbidity which have higher incidence rates.

In Figure 3.12, we follow the same layout as Figure 3.11 but instead use maternal morbidity and infant morbidity as our outcomes. The shape of these relationships is quite distinct from Figure 3.11. In particular, for three of the four graphs (the exception is infant morbidity for 2015-2017), singleton first births highly appropriate for C-section (i.e., predicted C-section need exceeding 0.8) exhibit nearly the lowest rates of maternal or infant morbidity across the appropriateness distribution. It is plausible, however not

[31]We have investigated the odd spike in the 1989-1991 figure; it is due to the fraction of breeched births discretely changing starting with this percentile.

testable with our data, that there is a lot less ambiguity about the usefulness of C-section for these births, and because these mothers receive appropriate care, they experience better outcomes.

Despite having the highest rates of C-section, high C-section counties do not have the highest incidence of maternal morbidity and infant morbidity. For 1989-1991, rates of maternal morbidity and infant morbidity are the lowest amongst high C-section counties across all levels of C-section need. In the later period, high C-section counties exhibit equal or lower rates of infant and maternal morbidity for most of the support of C-section need.

One caveat to interpreting the results in Figure 3.12 is that some of the conditions underlying the morbidity measures may be mechanically related to the delivery type. The only underlying measure that stands out in this respect is perineal laceration, which is very unlikely for women who deliver by C-section. However, this condition is not available during the early period (1989-1991), where we also observe an improvement in maternal morbidity for higher C-section counties, indicating that our results are not dependent on this condition alone. Nevertheless, because these potentially more mechanical measures are related to welfare, we include them in our analysis.

Taking Figures 3.11 and 3.12 together with the caveat of a correlational analysis, the higher rates of C-section in high C-section areas do not appear to come at the cost of worse outcomes, a pattern that holds throughout the C-section need distribution. These conclusions are generally robust to only adjusting for medical covariates (see Appendix Figures A.32 and A.33). Of course, the outcomes we study are limited. Specifically, a C-section delivery may affect the probability of breastfeeding, future pregnancy issues such as those relating to the placenta, a mother's long-term physical health, the infant's ability to breathe, etc., which are not covered here.[32] A simple view of our findings

[32]Source: `https://www.nhs.uk/conditions/caesarean-section/risks/`.

is that there is limited evidence, in terms of minimizing morbidity risks, that there is overuse of C-sections among first-birth singleton mothers, though the generally higher costs of C-sections as well as the possible effects on outcomes we do not observe color this interpretation.[33]

Characterizing Low, Medium, and High C-Section Counties

The intent of our use of individual covariates is to remove the part of the variation due to the selection of people across geographic space. However, it is possible that low, medium and high C-section counties differ along other dimensions not yet explored, explaining why high C-section counties do not generally experience worse outcomes. To investigate selection in more detail, we contrast these counties in Tables 3.4 and 3.5. In Table 3.4, we investigate disparities across these three groups of counties based on data available in the natality files. In Table 3.5, we instead use data from the Area Health Resource File and the National Practitioner Data Bank to contrast counties.

In Table 3.4, counties with the highest C-section rates have fewer births than medium and low C-section counties. However, with the exception of preterm birth (gestational age < 37 weeks) where high C-section counties have elevated rates, the underlying risk factors for a C-section are similar. High C-section counties are more economically disadvantaged (higher fraction of Medicaid births) and have higher fraction of Black residents than do counties with lower C-section rates.

On economic dimensions, high C-section counties, relative to low and medium C-section counties, have depressed income per capita and higher rates of unemployment and poverty as shown in Table 3.5. These counties also have a higher share of minority

[33] Other studies, including Currie and MacLeod (2017) and Johnson and Rehavi (2016), conclude that there is overuse. Of course, a full accounting of the total welfare analysis should take into account the health consequences plus the resource cost (e.g., higher health care costs – Podulka, Stranges and Steiner (2011); Corry, Delbanco and Miller (2013)), which are unaccounted for here.

residents and fewer college graduates than low C-section counties. Collectively, area-level omitted variables bias arising from the economic and demographic factors of high C-section counties is correlated with worse outcomes for their mothers and their babies. Controlling for such selection would likely lead to even better comparative outcomes for high C-section counties.

Health care resources also differ by the three county types. However, it is less clear the direction of selection as the effects of the resources on outcomes are more ambigious. Hospital resources, measured in terms of hospitals per birth and newborn beds/bassinets, are more abundant in high C-section counties. In contrast, the per-capita number of physicians and obstetricians/gynecologists is lower in high C-section counties. Per-capita malpractice payments are also higher in the high C-section counties. On net, the answer to why and if the high C-section counties experience better outcomes is not simply explained away by selection, at least on the basis of the characteristics in Tables 3.4 & 3.5.

3.4.5 Appropriateness of C-Section and Its Use Across Racial Groups

Racial disparities in C-section use are significant. In 2017, the C-section rate for Black mothers was 35 percent whereas for white mothers it was 31 percent (Yang and Mullen, 2022). This gap has grown over the last decade. A key question is the degree to which Black and white mothers with similarly-measured C-section need have different propensities of having a C-section. Akin to the earlier figures, we plot our C-section need measure against the actual C-section rate in Figure 3.13 (a).[34] We exclude race from the prediction model and do not consider variation across counties by their underlying

[34]We display this distribution based on adjustment using only medical covariates in Appendix Figure A.30.

C-section rates because of concerns about sample size. We display the distribution of individual predicted probabilities in Appendix Figure A.24, divided by race. The distributions in panel (a) resemble the overall pattern. Looking across the two racial groups, non-Hispanic Black mothers' distribution has more density at smaller C-section need values. This pattern holds when body mass index (BMI) is added to the prediction model in panel (b), but for both groups, the inclusion of BMI spreads out the distribution —placing more density between 0.4 and 0.8.

Our first finding in Figure 3.13 is that, at the highest levels of C-section need, non-Hispanic Black mothers have lower rates of C-section than non-Hispanic white mothers. The pattern flips for predicted C-section rates below 0.6. Using 2016 natality data along with the 10-point scale of the Robson Ten-Group Classification System, Valdes (2021) also document higher rates of C-section delivery among Blacks at the high end of the "C-section risk" distribution and lower rates of C-section delivery at the lower end of the "C-section risk" distribution. More generally, our finding for Black mothers is consistent with the notion that minorities experience less tailored treatment choices, a finding that complements those in Johnson and Rehavi (2016) and Card, Fenizia and Silver (2023), among others.

In the remaining panels of Figure 3.13, we examine how these C-section differences relate to outcomes. Across the spectrum of C-section need, the rates of neonatal mortality for non-Hispanic Black mothers exceed that of non-Hispanic white mothers (panel (b)).[35] The gap is largest for mothers with highest C-section need. The patterns for infant morbidity resemble that for neonatal mortality but the gap emerges at a higher level of appropriateness (0.5 for infant morbidity and 0.3 for neonatal mortality). Like the infant health outcomes, maternal morbidity among Black mothers at high levels of C-

[35] This racial gap in neonatal mortality also appears when contrasting mothers of similar incomes (Kennedy-Moulton et al., 2022).

section exceeds that for white mothers, but unlike the infant health outcomes, maternal morbidity for Black mothers is lower than that for white mothers at predicted probabilities of C-section below 0.4. The sizable elevated risk of morbidity and mortality for Black mothers with a high C-section need holds when the prediction model includes only medical risks (Appendix Figure A.34) or additionally includes body mass index (Appendix Figure A.35), a factor mentioned as a contributing factor in the rise in C-section use (Chu et al., 2007). Taken literally and ignoring the issues of patient and provider selection, these results are suggestive that the rate of C-section is too low for Black mothers with the highest C-section need. High-risk Black mothers and their infants might fare better if their C-section rate was increased. This particular form of potential underuse among Black patients mirrors findings from the literature (e.g., Abaluck et al. (2016)).

3.5 Discussion and Conclusion

In this paper, we use data on births and infant deaths covering the entirety of the United States (1989-2017) to study county-level variation in the use of C-section, the most common surgical procedure. We document how county-level C-section rates have recently leveled off after a dramatic rise in the 2000s and have had strong persistence over the three decades. Looking through a lens of appropriateness of care, we detail how C-section rates in high C-section counties are higher than those in counties with lower C-section rates, both for high-risk and low-risk mothers. Nevertheless, there is not strong correlational evidence that these higher rates of C-sections in counties with greater propensities for C-section lead to worse health outcomes as measured in the birth and infant death data.

This observation that high C-section counties exhibit higher rates of C-section along the continuum of C-section appropriateness is more consistent with surgical skills, rather than diagnostic skills, driving the high C-section rates (Currie and MacLeod, 2017). Under a goal of reducing C-sections for low-risk mothers, our results imply potentially adverse consequences for high-risk mothers. These mothers might be expected to fare worse if the target leads to spillover effects onto higher C-section risk groups. In particular, the result of such a policy could be a fall in C-sections for mothers more appropriate for C-section or a drop in physicians' surgical performance through a detrioration of a physician's surgical skills.

Together these facts provide some insights on the key question in the geographic variation literature: are these geographic differences attributable to allocative inefficiencies (Chandra and Staiger, 2020) with some counties performing too many C-sections and others too few? Our answer to that question is not straightforward. As best we can measure, the outcomes of high C-section counties would worsen under a national standardization of care with a policy target of X percent of C-sections, where X is below the current rate in high C-section counties. On the other hand, setting a target above the current C-section rate of low C-section areas could improve outcomes. But these conclusions are subject to the caveat that high C-section counties are different. Nevertheless, on the basis of demographic and economic characteristics alone, these counties should fare worse in terms of outcomes than low and medium C-section counties, suggesting a muted role of selection in explaining our results.

One policy effort to standardize care and thus potentially remove geographic disparities, is the use of decision aids to guide treatment decisions (e.g., CHADS_2 for atrial fibrillation, (Abaluck et al., 2020)). However, the complexity of the C-section decision process does not lend itself well to the easy adoption of such decision aids. For example, in 2015, the World Health Organization adopted the Robson classification, a 10-group classification based on 5 maternal characteristics —parity, the presence of multiples, previous C-section, onset of labour, gestational age, and fetal presentation (e.g., breech is one presentation).[36] The Robson classification has some embedded complexities —the interaction of these 5 characteristics determine the grouping. Foreseeably, this classification is only used to standardize reporting of C-sections across risk groups rather than inform C-section decisions. The ambiguity and complexity of the environment instead leads to very simple decision aids allowing for little or no provider discretion, typically involving only one present underlying condition. For instance, two examples are "automatic" C-sections following a previous C-section birth or a breech presentation. With the development of more complicated algorithms for the allocation of care, our work suggests the environment of C-section is ripe for innovation.

While our paper focuses on county-level variation, it misses important within-area variation (Epstein and Nicholson, 2009). We abstract from the rich variation across individual hospitals and doctors highlighted in other studies (Epstein and Nicholson, 2009; Currie and MacLeod, 2017; Card, Fenizia and Silver, 2023). A natural next step is to understand whether the documented observations of (1) persistence in C-section usage over time and (2) the elevation of C-section rates across all risk propensities among higher C-section areas are replicated when the unit of observation is a hospital or a physician.

3.6 Tables & Figures

[36] https://www.who.int/publications/i/item/9789241513197

Table 3.1: Covariate Sets and Explanatory Power

	High-risk (vs. low-risk) births	*Predicting C-section need & adjusting county rates* All covariates	Excluding race	Excl. race + BMI	Medical only
			Racial disparities	*Appendix A.3.4*	*Appendix A.3.3*
Maternal age	<18 or >35	5-year bins	5-year bins	5-year bins	5-year bins
Gestational age	<37 weeks	<37 weeks	<37 weeks	<37 weeks	<37 weeks
Prenatal visits	≥ 19	≥ 19	≥ 19	≥ 19	≥ 19
Growth restrictions*	✓	✓	✓	✓	✓
Breech	✓	✓	✓	✓	✓
Eclampsia	✓	✓	✓	✓	✓
Pre-eclampsia	✓	✓	✓	✓	✓
Diabetes	✓	✓	✓	✓	✓
Maternal education	–	✓	✓	✓	–
Maternal birth country	–	✓	✓	✓	–
Birth month	–	✓	✓	✓	–
Birth day of week	–	✓	✓	✓	–
Sex of child	–	✓	✓	✓	–
Father's info. present	–	✓	✓	✓	–
Race-by-Hispanic origin	–	✓	–	–	–
Body mass index	–	–	–	cubic polynomial	–
Adj. R^2 1989-1991	–	0.1383	0.1375	–	0.1342
Adj. R^2 2015-2017	–	0.1286	0.1272	0.1483	0.1221

Notes: This table describes the sets of covariates used to define high- vs. low-risk births; to adjust county rates of c-section, maternal & neonatal mortality, and maternal & infant morbidity; and to predict C-section need for individuals. For each model, the adjusted R^2 is from a regression of C-section on the listed covariates and county fixed effects among singleton first births. See Table A.9 for comparison with select literature. *Growth restrictions are defined as below the 5^{th} percentile of birthweight for gestational age. These cutoffs come from `https://srhr.org/fetalgrowthcalculator`.

Table 3.2: Distribution of Births Across Low- and High-Riskby Predicted C-Section Need

1989-1991

	Predicted C-section need (using all covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.33 %	0.00 %	60.74 %
High-risk births	33.27 %	100.00 %	35.99 %
Unknown risk births	3.40 %	0.00 %	3.27 %
	100%	100%	100%

2015-2017

	Predicted C-section need (using all covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.68 %	0.00 %	60.41 %
High-risk births	34.01 %	99.99 %	37.40 %
Unknown risk births	2.31 %	0.01 %	2.19 %
	100%	100%	100%

Notes: This table shows the relationship between two different approaches for assessing risk of C-section: (1) the categorization of births as low- or high-risk, a binary assignment based on medical factors only, and (2) the predicted C-section need, a continuum of risk estimated using all covariates. See Table 3.1 for the specific covariates used in each model. For this table, the cutoff of ± 0.6 was chosen due to the bimodal distribution exhibited by the predicted C-section need (see Figure A.23). Only singleton first births are represented. A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either low- or high-risk.

Table 3.3: Descriptive Statistics

	1989-1991				2015-2017			
	mean	sd	p25	p75	mean	sd	p25	p75
# Counties	2,346				2,346			
Population	105,260	307,227	18,549	77,921	135,564	377,944	21,115	102,119
Births	1,739	6,078	265	1,169	1,651	4,823	237	1,177
C-section rate	0.235	0.052	0.201	0.266	0.319	0.057	0.281	0.354
Fraction of births that are singleton 1st	0.327	0.046	0.298	0.355	0.294	0.038	0.272	0.316
C-section rate for singleton 1st births	0.243	0.063	0.202	0.280	0.283	0.065	0.242	0.319
Fraction of births that are low-risk singleton 1st	0.195	0.037	0.172	0.219	0.177	0.031	0.159	0.196
C-section rate for low-risk singleton 1st births	0.208	0.071	0.162	0.247	0.226	0.074	0.179	0.266
Fraction of births that are high-risk singleton 1st	0.125	0.032	0.102	0.144	0.111	0.025	0.096	0.125
C-section rate for high-risk singleton 1st births	0.303	0.087	0.247	0.353	0.376	0.093	0.324	0.429
Neonatal mortality rate per 1,000 births	5.488	4.269	2.865	7.375	4.097	3.763	1.406	5.780
Infant morbidity rate	0.079	0.045	0.052	0.097	0.048	0.031	0.027	0.062
Maternal morbidity rate	0.016	0.013	0.008	0.021	0.014	0.011	0.007	0.018

Notes: This table includes summary statistics for our sample of counties. We use a balanced panel and exclude counties that ever have fewer than 100 births. A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either high- or low- risk.

Table 3.4: Characteristics of Low, Medium, and High C-Section Areas (1/2): Statistics Compiled from Natality Data

	1989-1991			2015-2017		
	Low CS	Medium	High CS	Low CS	Medium	High CS
# Counties	666	654	1,026	668	537	1,141
# Singleton 1st births per year	625	682	421	570	753	345
C-section rate (adjusted for all covariates)	0.192	0.232	0.284	0.238	0.283	0.336
Predicted C-section need	0.240	0.237	0.231	0.291	0.287	0.279
Maternal age	24.3	23.9	23.3	26.7	26.4	25.5
Share maternal age <18 or > 35	0.134	0.139	0.150	0.103	0.103	0.101
Share gestational age <37 weeks	0.092	0.097	0.100	0.087	0.089	0.102
Share w/ any prenatal visits	0.985	0.984	0.984	0.987	0.987	0.982
Share prenatal visits ≥ 19	0.032	0.032	0.034	0.033	0.033	0.030
Share w/ growth restrictions	0.080	0.080	0.080	0.084	0.083	0.087
Share breech	0.041	0.040	0.039	0.053	0.048	0.044
Share w/ eclampsia	0.006	0.006	0.007	0.004	0.003	0.003
Share w/ pre-eclampsia	0.042	0.040	0.041	0.076	0.072	0.076
Share w/ diabetes	0.019	0.018	0.017	0.056	0.054	0.054
Share non-Hispanic white	0.681	0.668	0.611	0.573	0.517	0.555
Share non-Hispanic Black	0.133	0.154	0.141	0.119	0.137	0.150
Share Hispanic	0.128	0.140	0.210	0.197	0.247	0.231
Share occurring in the county of residence	0.781	0.768	0.761	0.760	0.742	0.660
Share Medicaid	.	.	.	0.319	0.382	0.425

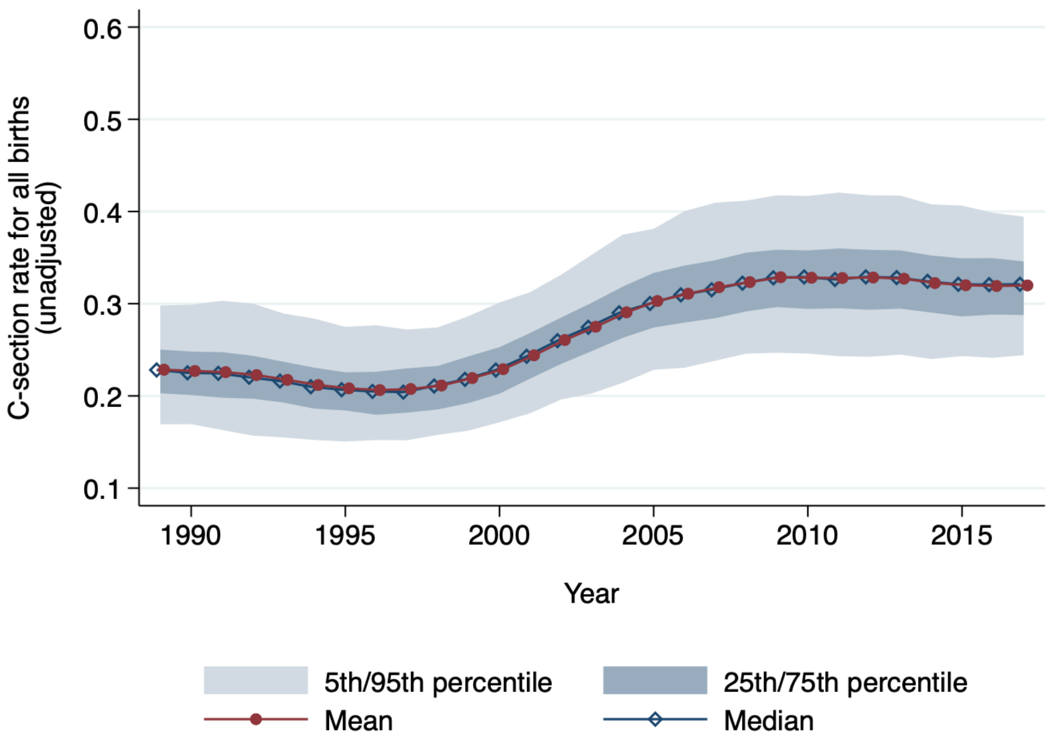
Notes: This table shows the average county characteristics for counties with low, medium, and high C-section rates. Counties are separated using the C-section rate for singleton first births (adjusted for all covariates), into terciles weighted by the number of singleton first births.

Table 3.5: Characteristics of Low, Medium, and High C-Section Areas (2/2): Statistics Compiled from the Area Health Resource File and National Practitioner Data Bank

	1989-1991			2015-2017		
	Low CS	Medium	High CS	Low CS	Medium	High CS
Income per capita (2020 $)	38,709	37,978	34,415	59,285	54,206	49,532
Unemployment rate	0.055	0.058	0.067	0.046	0.049	0.052
Poverty rate	0.122	0.122	0.152	0.132	0.144	0.153
Share urban	0.818	0.797	0.734	0.861	0.879	0.773
Share w/ at least HS diploma	0.784	0.756	0.712	0.888	0.861	0.858
Share w/ 4+ years college	0.231	0.206	0.181	0.351	0.313	0.276
Hospitals per 1,000 births	1.489	1.447	1.699	1.395	1.281	1.705
Hospitals w/ med. school affiliation per 1,000 births	0.365	0.288	0.215	0.486	0.397	0.361
Hospital beds per 1,000 births	303	297	289	238	231	241
Newborn bassinets per 1,000 births	16.99	16.50	17.15	14.55	13.86	14.60
MDs per 1,000 births	153	130	106	292	231	192
OB/GYNs per 1,000 births	7.989	7.356	6.055	12.034	9.927	8.861
Medical residents per 1,000 births	0.547	0.444	0.467	0.013	0.020	0.015
Surgeons per 1,000 births	36.89	32.72	27.18	53.17	43.51	37.44
Operating rooms per 1,000 births	.	.	.	9.91	8.63	8.59
Malpractice # payments per 1,000 MDs	25.77	25.69	25.30	8.02	9.55	10.51
Obstetrics malpractice # payments per 1,000 OB/GYNs	44.63	43.15	43.54	11.43	14.04	14.11
Malpractice liability per MD (2020 $)	7,436	7,816	7,109	3,527	3,843	3,975
Obstetrics malpractice liability per OB/GYN (2020 $)	21,085	20,625	16,870	9,046	9,423	9,147

Notes: This table shows the average county characteristics for counties with low, medium, and high C-section rates. Counties are separated using the C-section rate for singleton first births (adjusted for all covariates), into terciles weighted by the number of singleton first births.

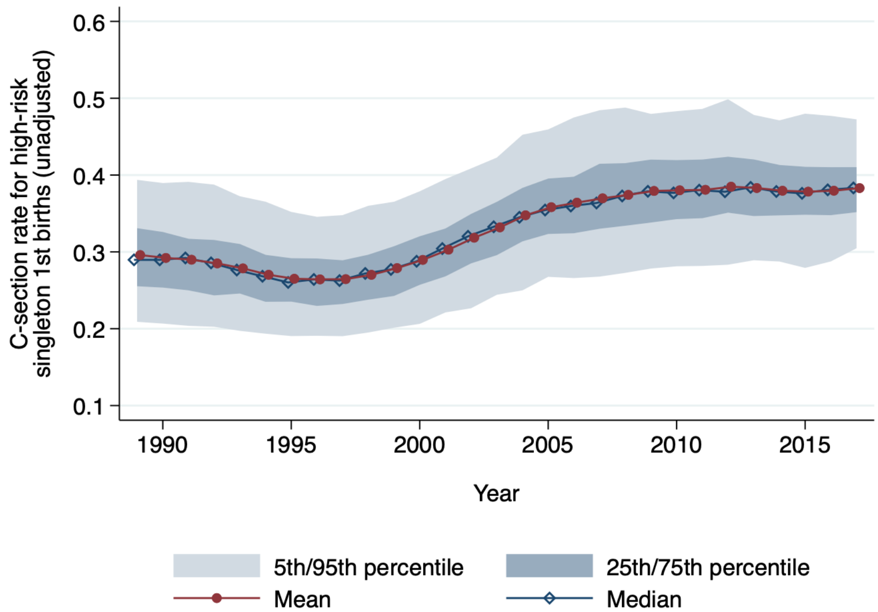
Figure 3.1: Time Trends in County C-Section Rates



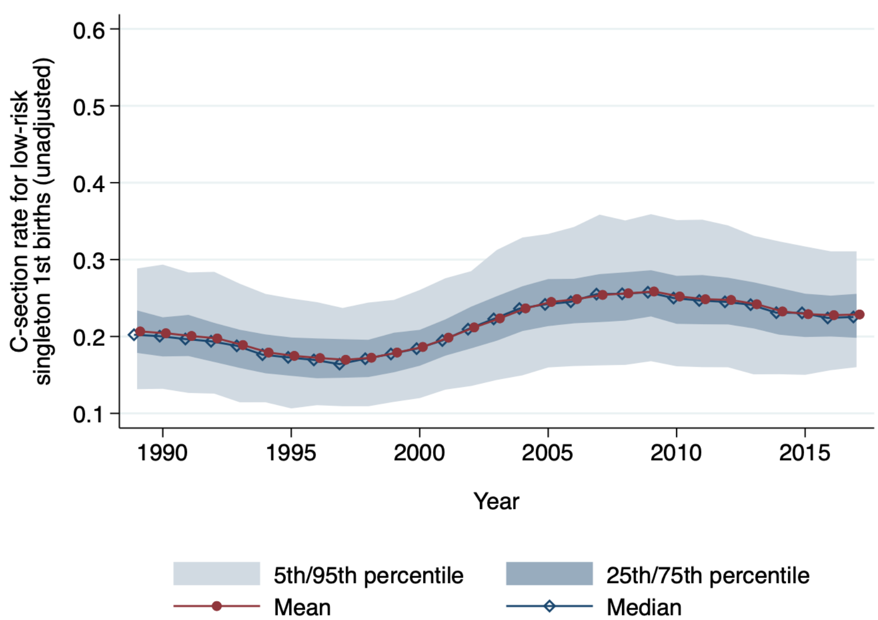
Notes: This figure shows the distribution of (raw) county C-section rates over time for all births (including higher order births). All statistics are weighted across counties by the number of births. The mean is the overall C-section rate and 50 percent of births occur in counties with C-section rates in the interquartile range.

Figure 3.2: Time Trends in County C-Section Rates by Risk Group

High-Risk Singleton 1st Births



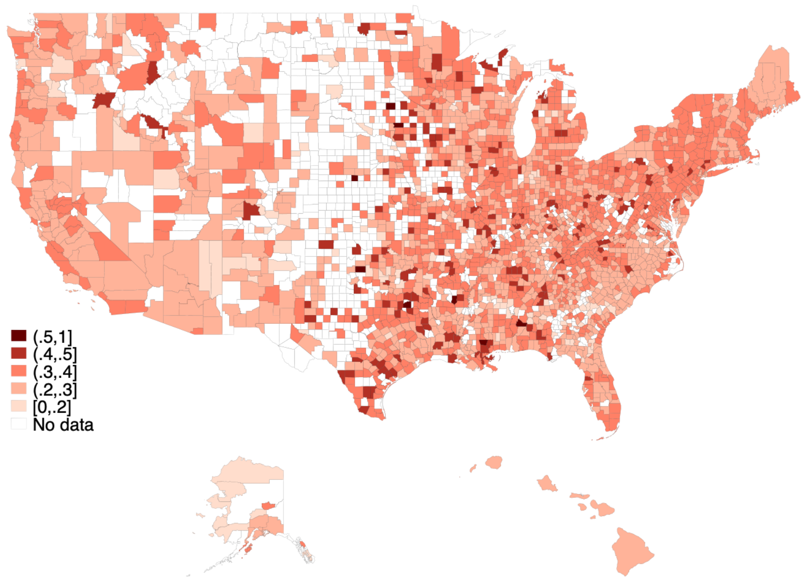
Low-Risk Singleton 1st Births



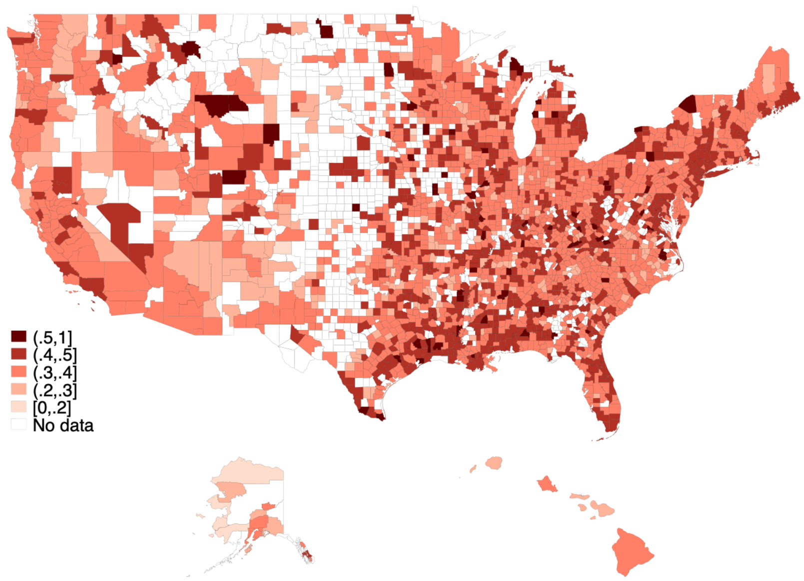
Notes: This figure shows the distribution of (raw) county C-section rates over time for low- and high-risk singleton first births. All statistics are weighted across counties by the number of relevant births.

Figure 3.3: Time Trends in County C-Section Rates Across U.S. for High-Risk Singleton 1st Births

1989-1991

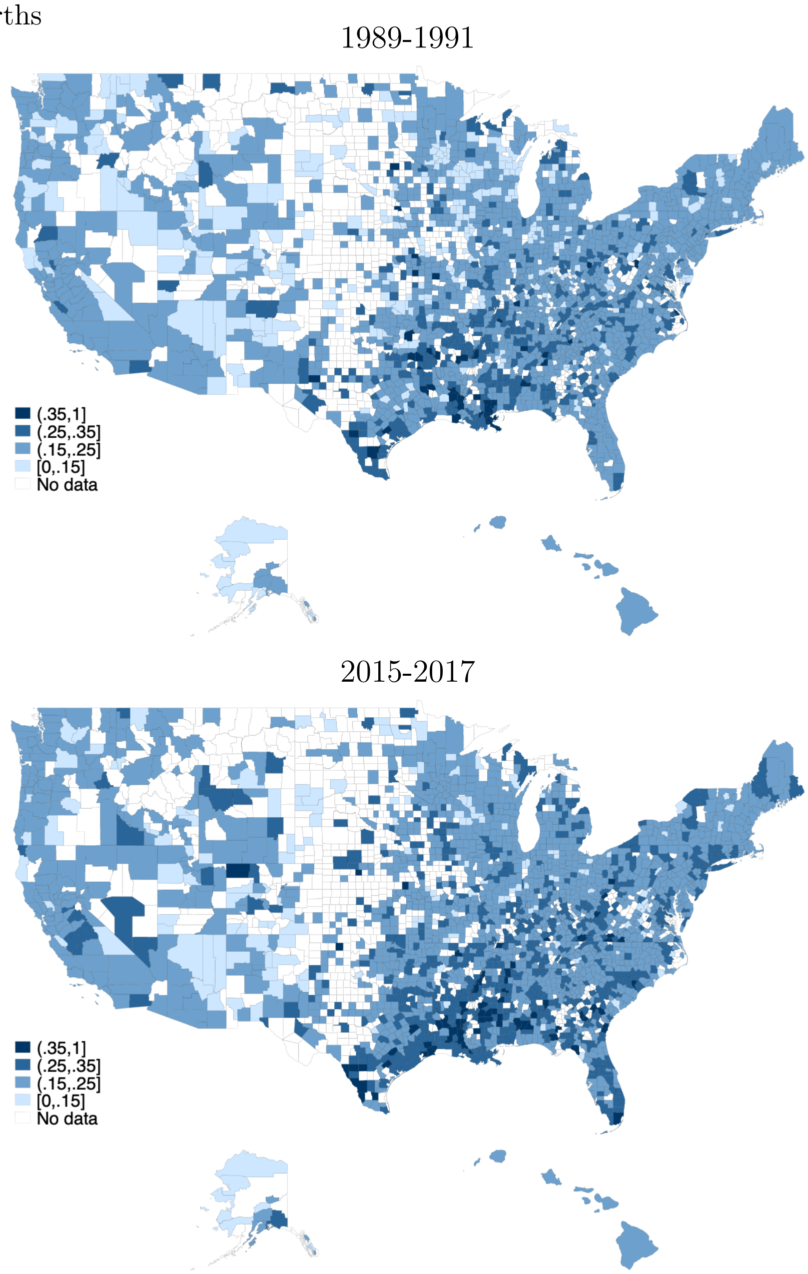


2015-2017



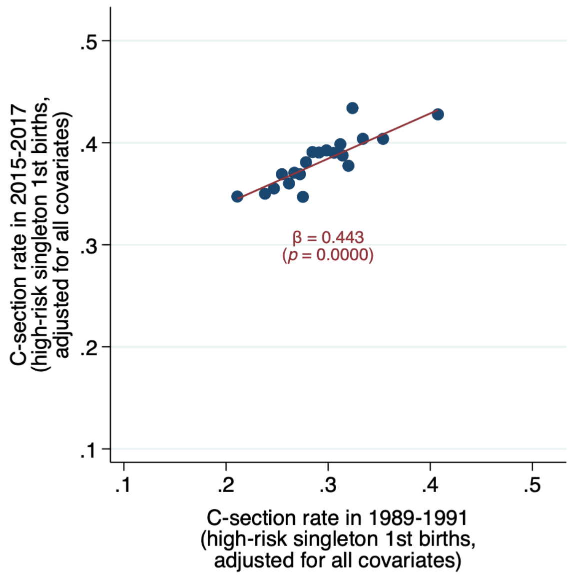
Notes: This figure shows the geographic distribution of (raw) C-section rates for high-risk singleton first births. Rates are the average within a county over the three-year period, weighted by the number of high-risk singleton first births in each year.

Figure 3.4: Time Trends in County C-Section Rates Across U.S. for Low-Risk Singleton 1st Births

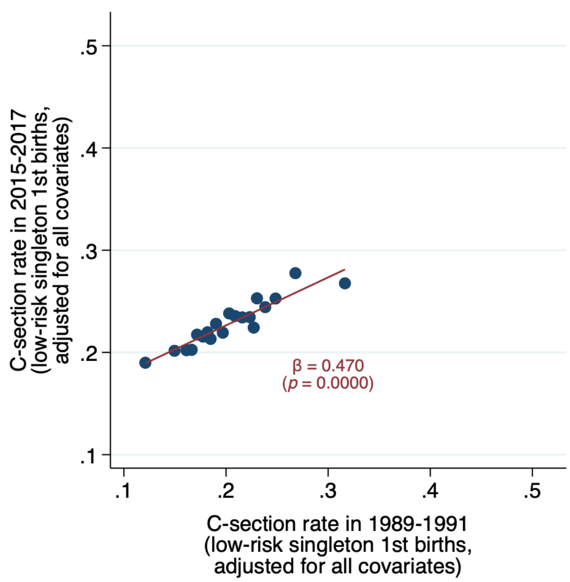


Notes: This figure shows the geographic distribution of (raw) C-section rates for low-risk singleton first births. Rates are the average within a county over the three-year period, weighted by the number of low-risk singleton first births in each year.

Figure 3.5: Persistence in Adjusted County C-Section Rates Over Time
High-Risk Singleton 1st Births



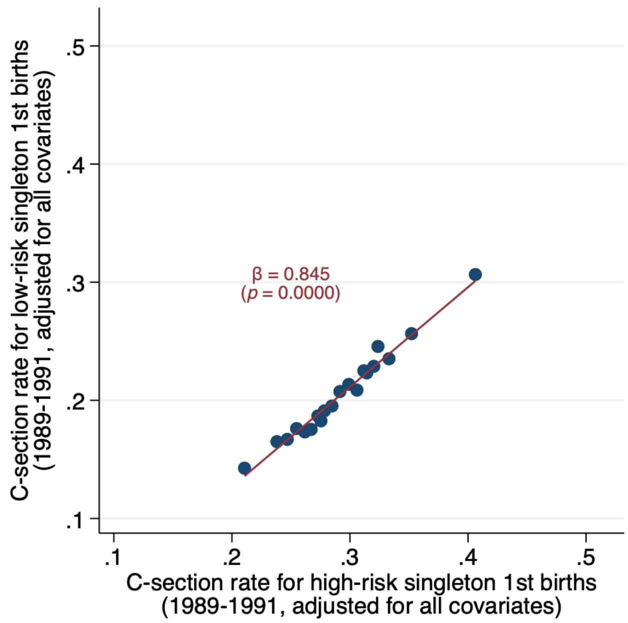
Low-Risk Singleton 1st Births



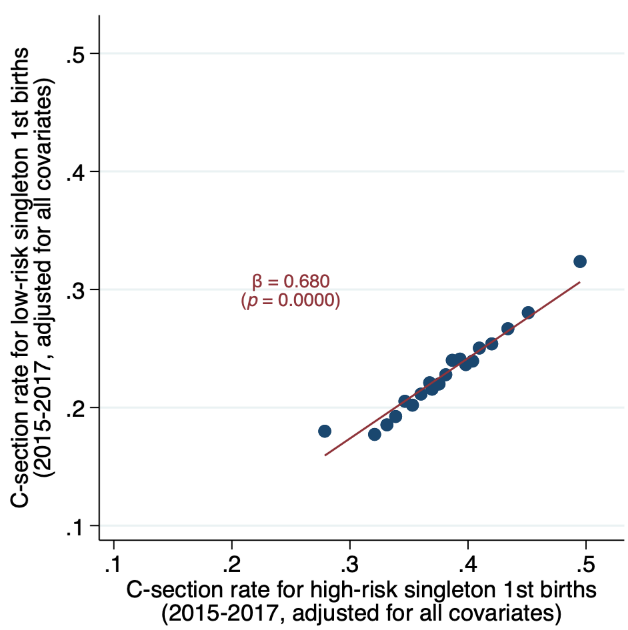
Notes: This figure shows binscatter plots of C-section rates across time periods. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over all six years.

Figure 3.6: Correlation in Adjusted County C-Section Rates Across Risk Type

1989-1991

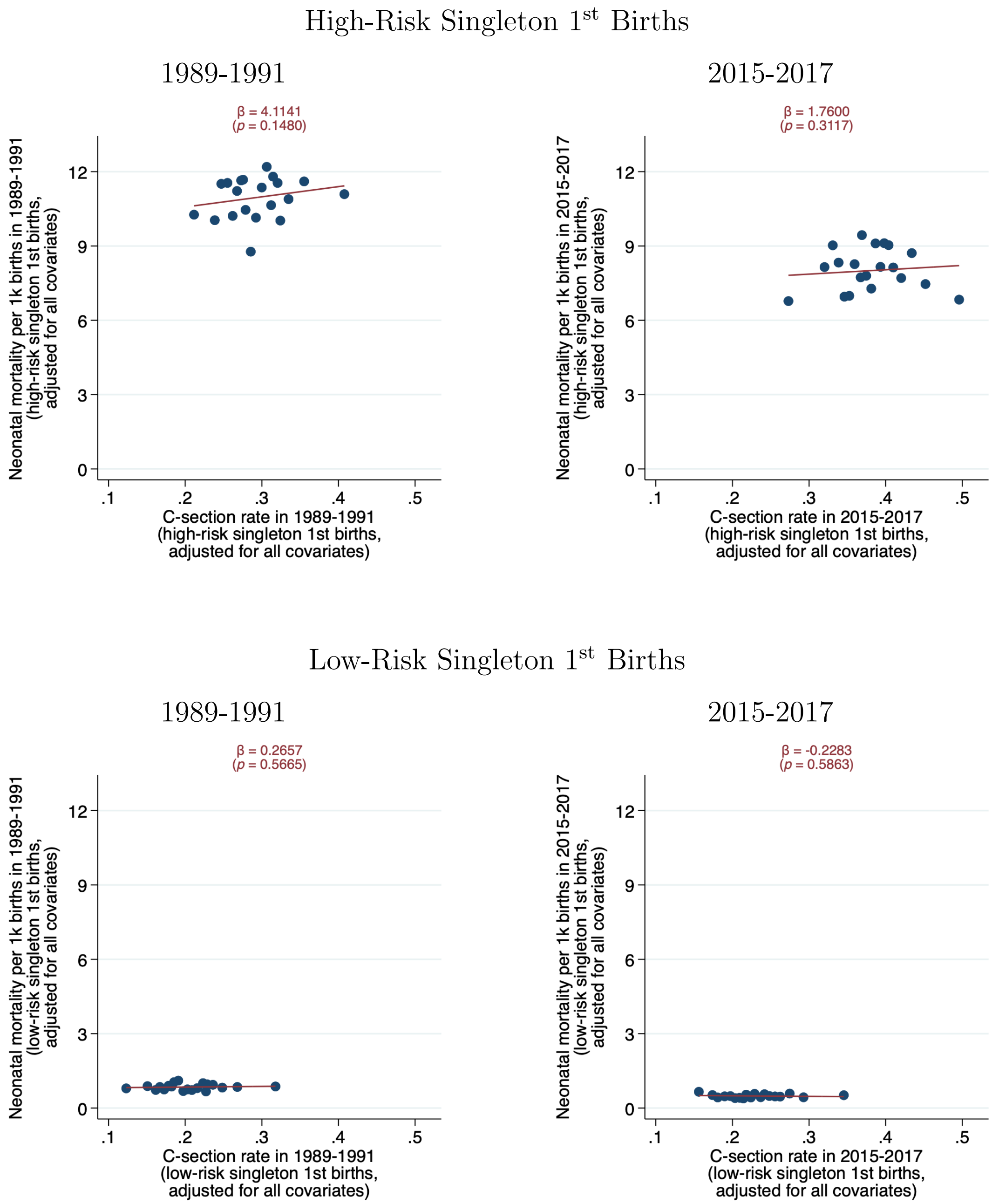


2015-2017



Notes: This figure shows binscatter plots of C-section rates across risk types. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

Figure 3.7: Correlation of Adjusted County C-Section Rates and Neonatal Mortality

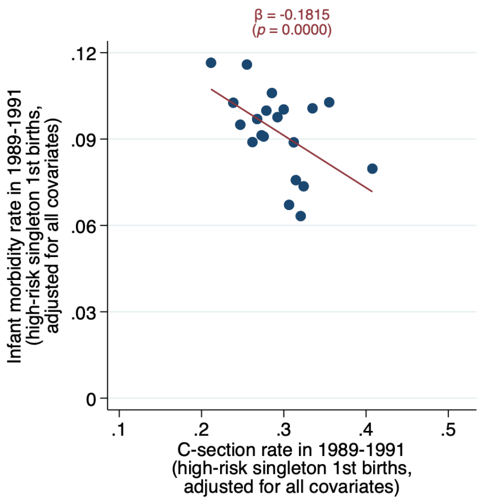


Notes: This figure shows binscatter plots of C-section rates with neonatal mortality rates. Linear fit and p-value are based on the underlying counties (prior to binning). C-section and neonatal mortality rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over the three years.

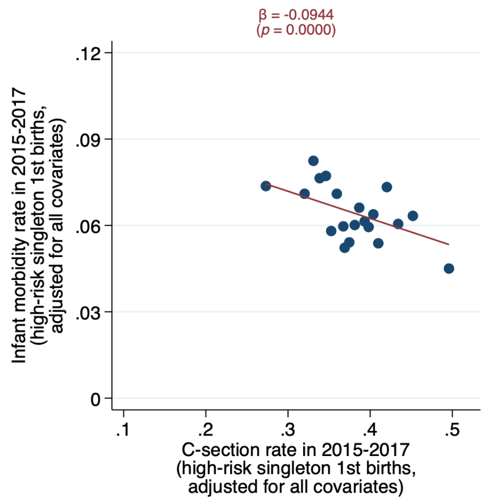
Figure 3.8: Correlation of Adjusted County C-Section Rates and Infant Morbidity

High-Risk Singleton 1st Births

1989-1991

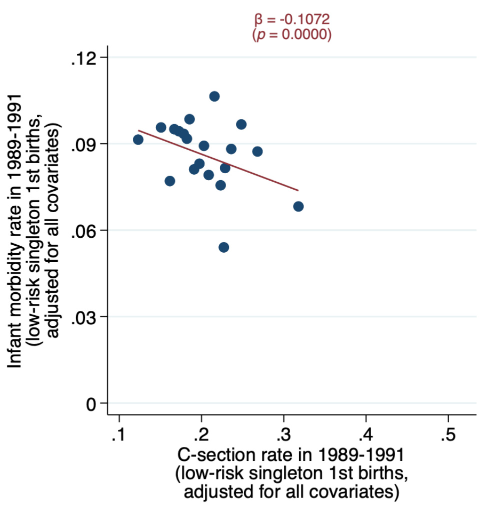


2015-2017

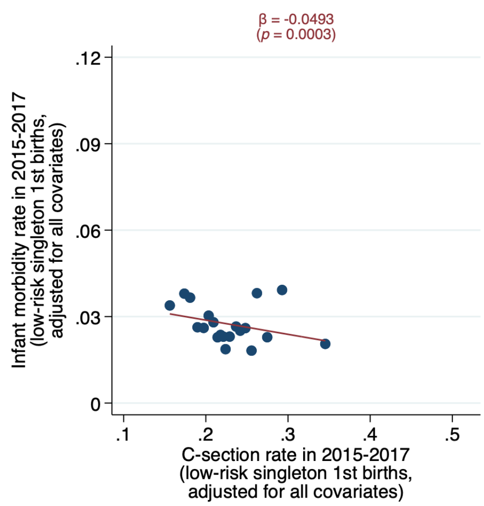


Low-Risk Singleton 1st Births

1989-1991



2015-2017

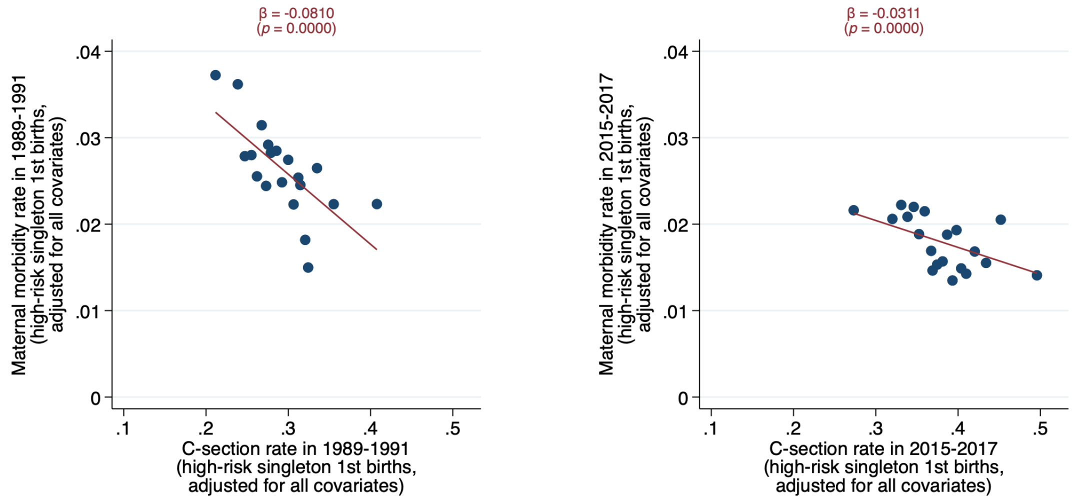


Notes: This figure shows binscatter plots of C-section rates for singleton first births with infant morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. C-section and morbidity rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of singleton first births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

Figure 3.9: Correlation of Adjusted County C-Section Rates and Maternal Morbidity

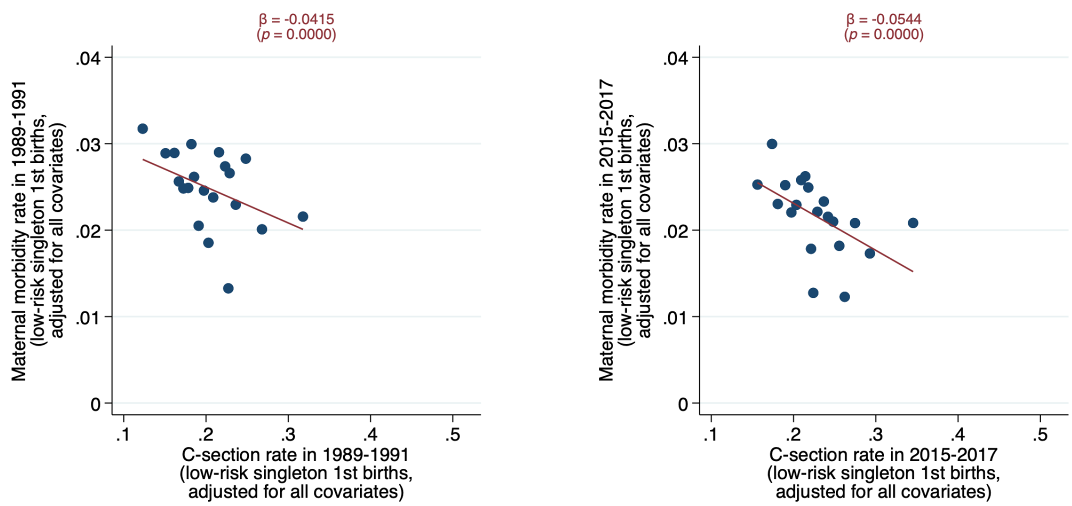
High-Risk Singleton 1^{st} Births

1989-1991 2015-2017



Low-Risk Singleton 1^{st} Births

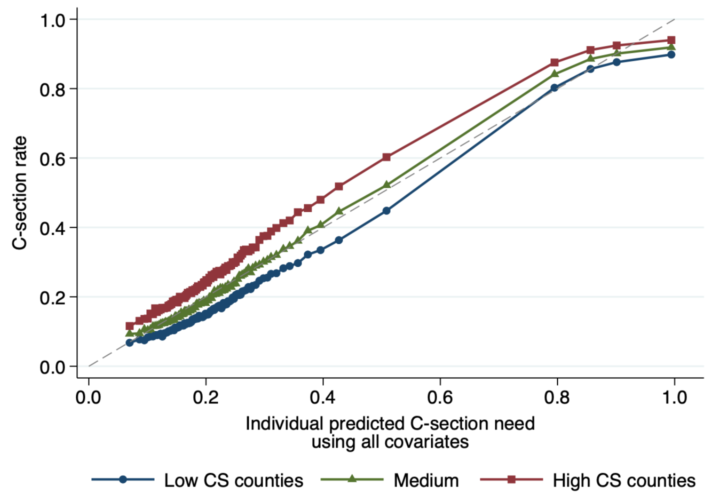
1989-1991 2015-2017



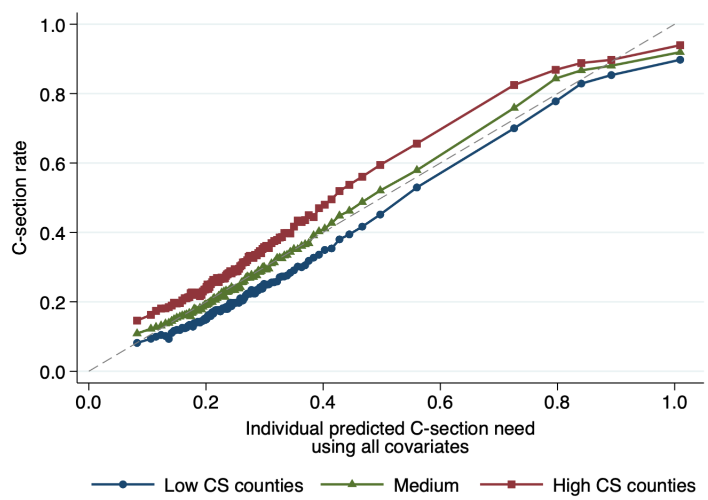
Notes: This figure shows binscatter plots of C-section rates for singleton first births with maternal morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. C-section and morbidity rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of high-risk singleton first births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

Figure 3.10: C-Section Rates by Predicted C-Section Needand County C-Section Rate

1989-1991



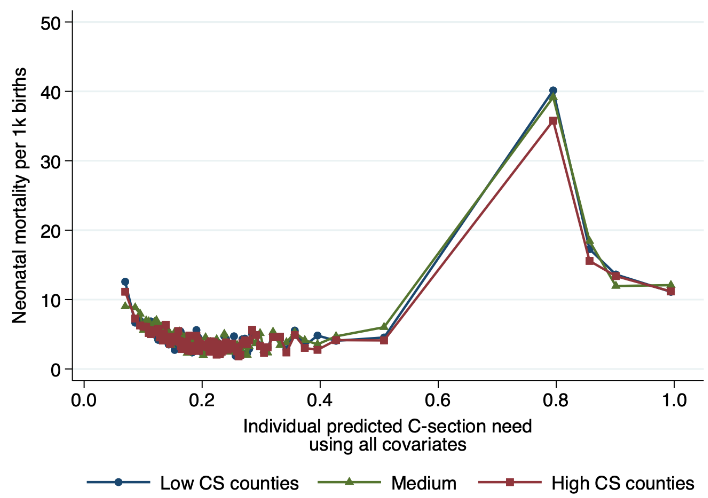
2015-2017



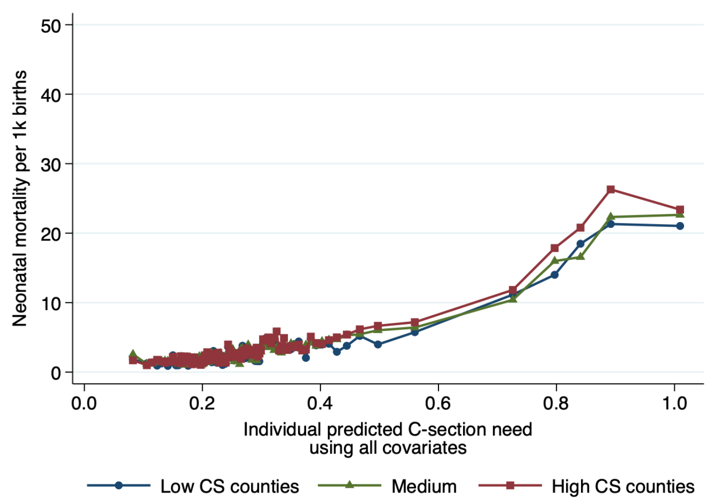
Notes: This figure shows (raw) C-section rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section need is derived from a regression based on all covariates and county fixed effects, but the prediction excludes the county fixed effects. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Figure 3.11: Neonatal Mortality by Predicted C-Section Needand County C-Section Rate

1989-1991

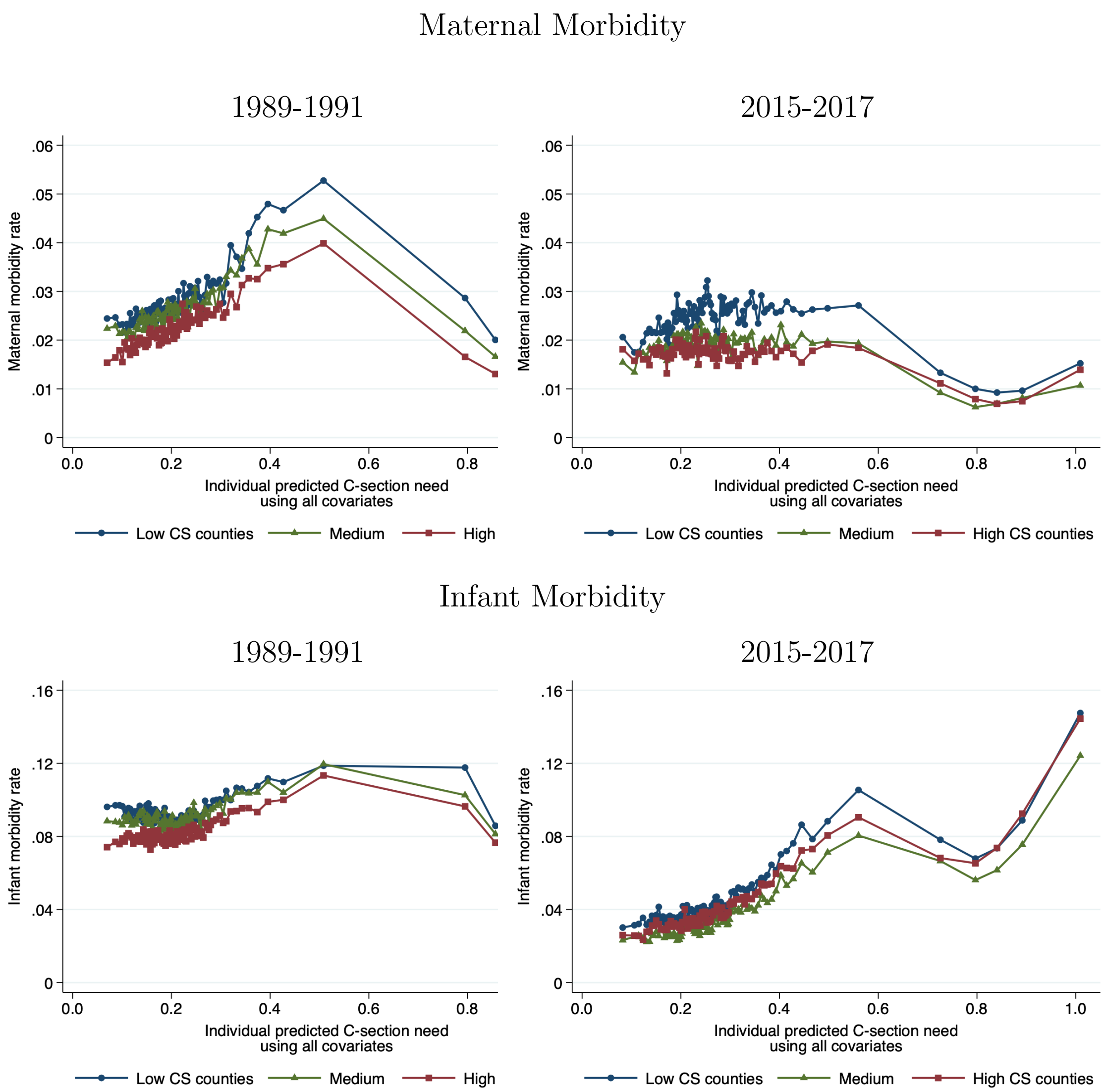


2015-2017



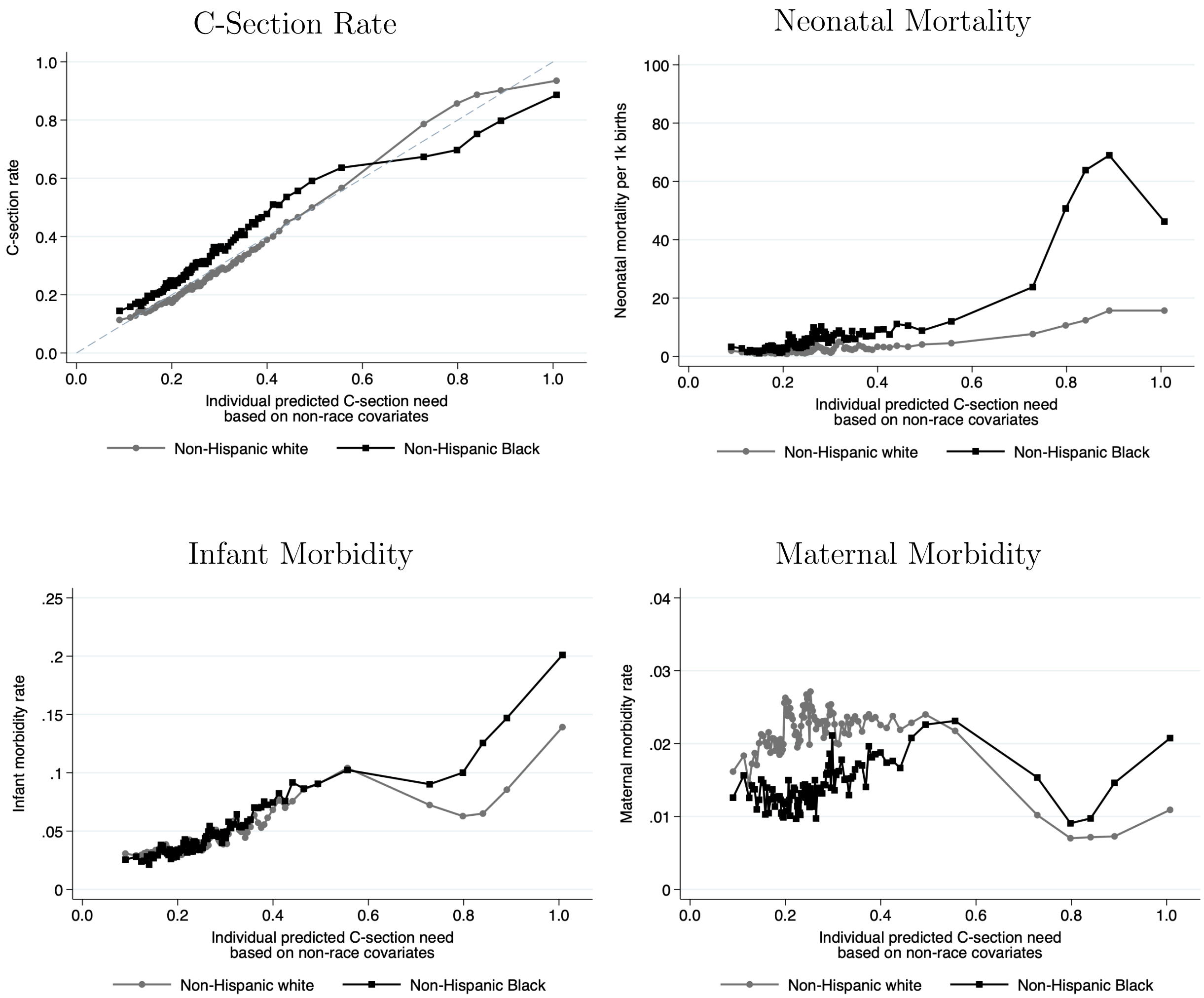
Notes: This figure shows (raw) neonatal mortality rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section need is derived from a regression based on all covariates and county fixed effects, but the prediction excludes the county fixed effects. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Figure 3.12: Maternal and Infant Morbidity by Predicted C-Section Need and County C-Section Rate



Notes: This figure shows (raw) maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section need is derived from a regression based on all covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Figure 3.13: C-Section Rates, Neonatal Mortality, and Maternal and Infant Morbidity by Predicted C-Section Need and Race, 2015-2017
(Based on Non-Race Covariates)



Notes: This figure shows (raw) c-section, neonatal mortality, and maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births with non-Hispanic white mothers and non-Hispanic Black mothers. Predicted C-section need is derived from a regression based on non-race covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Appendix A

Additional Information

A.1 Do Firms Avoid Health Insurance Mandates? Evidence from the Self-Funding of Employer Plans

Table A.1: Mandates Contributing to Treatment

Mandate	# Times Passed	# Times Passed in Treatment Year	# Times Passed Alone in Treatment Year
Acupuncturists	5	1	1
Alcoholism Treatment	2		
Ambulatory Surgery	1		
Chiropractors	3	3	
Contraceptives	26	13	7
Conversion to Non-Group	4	2	
Dentists	2	2	
Diabetic Supplies & Education	19	14	6
Handicapped Dependents	2	1	
Home Health Care	1		
Infertility Treatment	5		
Maternity	6	3	2
Mental Health (General)	9	5	2
Mental Health (Parity)	29	13	7
Optometrists	7	4	
Osteopaths	3	3	
Physical Therapists	3	1	
Psychologists	1	1	
Rehabilitation Services	3		
Social Workers	1	1	
Well Child Care	16	7	6

Notes: This table lists the (costly) mandates that occur during the time period of study. The first column details the number of times the mandate was passed overall. The second column details the number of times that the mandate was passed in the treatment year. The third column details the number of states where the mandate was the only mandate passed in the treatment year.

Table A.2: Effect of Mandates on Offering Health Coverage

	(1)	(2)
Mandate	-0.002	
	(0.007)	
Mandate * Small		-0.002
		(0.008)
Mandate * Large		-0.002
		(0.006)
N	185,261	185,261
State FE	Yes	Yes
Year FE	Yes	Yes
Control for Small		Yes
Mean	0.865	
Mean (100-249)		0.905
Mean (250+)		0.840

Notes: This table shows the estimated effect of mandated benefits on whether firms offer any health coverage. The sample includes all firms that report offering any welfare benefits through the Form 5500, for event times -5 to 3 (where 0 is the year of treatment). Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Effect of Mandates on Self Funding

	(1)	(2)
Mandate	0.009	
	(0.006)	
Mandate * Small		0.032**
		(0.013)
Mandate * Large		-0.006
		(0.006)
N	159,714	159,714
State FE	Yes	Yes
Year FE	Yes	Yes
Control for Small		Yes
Mean	0.274	
Mean (100-249)		0.239
Mean (250+)		0.298

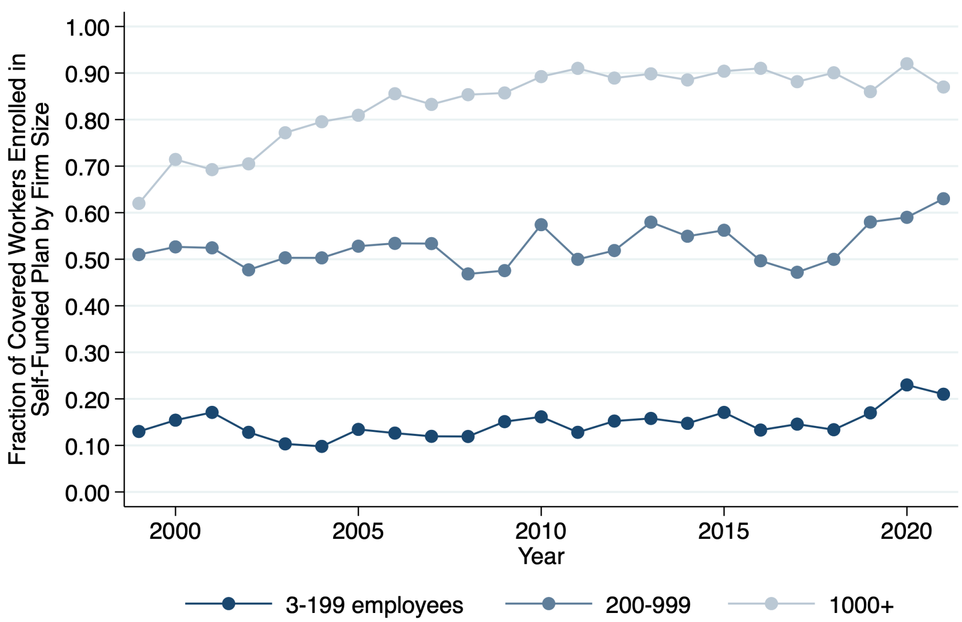
Notes: This table shows the estimated effect of mandated benefits on whether firms self-fund their health coverage. The sample includes all firms that report offering health coverage through the Form 5500, for event times -5 to 3 (where 0 is the year of treatment). Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Varying Size of Event Window

	(1)	(2)	(3)
Mandate * Small	0.032**	0.031**	0.031**
	(0.013)	(0.013)	(0.013)
Parallel trends test p-value	0.016	0.541	0.643
Mandate * Large	-0.006	-0.008	-0.009
	(0.006)	(0.005)	(0.007)
Parallel trends test p-value	0.006	0.005	0.084
N	159,714	152,986	144,384
State FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Control for Small	Yes	Yes	Yes
Window	-5 to 3	-4 to 2	-3 to 1

Notes: This table shows the estimated effect of mandated benefits on whether firms self-fund their health coverage. Each column considers a different event window, where 0 is the year of treatment. Results from the difference-in-differences specification are shown, as well as p-values from an F-test that all of the event study pre-period coefficients are equal to zero. The sample includes all firms that report offering health coverage through the Form 5500. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

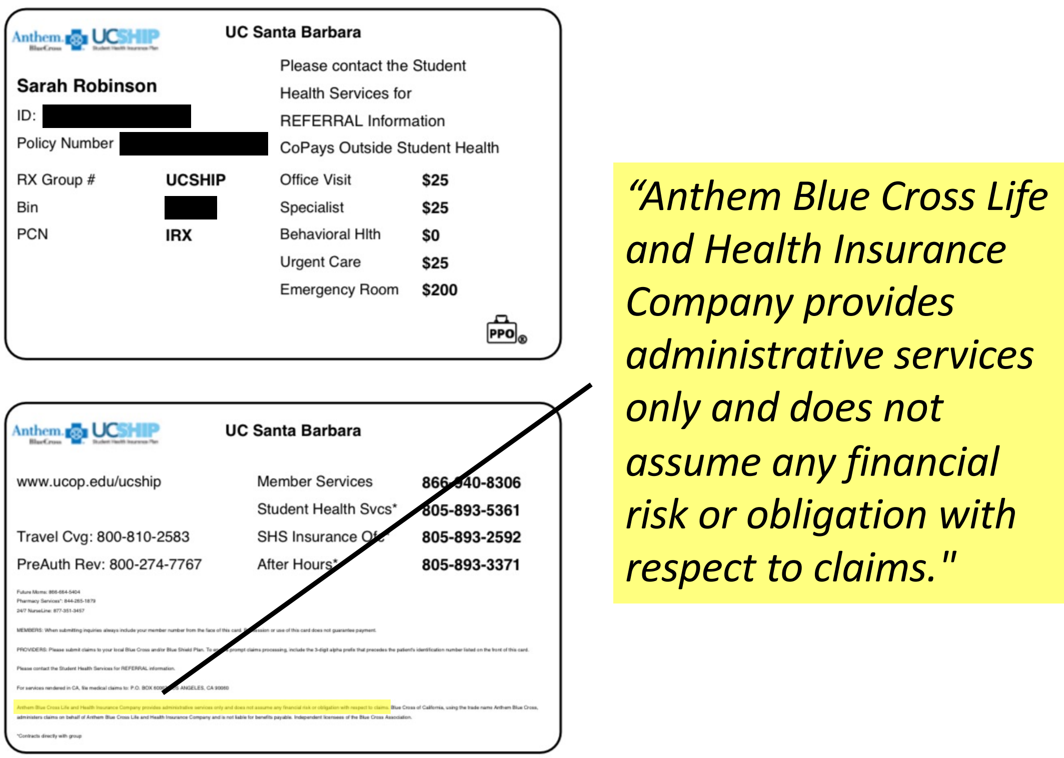
Figure A.1: Prevalence of Self-Funding Over Time by Firm Size



Notes: This figure shows the percent of workers who are enrolled in self-funded plans, among all workers who are covered by an employer-sponsored health plan, by the size of firm.

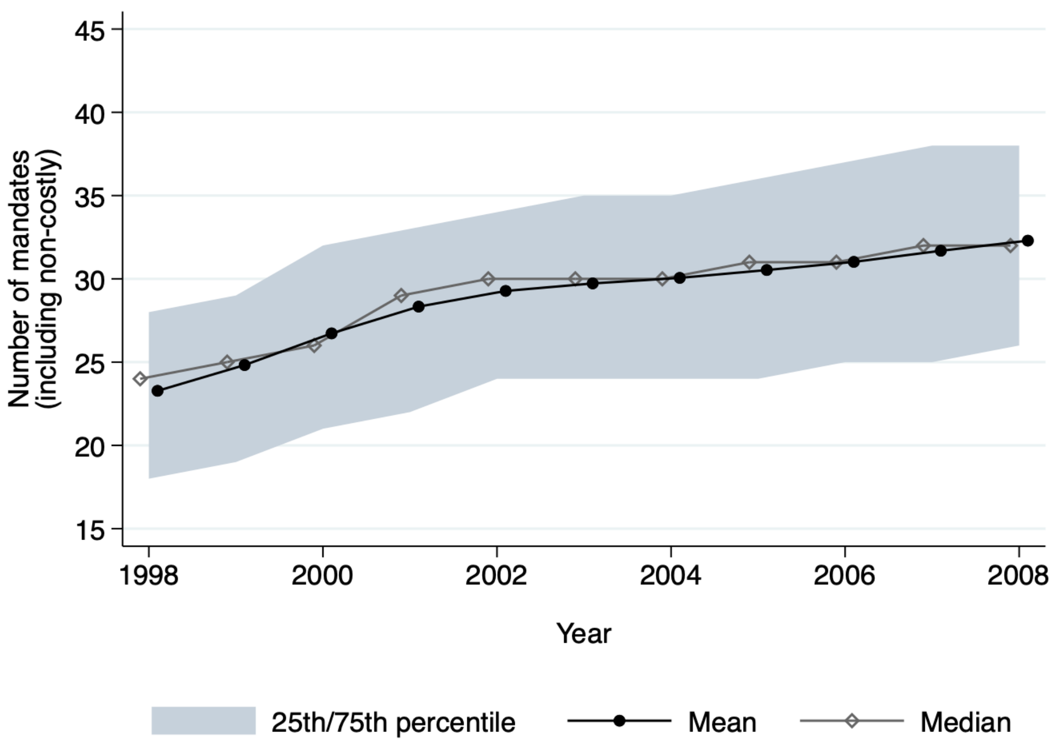
Source: Kaiser Family Foundation Employer Health Benefits Survey

Figure A.2: Sample Health Insurance Card



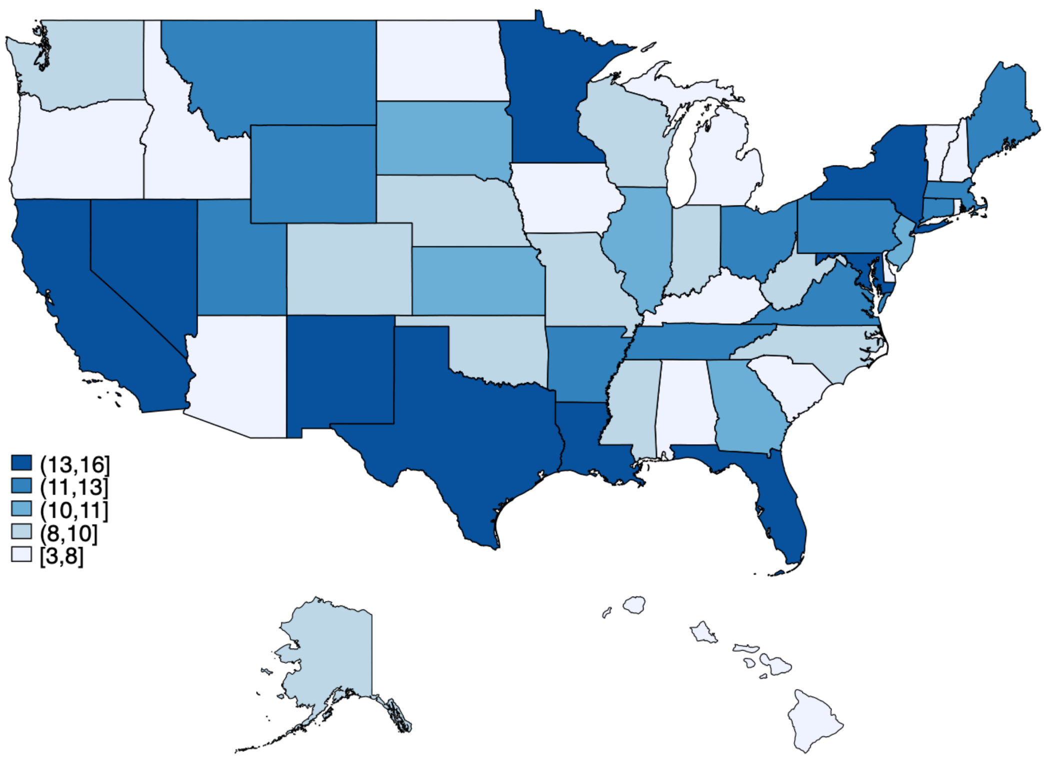
Notes: This figure shows an example of an insurance card. The highlight and expanded text is the indication that this insurance plan is self-funded.

Figure A.3: Mandates Over Time (Including Non-Costly Mandates)



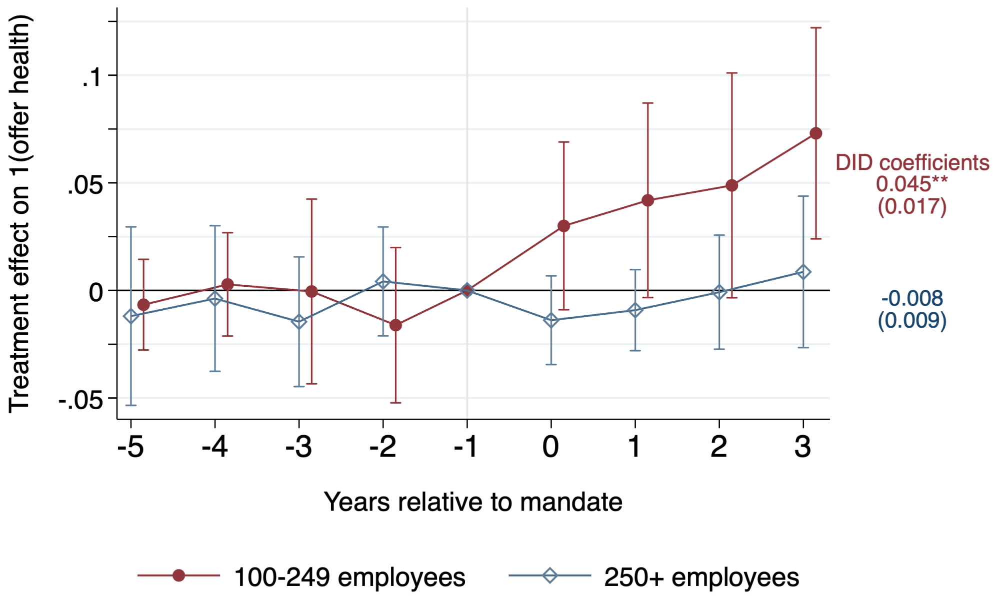
Notes: This figure shows the distribution of mandates across states over time. In each year, the mean number of mandates across states is shown. The median, 25th percentile, and 75th percentile are also shown. Mandates with negligible effects on costs are included.

Figure A.4: Number of Mandates by State in 1998



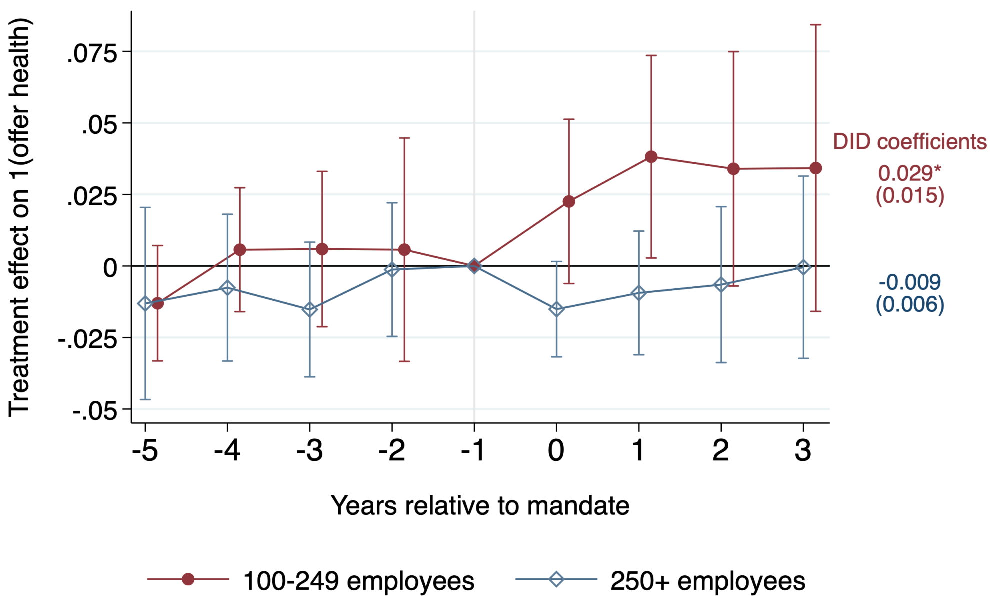
Notes: This figure shows the number of mandates in each state in the baseline year of 1998.

Figure A.5: Excluding States with AdditionalMandates After Treatment Year



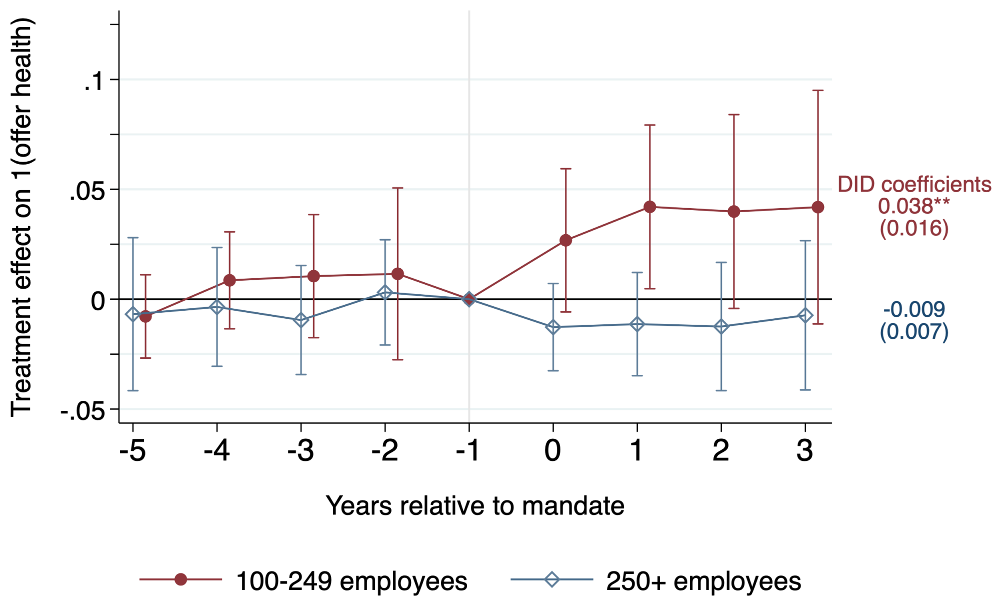
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, where states that passed additional mandates after the treatment year are excluded. The sample includes firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.6: Excluding States with More than OneMandate in Treatment Year



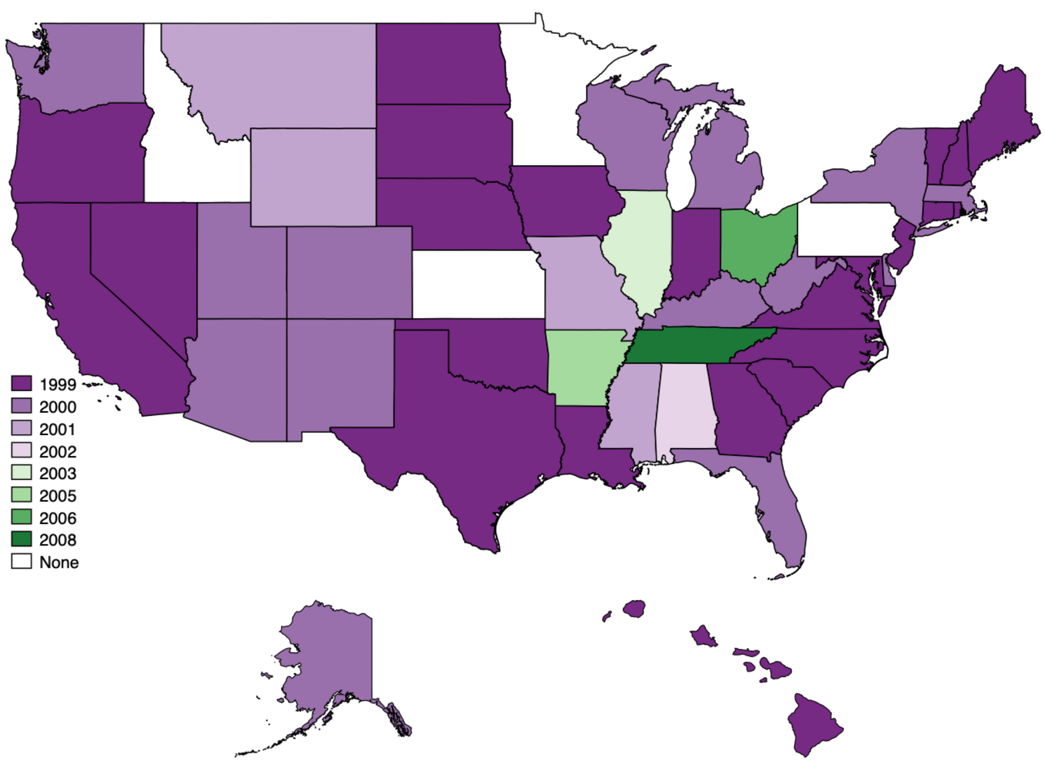
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, where states that passed more than one mandate in the treatment year are excluded. The sample includes firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.7: Excluding States with New Mandatesin Four Years Prior to Treatment



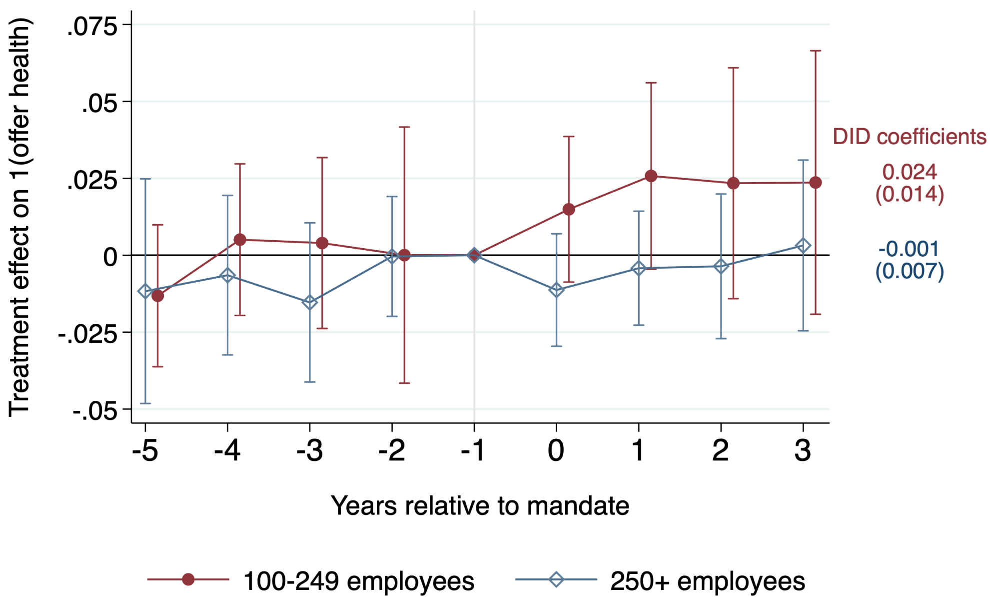
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, where states that passed mandates in the previous four years are excluded. The sample includes firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.8: Year of Treatment by State



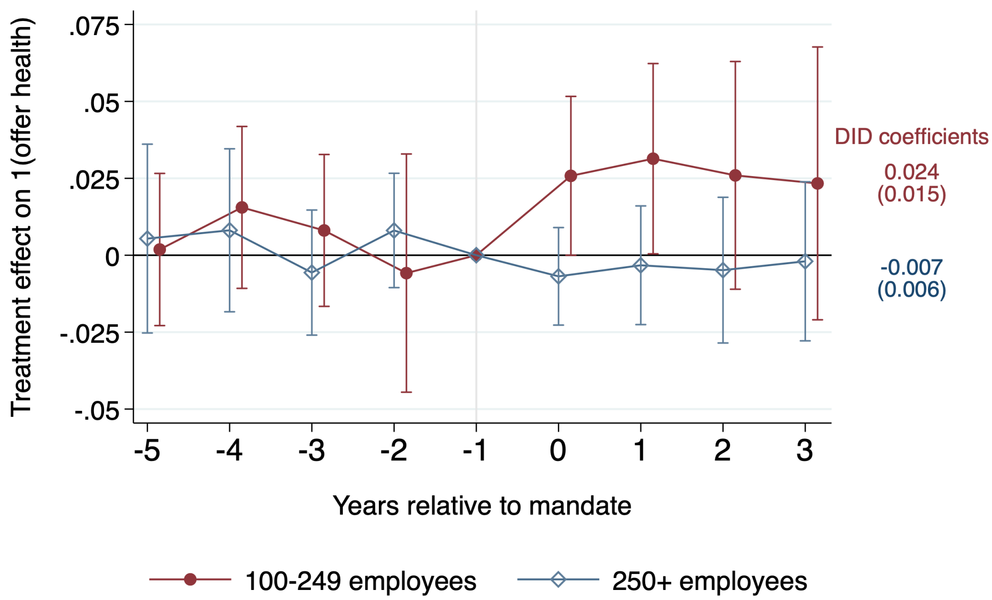
Notes: This figure shows the year of treatment for each state (the first year that state passes any costly mandate).

Figure A.9: Excluding States with Mandates in 1999



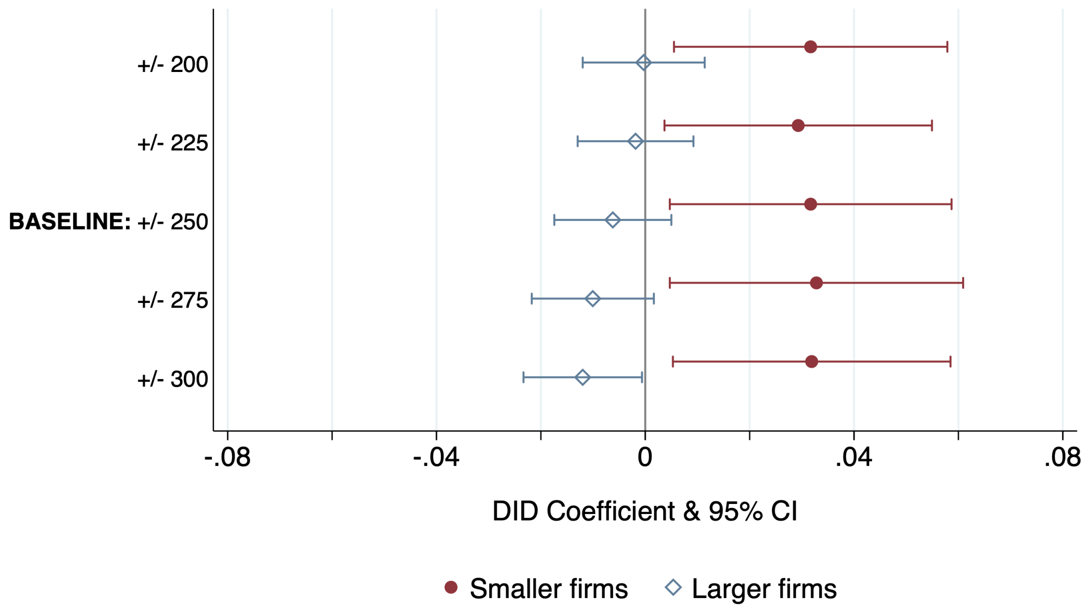
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, where states that passed mandates in 1999 are excluded. The sample includes all firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.10: Identifying Costly Mandates with CAHI Only



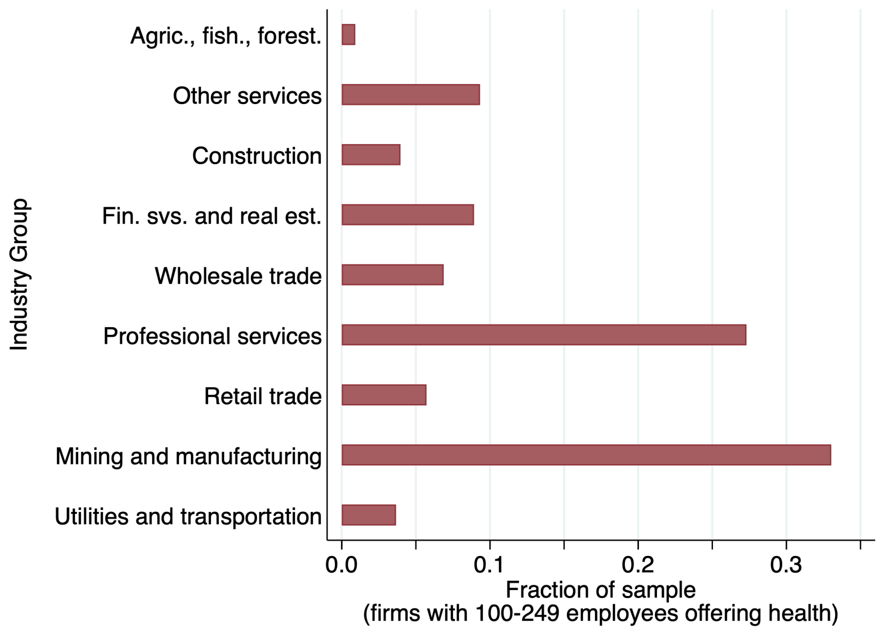
Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage, using only the 19 costly mandates identified by CAHI. The sample includes firms that report offering health coverage through the Form 5500. Event study and difference-in-differences estimates are from a regression that interacts treatment with firm size category, controls for size category, and includes state and year fixed effects. Standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.11: Varying Cutoff Between Small and Large Firms



Notes: This figure shows the estimated effect of mandated benefits on whether firms self-fund their health coverage. For each specification, difference-in-differences estimates and 95% confidence intervals are shown for smaller and larger firms. All estimates are from a regression that interacts treatment with firm size category and controls for size category, and standard errors are clustered at the state level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A.12: Distribution of Firms (100-249 Employees) Across Industry Groups



Notes: This figure shows how the sample of firms with 100-249 employees that offer health is distributed across industry groups.

A.2 What Drives Tax Policy? Political, Institutional and Economic Determinants of State Tax Policy

A.2.1 Data Notes

We rely on the previous literature, summarized in Table A.5 to identify the set of relevant economic, political and institutional variables that we use in our analysis. The resulting set of explanatory variables is available in Table 2.1. In this section, we describe how we construct these explanatory variables.

Political Affiliations. We collect detailed information on the political affiliation of state legislators, both in the upper and lower chambers of legislatures, and that of the governor. Our data also allows us to identify years in which the control of legislatures or governorship has changed, as well as episodes of divided governments. Previous work has shown these to be important determinants of state policy (e.g. McCubbins (1991); Alt and Lowry (1994); Bernecker (2016)). We complement party control data with information on election cycles for state upper and lower chambers, governorship, and federal presidential elections (e.g. Alesina, Roubini and Cohen (1997)). In addition, we collect information on states' pledges in presidential elections, and DW Nominate scores for state representatives and senators.

Southern Democratic states. In our analysis we distinguish Southern and Northern Democratic parties. We identify the following states as Southern Democratic states: AL, AR, FL, GA, KY, LA, MS, NC, OK, SC, TN, TX, VA, WV, for all years before 2015.

Safe Republican and Democratic states. In some of our analysis we break down states into three categories: "safe" Republican, "safe" Democratic, or Swing state. Safe Republican (resp. Democrat) states are defined as those who had only voted for a Republican (resp. Democratic) presidential candidate in the past six elections, i.e. starting with 2000 presidential elections. The remaining states are considered to be swing states. Table A.6 summarizes these groups.

Institutional Rules. Previous work has also shown that state policy is influenced by institutional features, such as the number of legislators in the legislatures (Gilligan and Matsusaka (2001), Egger and Koethenbuerger (2010)), term limits (Erler (2007); Besley and Case (1995*a*)), balanced budget provisions (e.g. Poterba (1994)), and legislative initiative rules (Matsusaka (1995), Matsusaka (2000), Asatryan, Baskaran and Heinemann (2017), Asatryan et al. (2017)). Therefore, in addition to the political affiliation of the state legislators and governors in each year, we collect information on institutional features of the state. The size of the legislatures – number of seats in each legislative chamber – has been obtained from Ballotpedia.[1] Information on the applicable term limits in state legislatures and when they were introduced has been obtained from the National Conference on State Legislatures (NCSL), while information on governor term limits was obtained from the Council of State Governments. We have identified all state-year observations during which an incumbent governor could no longer seek re-election because of the binding term limit. We also collect information on average durations of legislative sessions, as well as salaries and per-diem rates in 2019/2020 from NCSL. In contrast to the federal government, states are not allowed to carry budgetary deficits for prolonged periods of time. We collect information on the stringency of balanced budget rules as of 2010: whether the governor must submit a balanced budget, whether legislatures must enact a balanced budget, and whether deficit carry-forwards are allowed, all from NCSL (2010). We also identify states with separate capital budgets in addition to operating budgets using 2014 data from NASBO (2014). We also collect information as

[1] For Nebraska, we utilize the total number of seats as our measurement for both the number of upper chamber seats and the number of lower chamber seats.

to whether states have a rainy day fund and the year it was adopted.

States differ in who can propose new laws. We obtain information on voter initiatives from Matsusaka (1995): a number of states allow citizens to initiate and approve laws by popular vote, while other states only allow state legislators to do so. These rules remain unchanged during the studied period. We also identify states that require supermajorities in order to pass tax increases, and whether the state is a right-to-work state in a given year.

Neighbors. To investigate whether states change their tax rates in response to competition, we identify tax rates in the neighboring states. We use two approaches to defining neighbors. First, we consider states as neighbors if they share a geographical border. We believe this is the best approach for excise taxes since individuals can purchase goods by driving to a neighboring state. Second, we use migration flows as measure of neighborliness, following Baicker (2005). Since tax competition is primarily concerned with out-migration, for each state, we identify five "neighbor" states that accept the largest number of migrants from that state, using 2010 state-to-state migration data from the IRS. While migration flows vary from year to year, the ranking of states, especially at the very top, appears to be fairly stable. For this reason – and due to the lack of consistent yearly data throughout the 70-year period – we use 2010 neighbors for all years. We calculate average tax rates in these five neighboring states, and we consider neighbors to change taxes if at least one of five states changed their tax rate. We use this approach to identify neighbors for all other tax types.

Recessions. We identify state recessions by applying the Bry-Boschan Method to Federal Reserve Bank of Philadelphia State Coincident Index. Since the Index is available from 1979 onward, we supplement our measure with equivalent calculations based on yearly state GDP values for years 1963-1979, and with federal recessions using NBER datings for years 1949-1962. Method identifies the peaks and troughs in the level of a time series, thus marking the beginning and ends of expansions and contractions. Our specification uses a window of 12 month, with a phase of at least 6 months and a complete cycle of 24 months. For 1949-1962, we rely on federal recessions using NBER datings. We also obtained information on natural resource prices (oil, natural gas and coal). We include inflation in our set of explanatory variables.

Mandates. Many federal policy changes impose substantial fiscal costs on state and local governments, as well as on the private sector. These federal mandates come in many different forms: from federal statutes that "order" costly changes (e.g. minimum wage mandates, or improving accessibility for the disabled), to federal policies that influence state spending by offering matching grants or other forms of incentives. Importantly, many of these mandates are unfunded and thus require states to raise more tax revenue or cut other expenditures in order to balance their budgets.

We use three sources to identify the federal mandate changes that are likely to have important economic consequences for state budgets. First, we use Congressional Budget Office (CBO) reports to identify mandates that exceed the "UMRA" threshold. A rapid increase in federal unfunded mandates led to the introduction of the Unfunded Mandates Reform Act of 1995 (UMRA), which required the CBO to estimate the costs of mandates to state and local governments, as well as the private sector, for new legislative proposals. While UMRA applies to most legislation that can impose enforceable duties, it typically does not apply to existing programs, Social Security, and legislation that cover national security and constitutional rights. Since UMRA's introduction in 1996, 15 laws have been enacted that have costs estimated exceed the 50 million 1996$ threshold (Congressional Research Service (2020)). Second, because UMRA did not apply before 1996, we look for costly mandates in the U.S. Advisory Commission on Intergovernmental Relations (ACIR) reports and National Conference of State Legislatures Mandate Monitor. Finally, we supplement these sources by hand-collecting information on historical changes to existing social welfare programs that are jointly funded by federal and state governments: AFDC/TANF, Food Stamp Program /SNAP, and Medicaid.[2]

Since our goal is to identify federal changes that may influence state tax policy, we focus on mandates that are large (i.e. likely to exceed the UMRA threshold) and are persistent in nature (i.e. affect state expenditures in all future years rather than impose a one-time burden). With these requirements in mind, we have identified 27 mandates summarized in Table A.8 that were enacted in 1950 or later. For each mandate, we record the year of mandate enactment and the year it became effective, as well as the list of states the mandate affected. While federal mandates apply to all states, they are not binding if a state had already satisfied the mandate prior to enactment.[3]

Demographics. We augment the political and institutional data with information on the demographics of each state. We obtain the poverty rate for 1980-2019 and population measures along with race and ethnicity breakdowns for 1969-2019 from the Census Bureau. We collect the unemployment rate, employment to population ratio, and labor force participation rate for 1976-2020 from the Bureau of Labor Statistics. Earlier year observations are collected from the Statistical Abstracts of the United States. Finally, we obtain information on state tax revenues, expenditures and total outstanding debt from Census Annual Survey of State Governments.

We obtain population measures along with race and age breakdowns for 1969-2019 from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. Population totals for 1949-1969 are obtained from the Statistical Abstracts of the United States. Breakdowns by race and age were obtained from the Sta-

[2]We do not include SSI, SSDI, and Medicare in our collection process because these programs are fully federally funded. Food Stamp / SNAP benefits are funded by the federal government, but administrative costs are shared with the states.

[3]For example, according to CBO calculations, federal minimum wage increases impose a substantial burden on state budgets through their direct effect on state employee salaries. However, any state with state minimum wage above the new federal wage was unaffected by this mandate.

tistical Abstracts of the United States for years 1950, 1960 and 1968. These values are then used in place of missing years, i.e 1950 value for years 1949-1955, 1960 value for years 1956-1963, and 1968 value for 1964-1968.

We obtain the poverty rate for 1980-2019 from Census and for years 1959, 1969 and 1975 from the Statistical Abstracts of the United States. These values are then used in place of missing years, 1959 for years 1949-1963, 1969 for years 1964-1972, and 1975 for years 1973-1979. Median household income values are available from Census for years 1979-2019, and are supplemented with values for 1950, 1959, 1969 and 1975 from the Statistical Abstracts of the United States. Again, the latter values (but inflation-adjusted) are used in place of missing data: i.e 1950 value for years 1949-1955, 1959 value for years 1956-1963, 1969 value for 1964-1972, 1975 value for 1973-1978.

We collect the unemployment rate, employment to population ratio, and labor force participation rate for 1976-2020 from the Bureau of Labor Statistics. Unemployment rate and total unemployment for 1957-1975 were obtained from the Manpower Report of the President and the Employment and Training Report of the President. For 1957-1970, employment to population ratio is estimated as the number of employed individuals (obtained by multiplying one-minus the unemployment rate by the size of the labor force, i.e. unemployment divided by the unemployment rate) divided by the the number of prime age-adults (i.e. age 19-65). Labor force participation rate is estimated as the number unemployment divided by the unemployment rate and divided by the number of prime age-adults (i.e. age 19-65). Values for earlier years (1949-1956) are filled with values from 1957.

Oil, gas and coal prices. Crude oil prices are represented by the historical free market (stripper) oil prices of Illinois Crude from Illinois Oil and Gas Association and Plains All American Oil. Natural gas price is based on Wellhead price until 2012, after which it is Citygate price minus 2.07; both from U.S. Energy Information Administration.

Coal prices were obtained from U.S. Bureau of Labor Statistics, Producer Price Index by Commodity: Fuels and Related Products and Power: Bituminous Coal and Lignite (WPS0512) (averaged over a year), retrieved from FRED, Federal Reserve Bank of St. Louis.

A.2.2 Additional Tables & Figures

Table A.5: Plausible Explanatory Variables Based on Previous Literature

Studies	**Suggested explanatory variables**
Election Cycles:	
Mikesell (1978), Rosenberg (1992), Foremny and Riedel (2014), Katsimi and Sarantides (2012), Nelson (2000), Chang et al. (2020)	election cycle year indicators
Ashworth, Geys and Heyndels (2006)	election cycle year indicators, neighbors' tax rates, coalition vs single-party in control indicator
Veiga and Veiga (2007)	election cycle year indicators, salience of tax instrument
Rose (2006)	election cycle year indicators, election cycle year indicators x deficit not allowed indicator
Political Structures:	
Alt and Lowry (1994)	divided government indicator, divided government indicator x deficit not allowed indicator
McCubbins (1991)	divided government indicator, party of the president
Bernecker (2016)	divided government indicator, governor election cycle year indicator, percent of female legislators in the legislature
Castanheira, Nicodème and Profeta (2012)	size of majority, election cycle year indicators, recession indicator, tax reform the year prior indicator
Roubini and Sachs (1989)	government tenure, coalition vs single party in control indicator
Institutional Rules:	
Erler (2007)	legislator term limit indicator
Besley and Case (1995*a*)	governor term-limited, governor term-limited x Democrat, governor term-limited x Republican
Gilligan and Matsusaka (2001), Egger and Koethenbuerger (2010)	size of senate, size of house
Matsusaka (1995), Matsusaka (2000), Asatryan, Baskaran and Heinemann (2017), Asatryan et al. (2017)	voter initiative indicator, voter initiative indicator x complexity of voter initiative requirements
Poterba (1994)	deficit not allowed indicator, tax limitations, general fund balance, divided government x deficit not allowed, governor election cycle year indicators

Table continues on next page.

Notes: This table summarizes variables that are likely to explain variation in state tax policies based on the previous studies.

Table A.5: Plausible Explanatory Variables Based on Previous Literature

Studies	**Suggested explanatory variables**
Competition:	
Besley and Case (1995*b*), Chirinko and Wilson (2017), Deskins and Hill (2010), Rork (2003)	neighbors' tax rates
Buettner (2003)	neighbors' tax rates, neighbors' tax rates x size of state
Case, Rosen and Hines (1993)	neighbors' spending, as defined based on economic and geographic similarities
Besley and Rosen (1998), Goodspeed (2000), Goodspeed (2002), Devereux, Lockwood and Redoano (2007),	neighbors' tax rates, federal tax rates
Geys (2006)	neighbors' ratio of the cost of public goods provision to the level of public goods actually provided by the government, also interacted with coalition vs single-party in control indicator
Baicker (2005)	neighbors' tax rates, defined based on degree of mobility between states
Bordignon, Cerniglia and Revelli (2003)	neighbors' tax rates x mayor term-limited, election year indicators, demographics: unemployment, elderly and young shares of population
Other:	
Inman and Fitts (1990)	income level, unemployment level, demands from special interest groups, share of young people in population, strength of party control
Bozzano et al. (2021)	gender equality level

Table A.6: "Safe" Republican and Democratic States

Safe Republican States	AL, AK, AR, ID, KS, KY, LA, MO, MS, MT, NE, ND, OK, SC, SD, TN, TX, UT, WV, WY
Swing States	AZ, CO, FL, GA, IA, IN, MI, NC, NH, NM, NV, OH, PA, VA, WI
Safe Democratic States	CA, CT, DE, HI, IL, ME, MD, MA, MN, NJ, NY, OR, RI, VT, WA

Notes: Safe Republican (resp. Democrat) states are defined as those who had only voted for a Republican (resp. Democratic) presidential candidate in the past six elections, i.e. starting with 2000 presidential elections. The remaining states are considered to be swing states.

Table A.7: Correlation Matrix

	Top Personal	Min Personal	Top Corporate	Min Corporate	Sales	Cigarette	Gasoline	Alcohol Spirit
Top Personal	1	0.49	0.58	0.46	-0.04	0.06	-0.02	0.01
Min Personal	0.49	1	0.46	0.51	0.17	0.2	-0.12	-0.18
Top Corporate	0.58	0.46	1	0.8	0.18	0.19	-0.26	-0.03
Min Corporate	0.46	0.51	0.8	1	0.2	0.19	-0.17	-0.06
Sales	-0.04	0.17	0.18	0.2	1	0.29	-0.36	-0.26
Cigarette	0.06	0.2	0.19	0.19	0.29	1	-0.25	-0.06
Gasoline	-0.02	-0.12	-0.26	-0.17	-0.36	-0.25	1	0.27
Alcohol Spirit	0.01	-0.18	-0.03	-0.06	-0.26	-0.06	0.27	1

Notes: This table shows the correlation matrix of 6 tax rates. Personal and corporate income taxes are represented by top rates, all 50 states included.

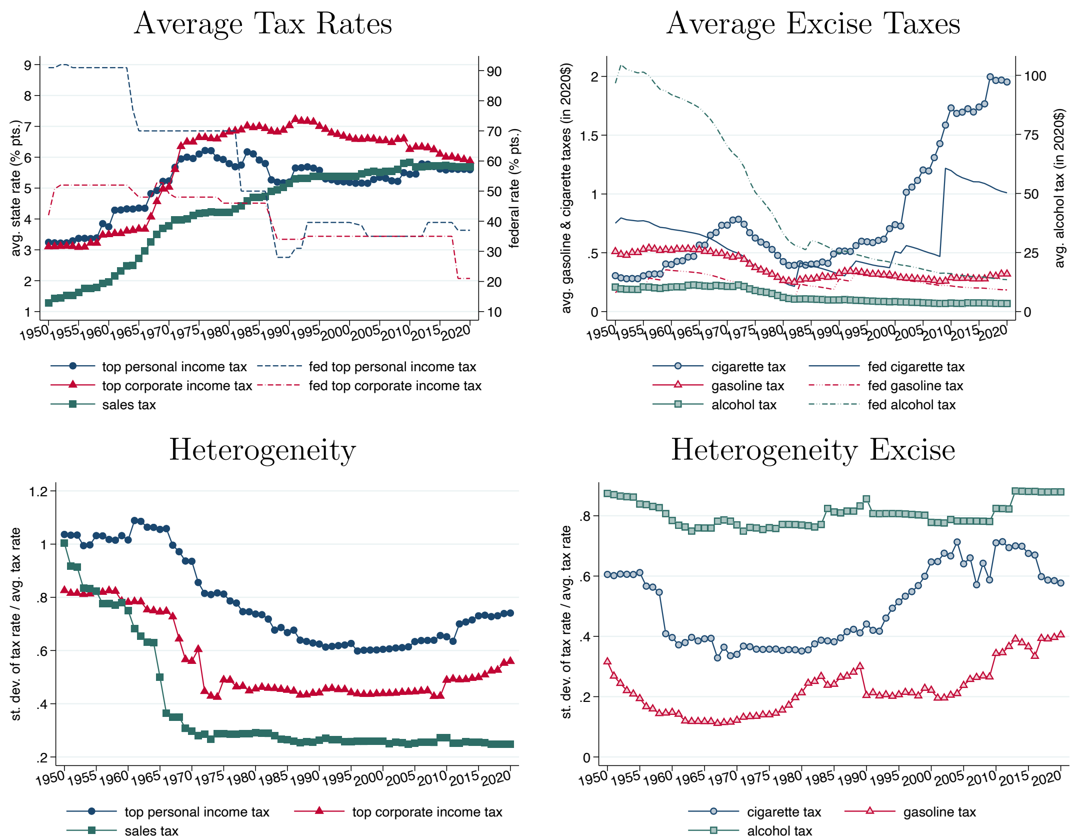
Table A.8: Federal Mandates

Mandate	Enacted	Effective	States affected
Medicaid: Mandatory preventative services for children	1967	1973	All states except AL, AK, AZ, AR, CO, FL, IN, MS, NJ, NC, SC, TN, VA
FSP/SNAP: Mandatory expansion	1973	1974	All states
FSP/SNAP: Expanded eligibility	1977	1979	All states
Medicaid: Mandatory coverage for pregnant women and infants up to 100% FPL	1988	1989	CO, ID, IN, MT, ND, NH, NV, NY, WI
AFDC: Mandatory coverage for 2-parent families w/ unemployed primary earner	1988	1990	AK, AL, AR, AZ, CO, FL, GA, ID, IN, KY, LA, MS, ND, NH, NM, NV, OK, SD, TN, TX, UT, VA
Medicaid: Requirement to cover pregnant women and young children up to 133% FPL	1989	1990	All states except: CA, CT, IA, ME, MA, MI, MN, MS, RI, VT, WV
AFDC: AFDC ended; replaced by Temporary Assistance for Needy Families (TANF) w/ looser spending restrictions	1996	1997	All states
FSP/SNAP: Reduced reimbursement of state administration costs	1998	1998	All states
Min wage increase	1950	1950	All states except: AK not affected
Min wage increase	1956	1956	All states except: AK not affected
Min wage increase	1961	1961	All states except: AK not affected
Min wage increase	1963	1963	All states except: AK not affected
Min wage increase	1967	1967-1968	All states except: AK, CA not affected
Min wage increase	1974	1974-1976	All states except: AK, HI not affected
Min wage increase	1977	1979-1981	All states except: AK, CT not affected
Min wage increase	1990	1990-1991	All states in 1990, except: HI, IA, ME, MN, VT, WA in 1991; AK, CA, CT, OR, RI not affected
Min wage increase	1996	1996-97	All states in 1996, except: NJ and WA in 1997; AK and HI not affected.
Min wage increase	2007	2007-09	All states in 2007 except: AR, MN, NV in 2008; AK, AZ, DE, FL, NJ, NY in 2009; CA, CT, HI, IL, ME, MA, MI, OR, RI, VT, WA, WV not affected.
Clean Air Act	1963, 1967, 1970, 1977, 1990	1963, 1967, 1970, 1977, 1990	All states
Occupational Safety and Health Act	1970	1970	All states
Federal Water Pollution Control Act	1972, 1977, 1987	1972, 1977, 1987	All states
Marine Protection Research and Sanctuaries Act	1972	1972	All states
Endangered Species Act	1973	1973	All states
Safe Drinking Water Act	1974, 1986, 1996	1974, 1986, 1996	All states
Surface Mining Control and Reclamation Act	1977	1977	All states
Internet Tax Freedom Act	1998	2020	HI, NM, ND, OH, SD, TX, and WI.
Healthy, Hunger-Free Kids Act	2010	2012	All states

Notes: This table summarizes federal mandates enacted in 1950 or later that are likely to impose a substantial burden on state budgets, i.e. have projected costs that exceed the UMRA threshold ($50 million 1996 dollars). See Section A.2.1 for details.

Figure A.13: State Tax Rates Over Years Weighted by Population

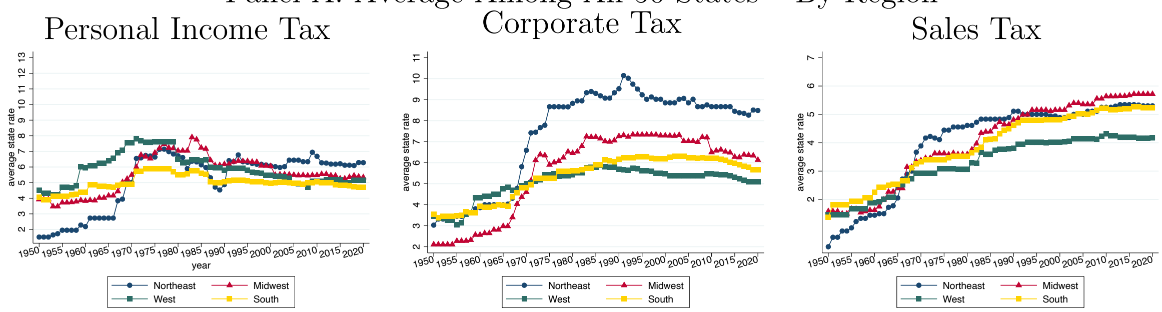
All 50 States



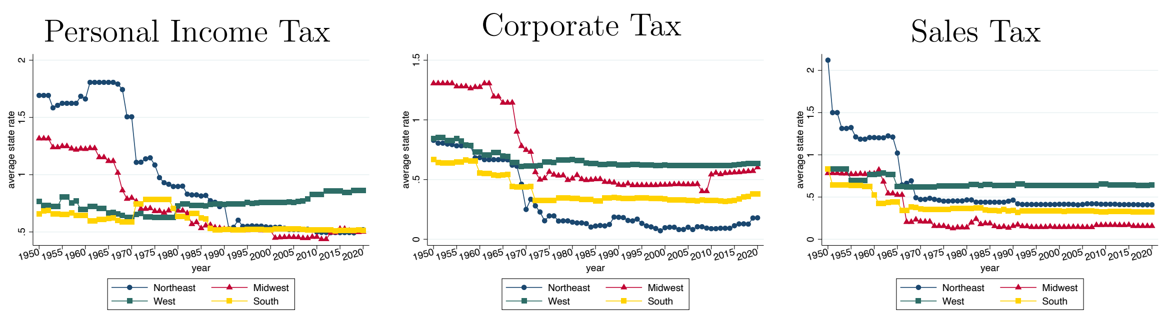
Notes: Figures (a) and (b) show average top personal income and corporate tax rates, sales tax rates, and average cigarette, alcohol (spirit) and gasoline tax rates, as well as corresponding federal tax rates. Figures (c) and (d) show the standard deviation of the state taxes divided by average tax rate (coefficient of variation). All states included, including those with zero rates. In all figures observations are weighted by population.

Figure A.14: Long Term Trends by Region

Panel A: Average Among All 50 States – By Region

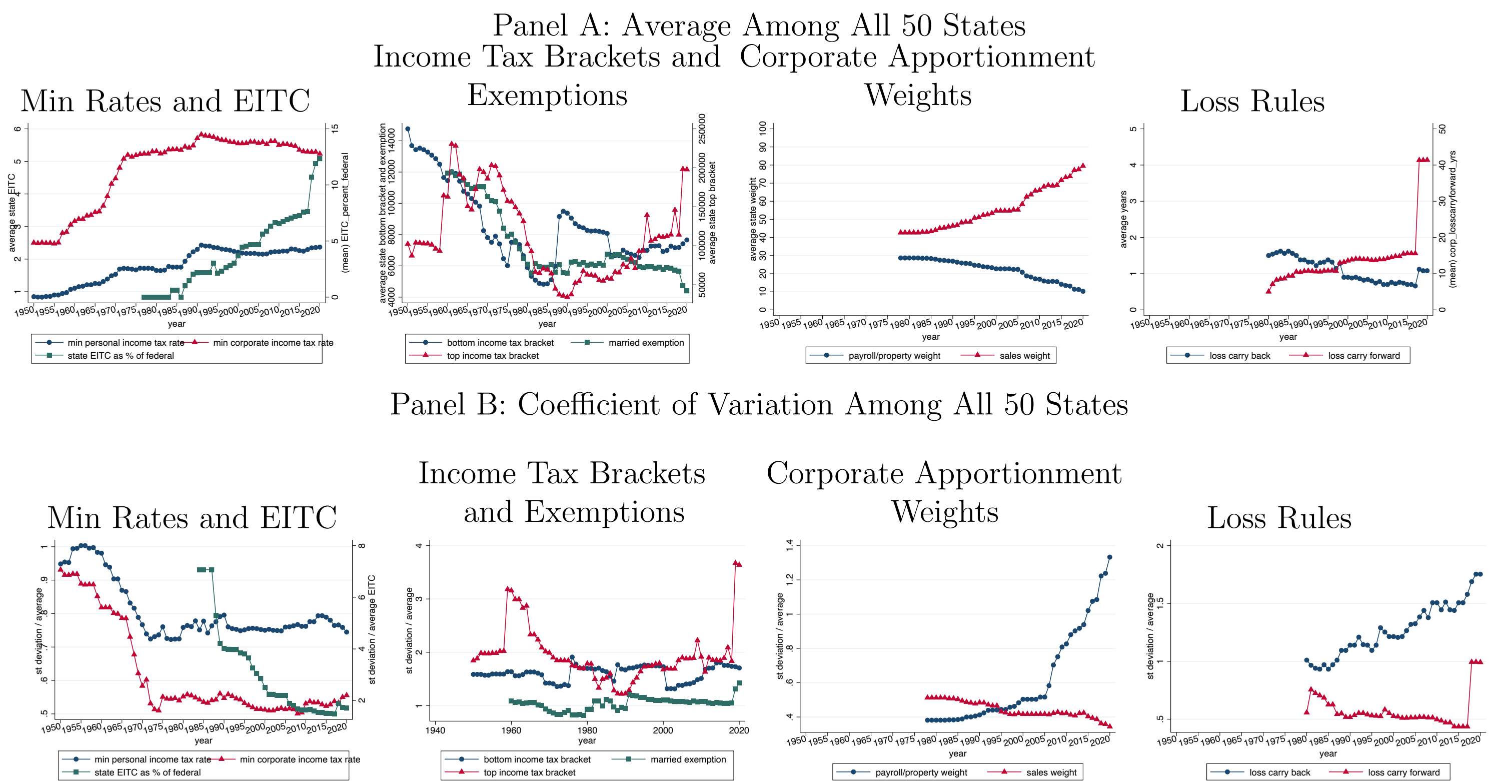


Panel B: Coefficient of Variation All 50 States – By Region



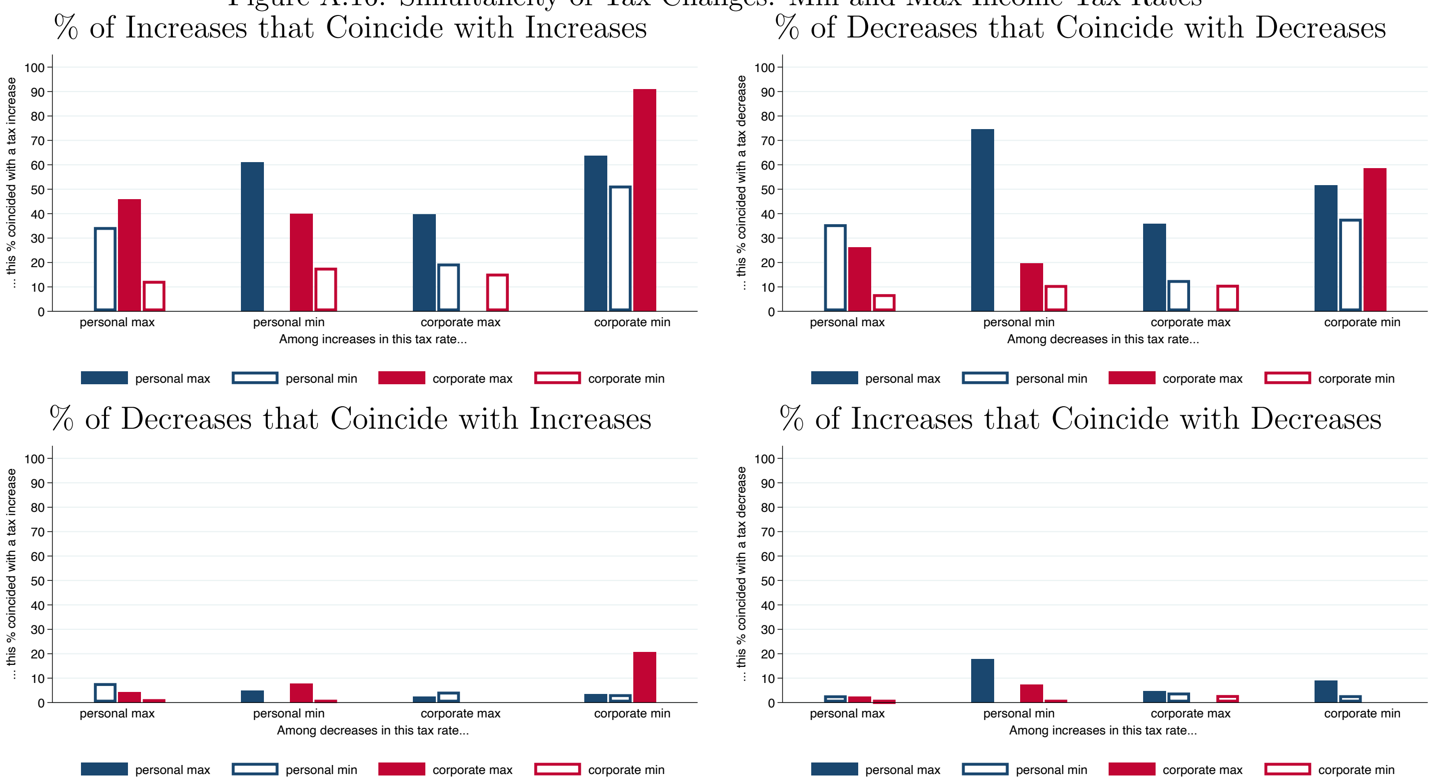
Notes: These figures show the average as well as the standard deviation of the state values divided by average value (coefficient of variation). All states included, including those with zero rates, but broken down by regions.

Figure A.15: Long Term Trends: Additional Tax Rules



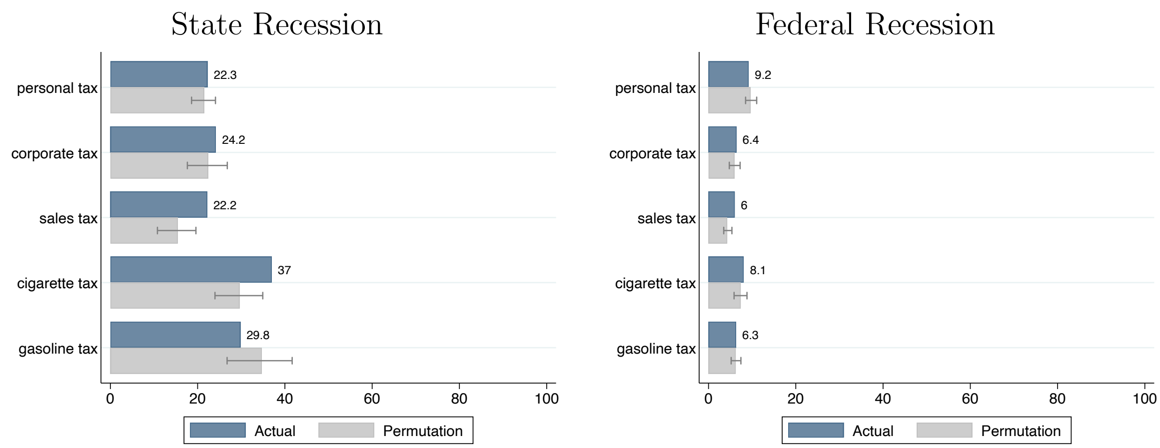
Notes: These figures show the average as well as the standard deviation of the state values divided by average value (coefficient of variation). All states included, including those with zero rates. Unlimited loss carryforwards are coded as 100 years.

Figure A.16: Simultaneity of Tax Changes: Min and Max Income Tax Rates



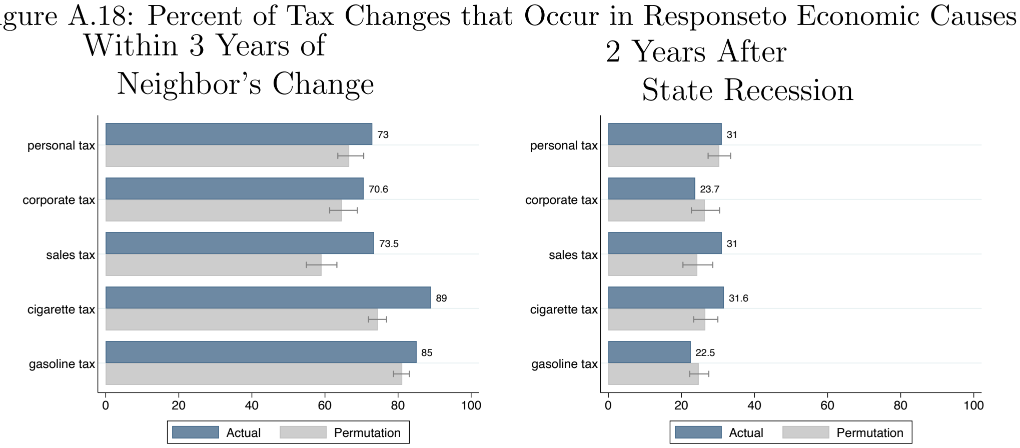
Notes: These figures explore the extent to which states change one tax rate while simultaneously changing another tax type (i.e., in the same year). Among the increases (or decreases) in each tax on the x-axis, the vertical bars specify the share that coincides with an increase (or decrease) in another tax type in the same state and year. These other tax types are identified by the color of the bar (top income tax rates, top corporate tax rates, minimum income tax, minimum corporate tax).

Figure A.17: Percent of Recession Episodes that Result in Tax Changes



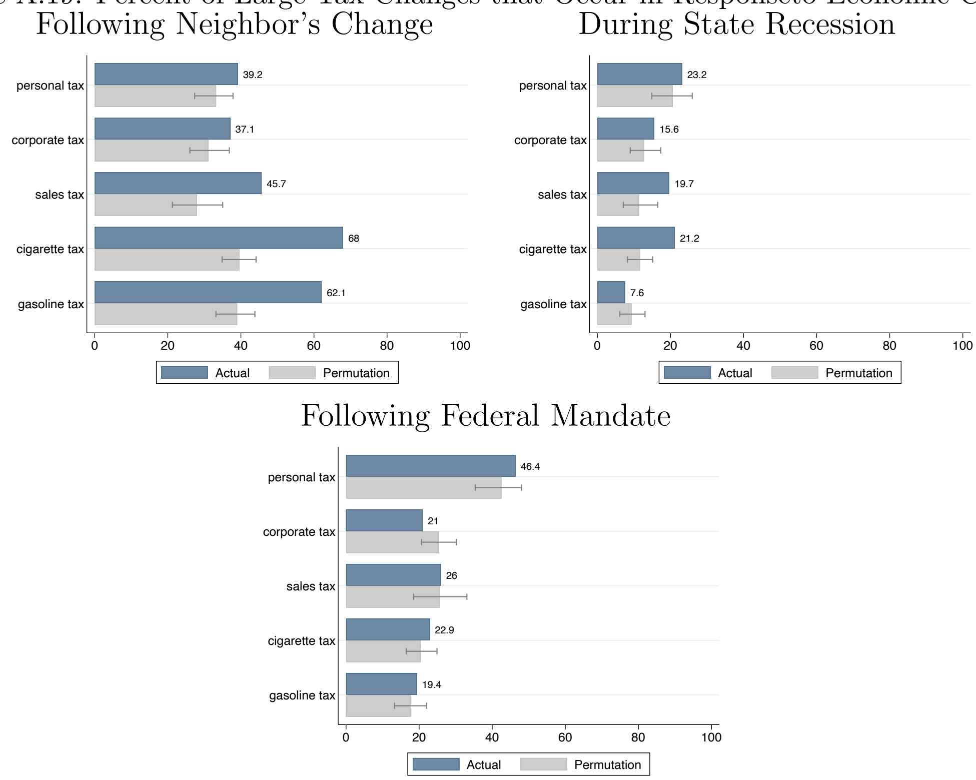
Notes: This figure shows the percent of (a) state recessions or (b) federal recessions that lead to a tax change. Each recession episode is counted as one recession and only one tax change (per tax rate type) is allowed per recession. In all figures, the top blue bars show actual observed percentages, while the bottom grey bars show the simulated average, calculated by randomizing the timing of tax changes 100 times. The thin interval bars show the 5th and 95th percentiles of the simulated percentages.

Figure A.18: Percent of Tax Changes that Occur in Responseto Economic Causes



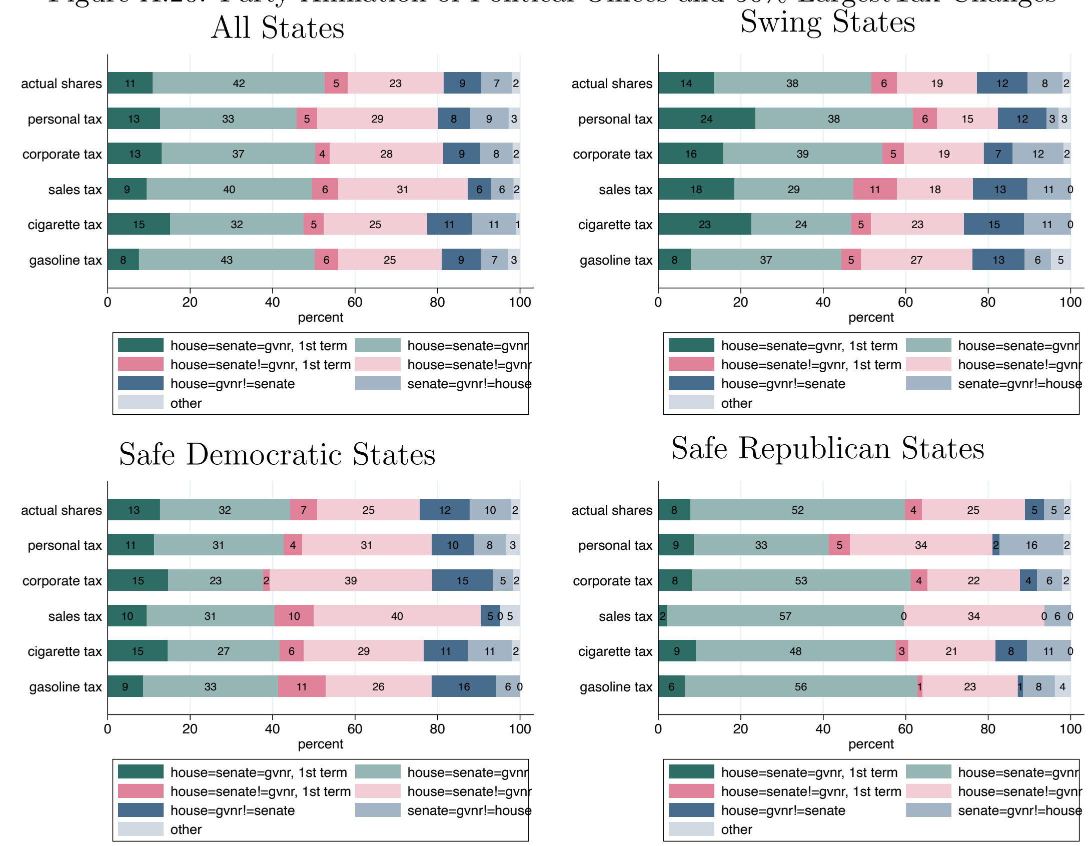
Notes: This figure shows the percent of tax changes that occur (a) within 3 years after neighboring state changes its tax rate; (b) during a state recession or a year after. In all figures, the top blue bars show actual observed percentages, while the grey bars show the simulated average, calculated by randomizing the timing of tax changes 100 times. The thin interval bars show the 5th and 95th percentiles of the simulated percentages.

Figure A.19: Percent of Large Tax Changes that Occur in Responseto Economic Causes



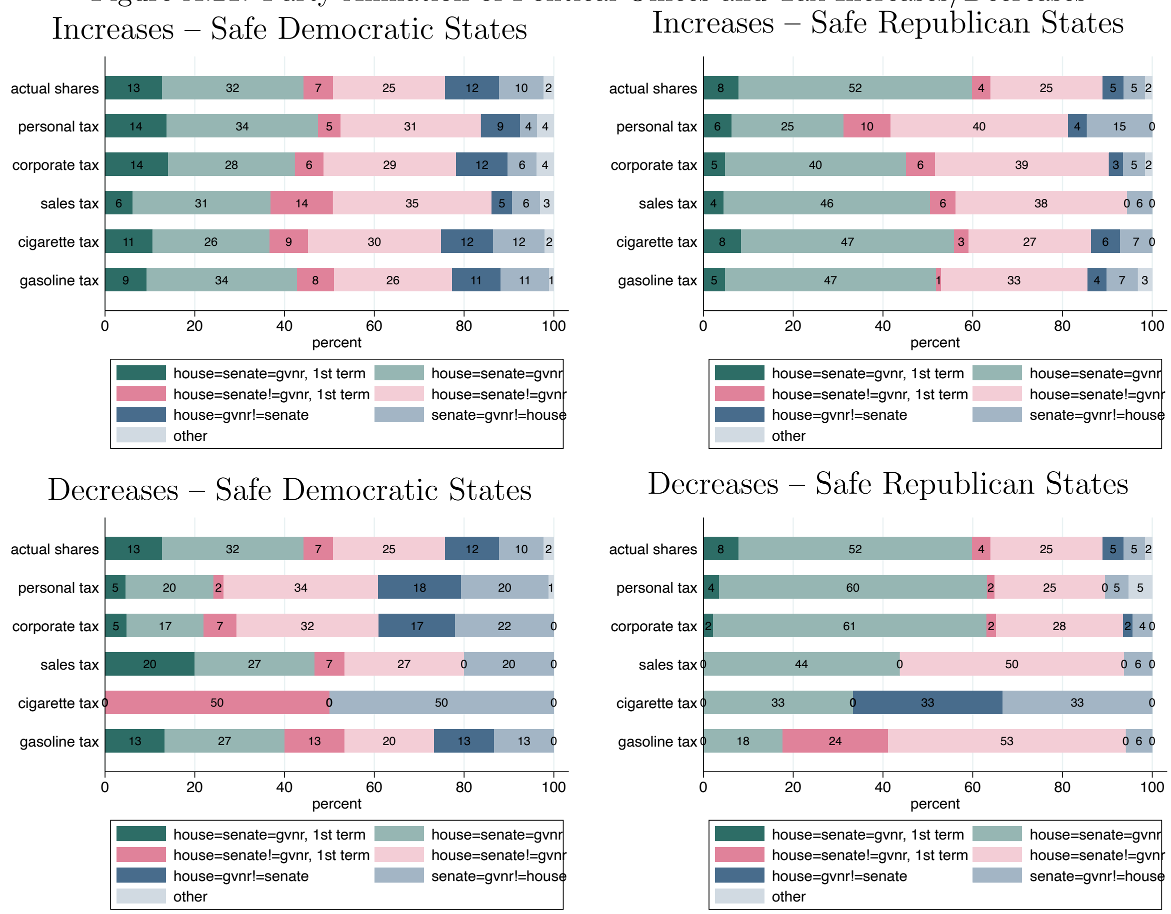
Notes: This figure shows the percent of large tax changes (top 50th percentile) that occur (a) in the same year or 1 year after neighboring state changes its tax rate; (b) during a state recession, or (c) in the year the federal mandate becomes enacted or effective. In all figures, the top blue bars show actual observed percentages, while the grey bars show the simulated average, calculated by randomizing the timing of tax changes 100 times. The thin interval bars show the 5th and 95th percentiles of the simulated percentages.

Figure A.20: Party Affiliation of Political Offices and 50% LargestTax Changes



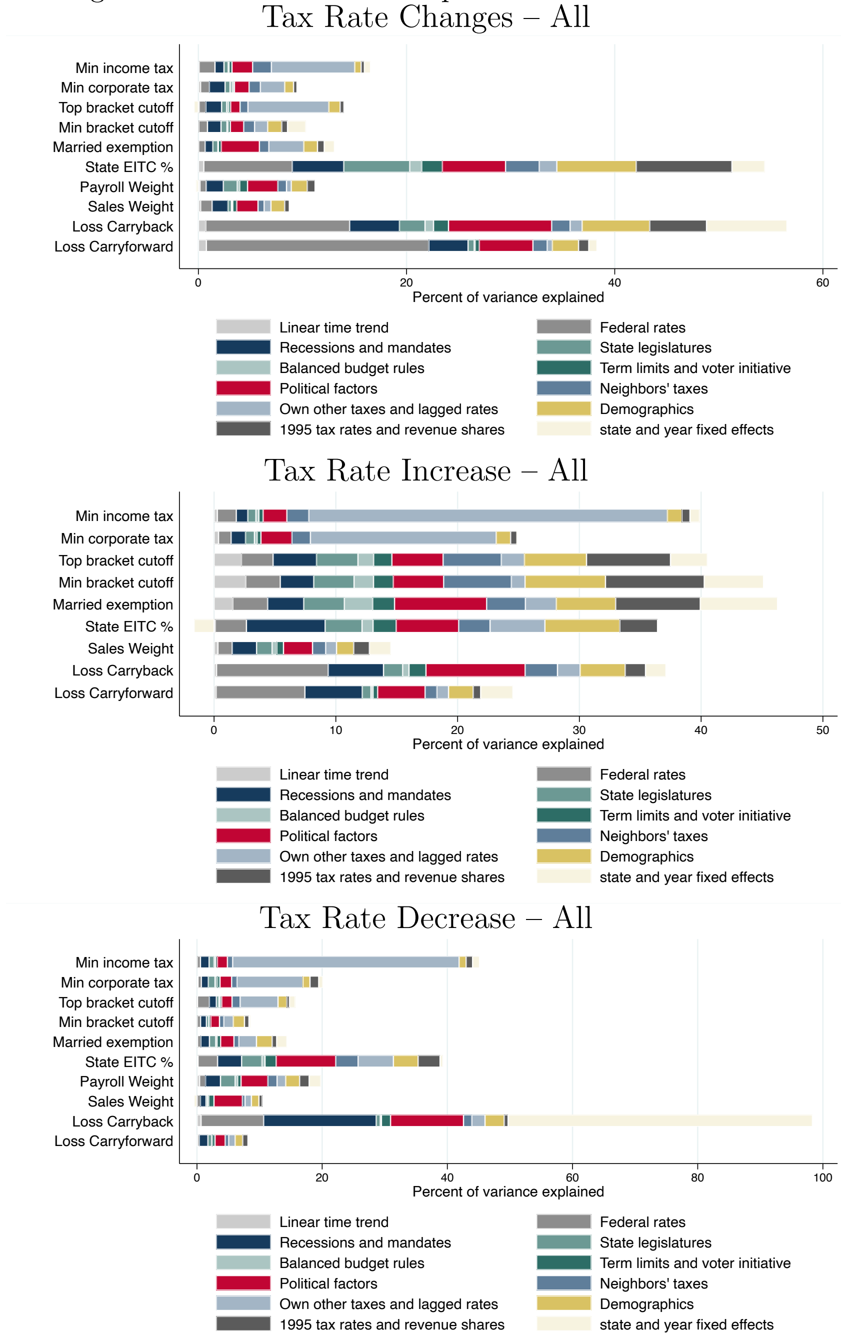
Notes: The top row of each figure shows the percent of yearly observations in which (i) the majority party of the House is the same as that of the Senate and of the Governor, and one of these three bodies switched party control; (ii) same as (i) but no party control change; (iii) House and Senate majorities are the same party, but Governor of a different party, and the joint majorities in House and Senate were obtained this term; (iv) same as (iii) but no party control change; (v) House majority matches Governor's affiliation but not Senate majority's; (vi) Senate majority matches Governor's affiliation but not House majority's; (vii) all other options (i.e. non-Democratic/Republican affiliations or lack of majorities). The next five rows show party affiliations in years when respective large (top 50% percentile) tax changes occur. Figures (c) and (d) provide these statistics separately for states that have only voted for a Democratic (Republican) presidential candidate since 2000 elections (see Table A.6), while Figure (b) for all other states.

Figure A.21: Party Affiliation of Political Offices and Tax Increases/Decreases



Notes: The top row of each figure shows the percent of yearly observations in which (i) the majority party of the House is the same as that of the Senate and of the Governor, and one of these three bodies switched party control; (ii) same as (i) but no party control change; (iii) House and Senate majorities are the same party, but Governor of a different party, and the joint majorities in House and Senate were obtained this term; (iv) same as (iii) but no party control change; (v) House majority matches Governor's affiliation but not Senate majority's; (vi) Senate majority matches Governor's affiliation but not House majority's; (vii) all other options (i.e. non-Democratic/Republican affiliations or lack of majorities). The next five rows show party affiliations in years when respective tax changes occur. These statistics are shown separately for states that have only voted for a Democratic (Republican) presidential candidate since 2000 elections (see Table A.6) and for tax increases and decreases.

Figure A.22: Variance Decomposition – Other Tax Rules



Notes: This figure shows the Shapley variance decomposition of adjusted R^2 for (a) tax rule changes (in \$, or pp, or otherwise), (b) all tax rule increases (indicators for years when a tax rule increase occurs), (c) all tax rule decreases (indicators for years when a tax rule decrease occurs). All decompositions use the 166 variables summarized in Table 2.1 plus state and year fixed effects.

A.3 Geographic Variation in C-Sections in the United States: Trends, Correlates, and Other Interesting Facts

A.3.1 Selection of Covariates

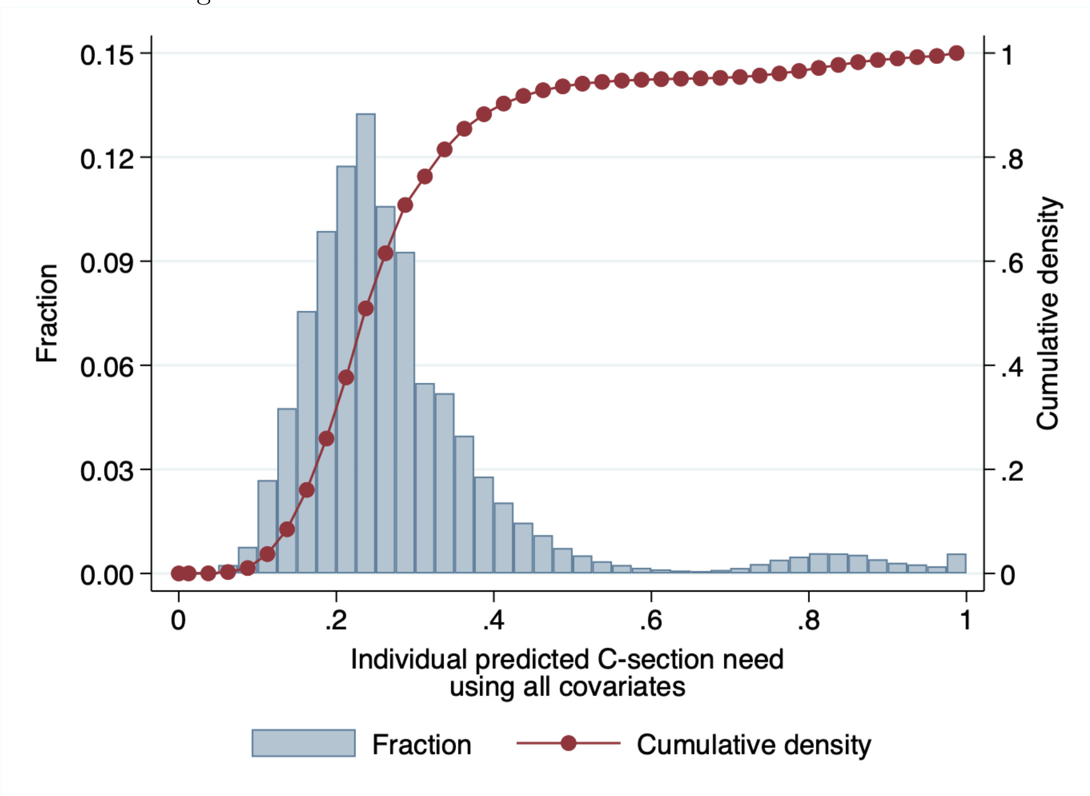
Table A.9: Comparison of Covariates with Select Literature

	This paper High-risk (vs. low-risk) births	*This paper* Predicted C-section need	Card, Fenizia and Silver (2023)	Currie and MacLeod (2017)
Maternal age	<18 or >35	5-year bins	<18 or >35	5-year bins
Gestational age	<37 weeks	<37 weeks	<37 weeks	–
Prenatal visits	≥ 19	≥ 19	>20	–
Growth restrictions	✓	✓	✓	–
Breech	✓	✓	✓	✓
Eclampsia	✓	✓	✓	–
Pre-eclampsia	✓	✓	✓	–
Diabetes	✓	✓	–	–
BMI (avail. $\geq$ 2009)	–	*See Table 3.1*	90^{th} pctile	–
Placenta previa (avail. $\leq$ 2006)	–	–	–	✓
Abruptio placenta (avail. $\leq$ 2006)	–	–	–	✓
Cord prolapse (avail. $\leq$ 2006)	–	–	–	✓
Multiple birth	N/A (singleton 1^{st} births)		✓	✓
Birth order	N/A (singleton 1^{st} births)		Non-first	✓
Previous C-section	N/A (singleton 1^{st} births)		N/A	✓
Previous large infant	N/A (singleton 1^{st} births)		N/A	✓
Previous preterm	N/A (singleton 1^{st} births)		N/A	✓
Non-medical factors	–	*See Table 3.1*	–	–

Notes: This table describes the sets of covariates used to define high- vs. low-risk births; to adjust county rates of c-section, maternal & neonatal mortality, and maternal & infant morbidity; and to predict C-section need for individuals. See Table 3.1 for additional information about the use of covariates in this paper.

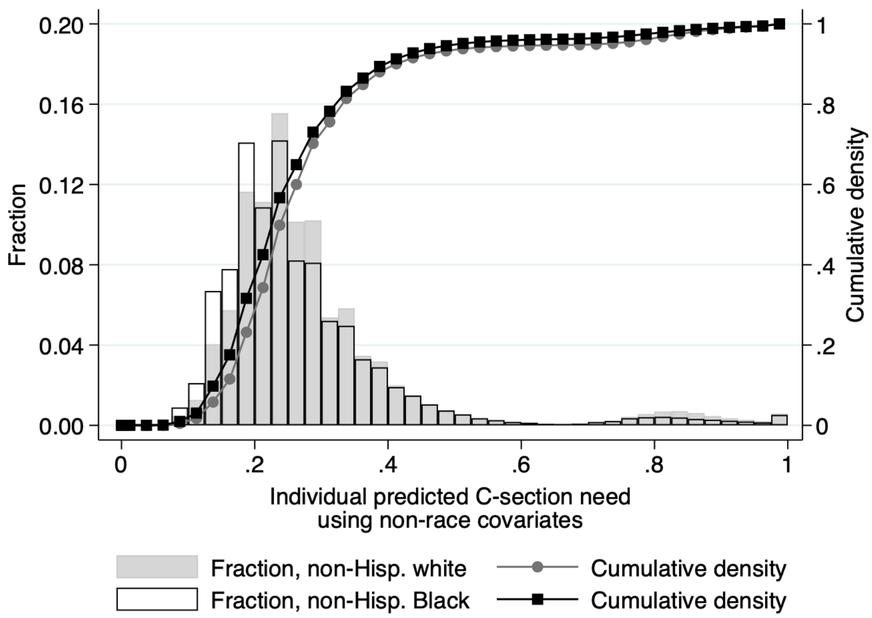
A.3.2 Predicted C-Section Need

Figure A.23: Individual Predicted C-Section Need 2015-2017

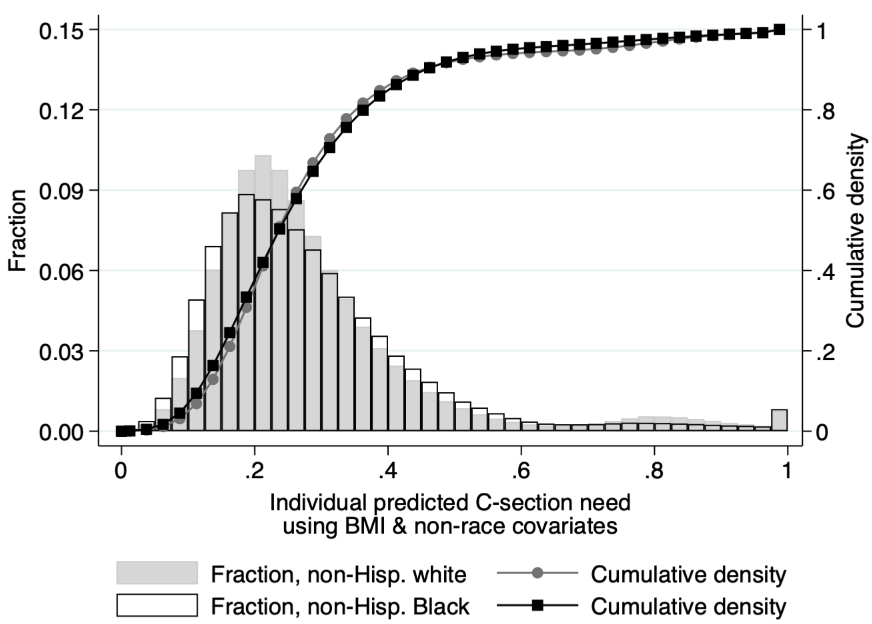


Notes: This figure shows the distribution of the individual predicted C-section need for singleton first births using all covariates (after controlling for county fixed effects).

Figure A.24: Individual Predicted C-Section Need by Race 2015-2017
Based on Non-Race Covariates



Based on Non-Race Covariates plus BMI



Notes: This figure shows the distribution of the individual predicted C-section need for singleton first births using non-race covariates and non-race covariates plus BMI (after controlling for county fixed effects).

A.3.3 Using Medical Covariates Only

Table A.10: Distribution of Births Across Low- and High-Risk
by Predicted C-Section Need
(Adjusted for Medical Covariates Only)

1989-1991

	Predicted C-section need (using medical covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.32 %	0.00 %	60.74 %
High-risk births	33.28 %	100.00 %	35.99 %
Unknown risk births	3.40 %	0.00 %	3.27 %
	100%	100%	100%

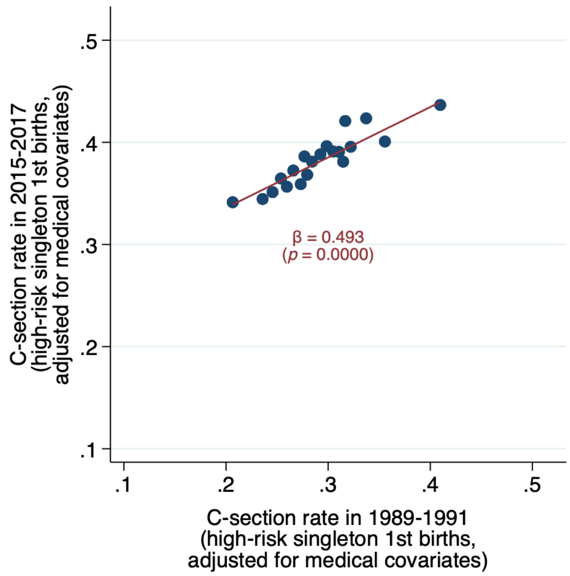
2015-2017

	Predicted C-section need (using medical covariates)		
	≤ 0.6	*> 0.6*	*Overall*
Low-risk births	63.64 %	0.00 %	60.41 %
High-risk births	34.05 %	100.00 %	37.40 %
Unknown risk births	2.31 %	0.00 %	2.19 %
	100%	100%	100%

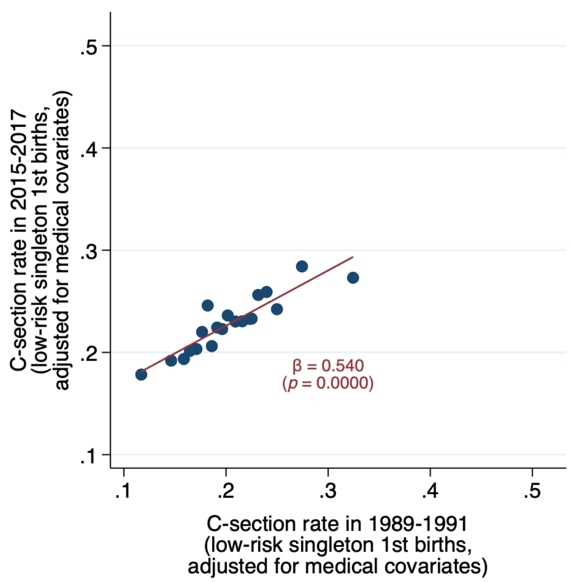
Notes: This table shows the relationship between two different approaches for assessing risk of C-section: (1) the categorization of births as low- or high-risk, a binary assignment based on medical factors only, and (2) the predicted C-section need, a continuum of risk estimated using the same medical factors. See Table 3.1 for the specific covariates used in each model. Only singleton first births are represented. A small portion of singleton first births have no observed high-risk characteristics but are missing data, and thus are not classified as either low- or high-risk.

Figure A.25: Persistence in County C-Section Rates Over Time (Adjusted for Medical Covariates Only)

High-Risk Singleton 1st Births



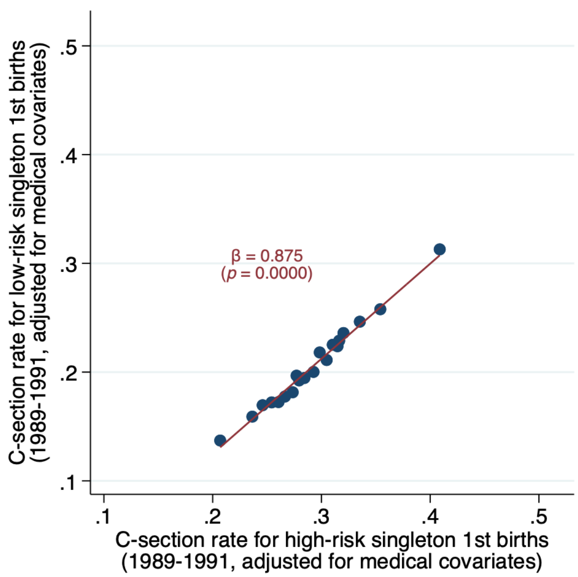
Low-Risk Singleton 1st Births



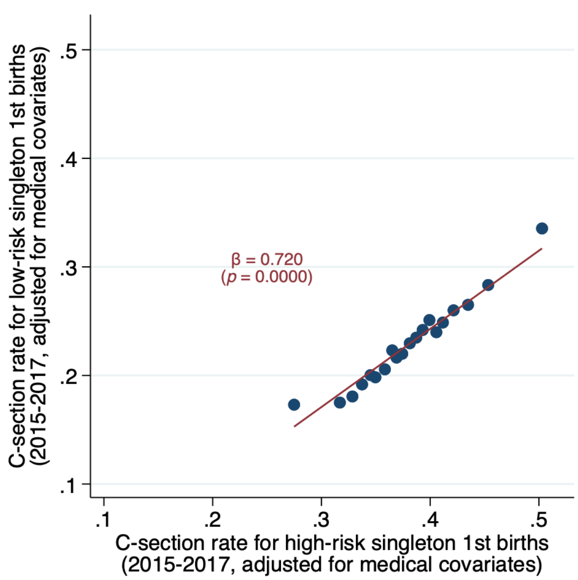
Notes: This figure shows binscatter plots of C-section rates across time periods. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over all six years.

Figure A.26: Correlation in County C-Section Rates Across Risk Type (Adjusted for Medical Covariates Only)

1989-1991

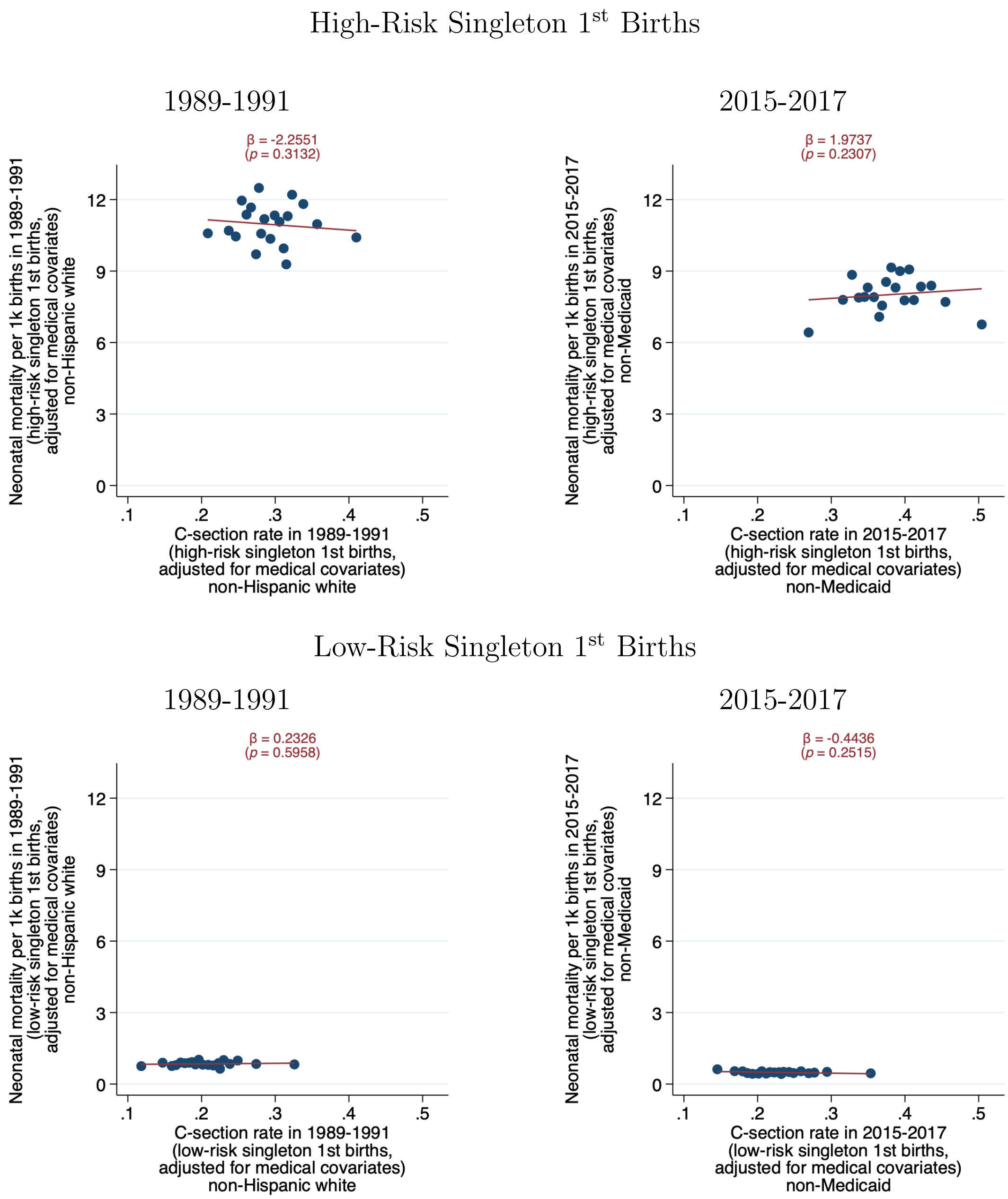


2015-2017



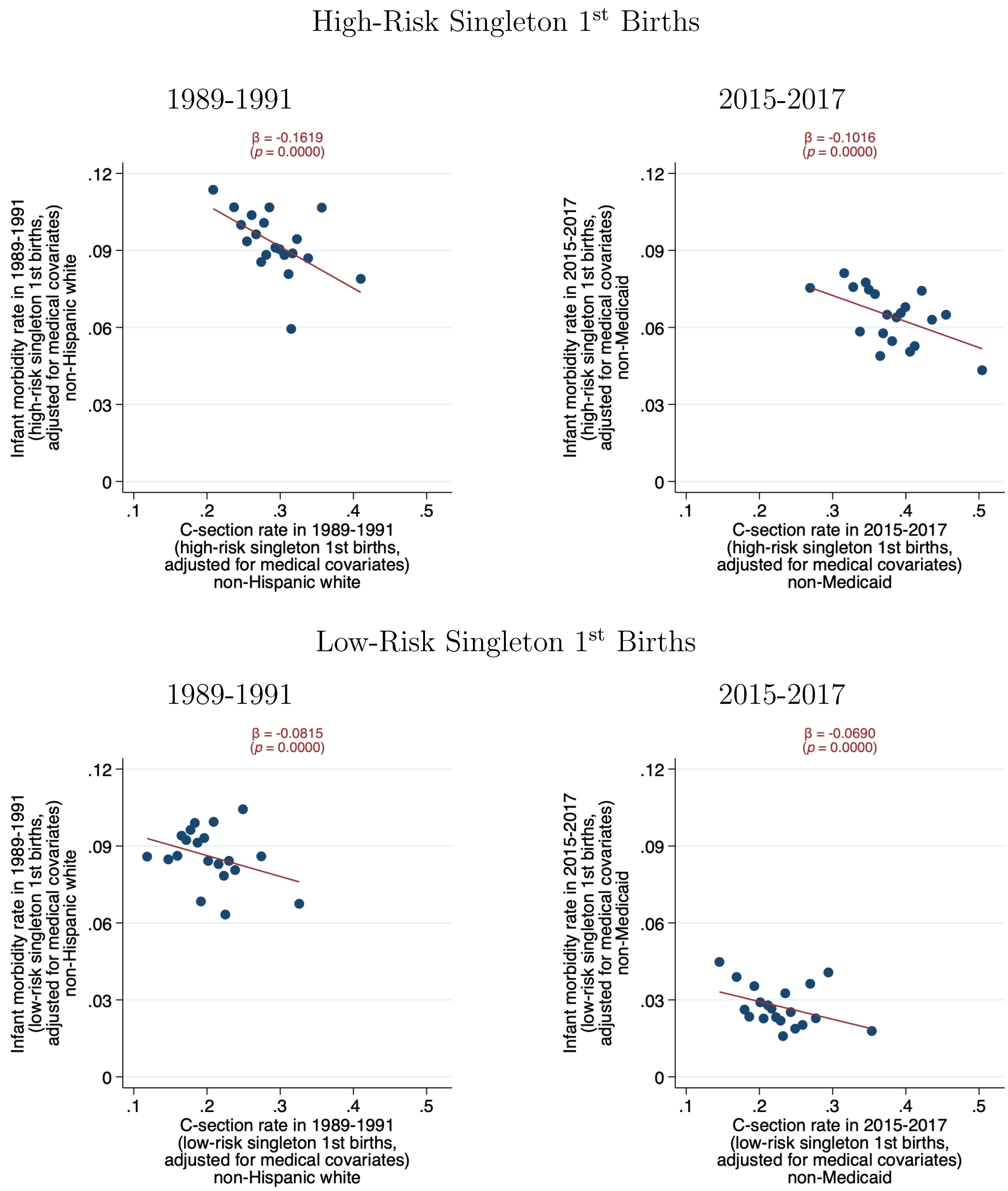
Notes: This figure shows binscatter plots of C-section rates across risk types. Linear fit and p-value are based on the underlying counties (prior to binning). C-section rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of singleton first births in the county over the three years.

Figure A.27: Correlation of County C-Section Rates and Neonatal Mortality (Adjusted for Medical Covariates Only)



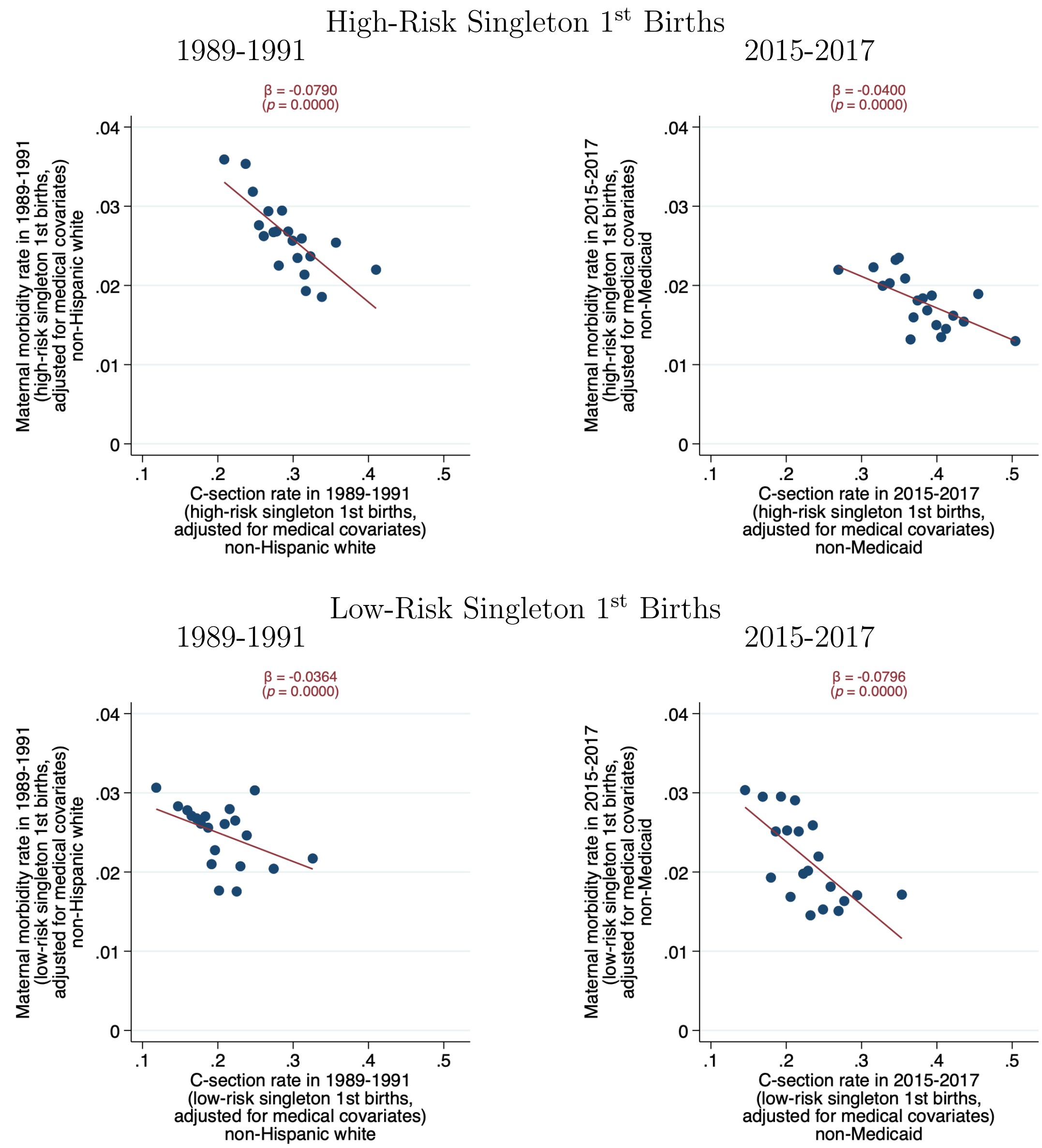
Notes: This figure shows binscatter plots of C-section rates with neonatal mortality rates. Linear fit and p-value are based on the underlying counties (prior to binning). C-section and neonatal mortality rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of relevant births in each year. Binscatter and linear fit are weighted by the number of relevant births in the county over the three years.

Figure A.28: Correlation of County C-Section Rates and Infant Morbidity (Adjusted for Medical Covariates Only)



Notes: This figure shows binscatter plots of C-section rates for low-risk singleton first births with maternal or infant morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. C-section and morbidity rates (adjusted for all covariates) are averaged over the three-year period, weighted by the number of low-risk singleton first births in each year. Binscatter and linear fit are weighted by the number of low-risk singleton first births in the county over the three years.

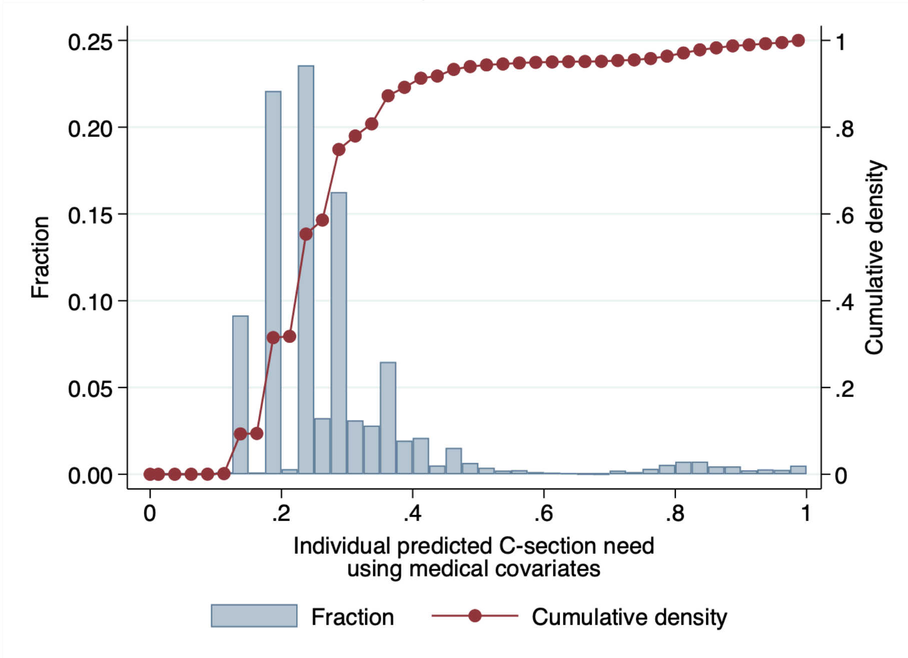
Figure A.29: Correlation of County C-Section Rates and Maternal Morbidity (Adjusted for Medical Covariates Only)



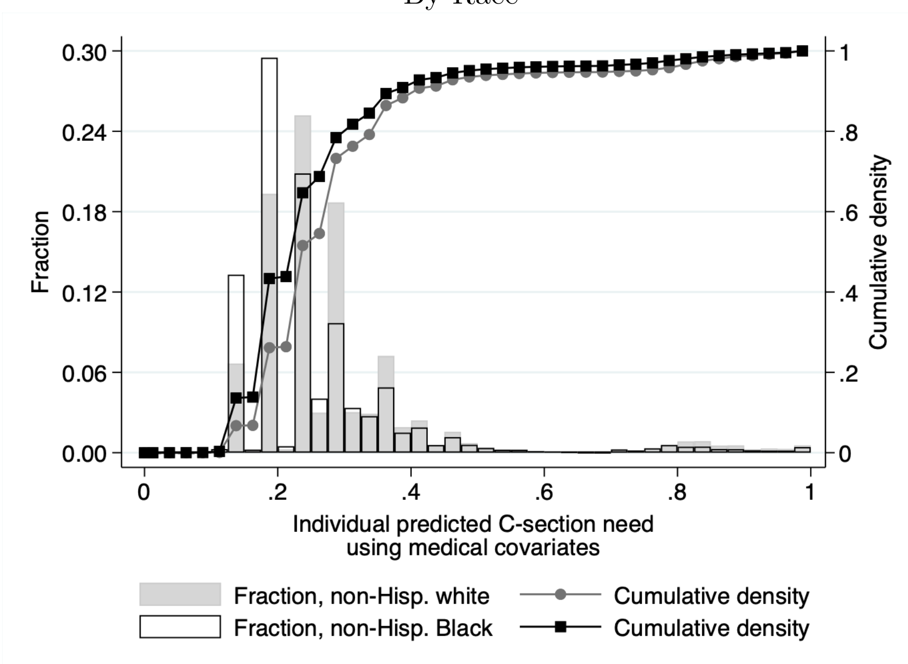
Notes: This figure shows binscatter plots of C-section rates for high-risk singleton first births with maternal or infant morbidity. Linear fit and p-value are based on the underlying counties (prior to binning). Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. C-section and morbidity rates (adjusted for medical covariates) are averaged over the three-year period, weighted by the number of high-risk singleton first births in each year. Binscatter and linear fit are weighted by the number of high-risk singleton first births in the county over the three years.

Figure A.30: Individual Predicted C-Section Need Overall and by Race 2015-2017 (Based on Medical Covariates Only)

Overall



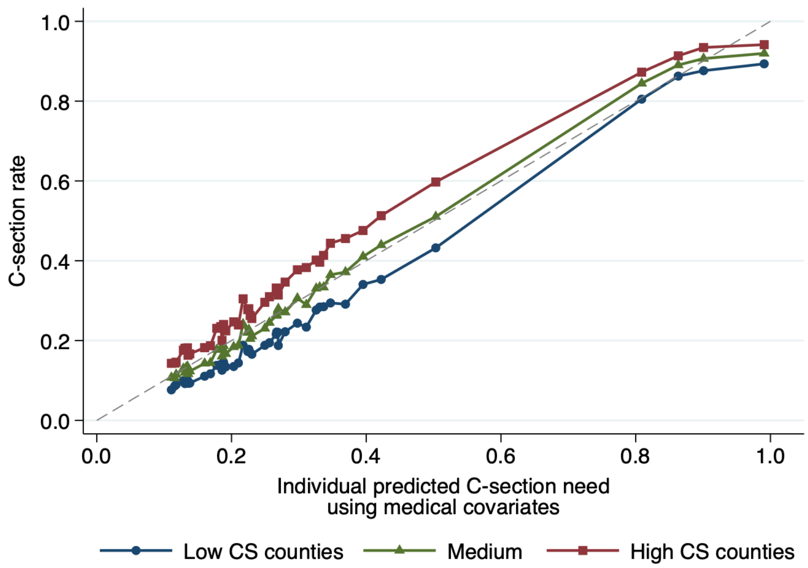
By Race



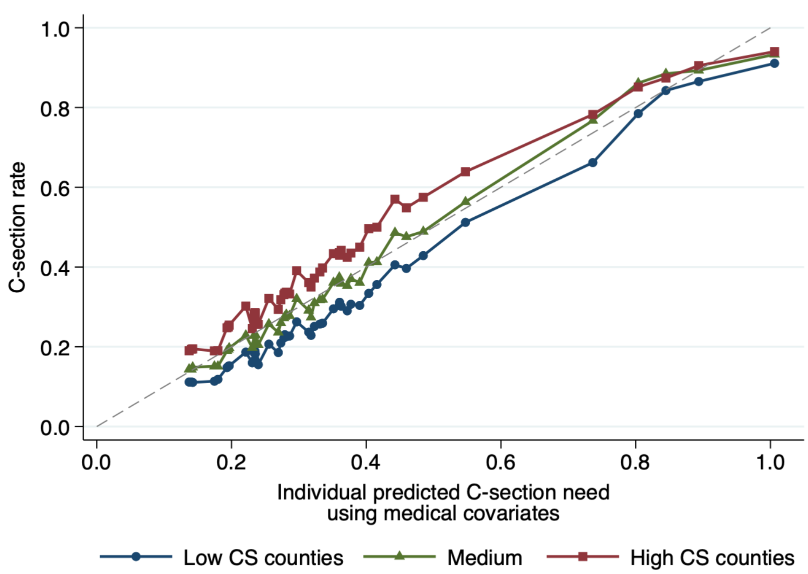
Notes: These figures show the distribution of the individual predicted C-section need for singleton first births using medical covariates (after controlling for county fixed effects).

Figure A.31: C-Section Rates by Predicted C-Section Need and County C-Section Rate
(Based on Medical Covariates Only)

1989-1991



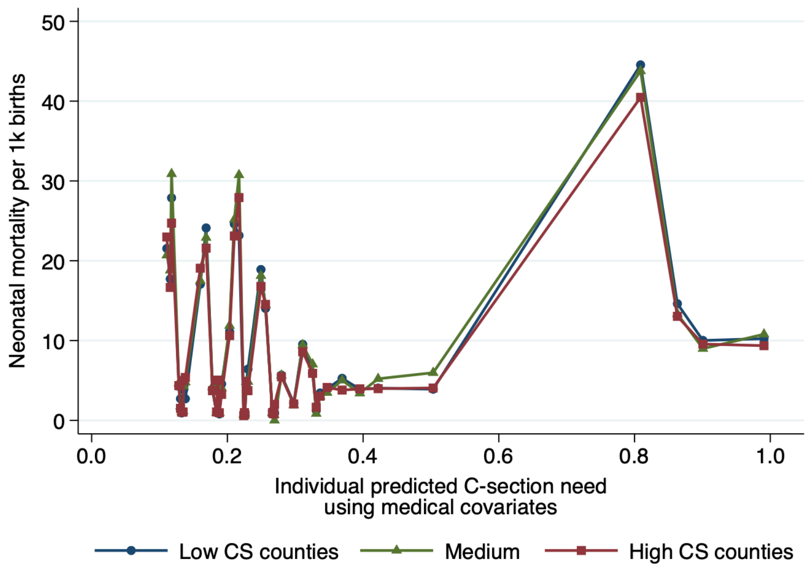
2015-2017



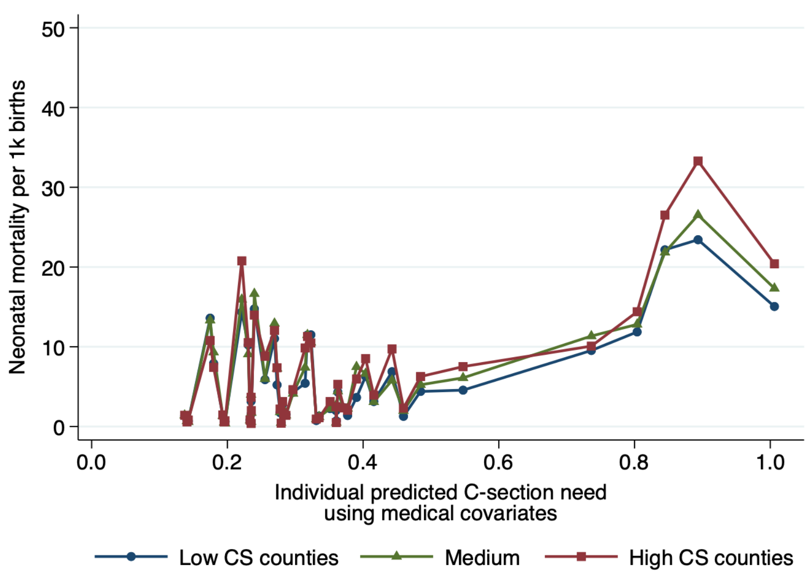
Notes: This figure shows (raw) C-section rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section need is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects.

Figure A.32: Neonatal Mortality by Predicted C-Section Need
and County C-Section Rate
(Based on Medical Covariates Only)

1989-1991

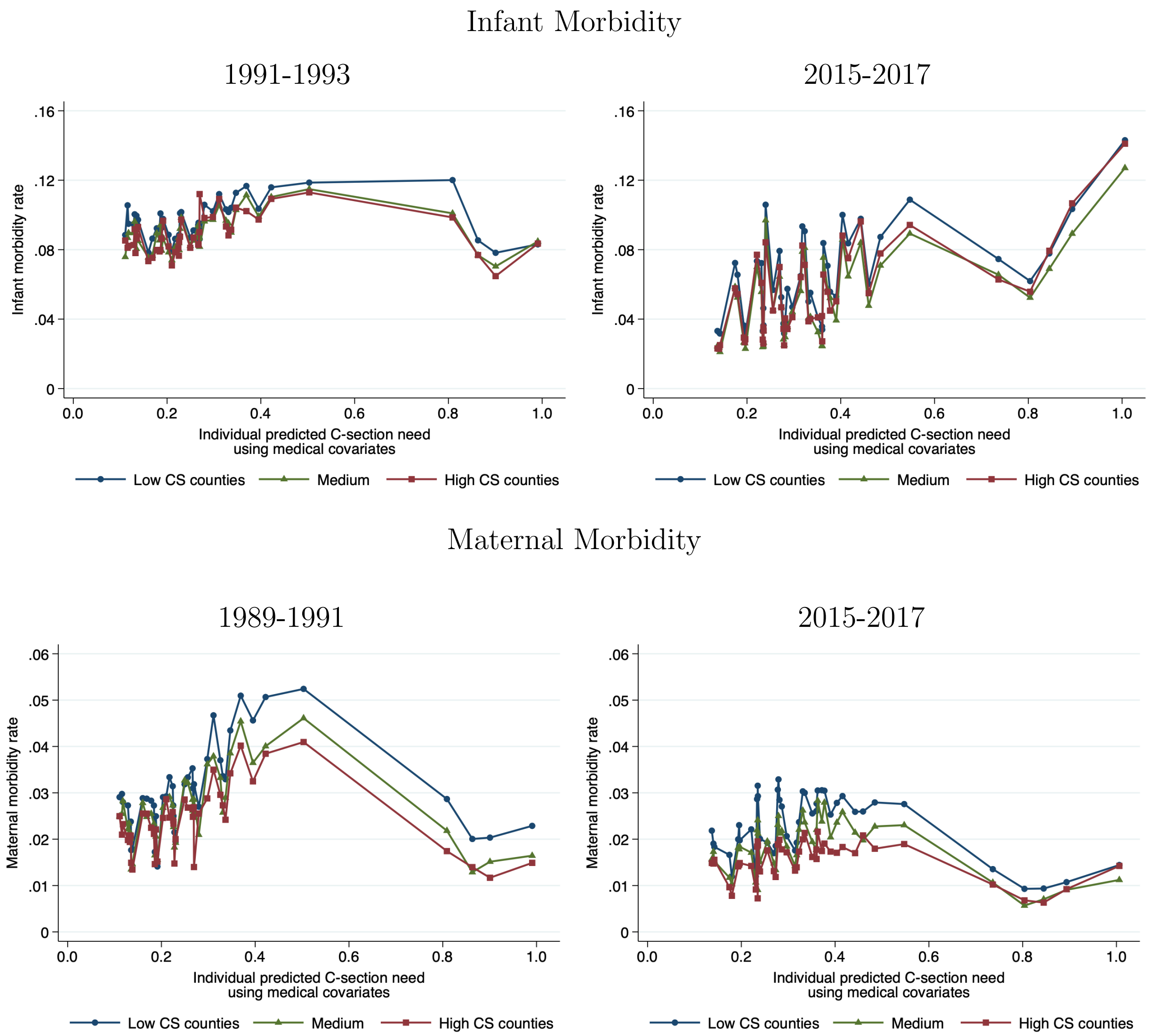


2015-2017



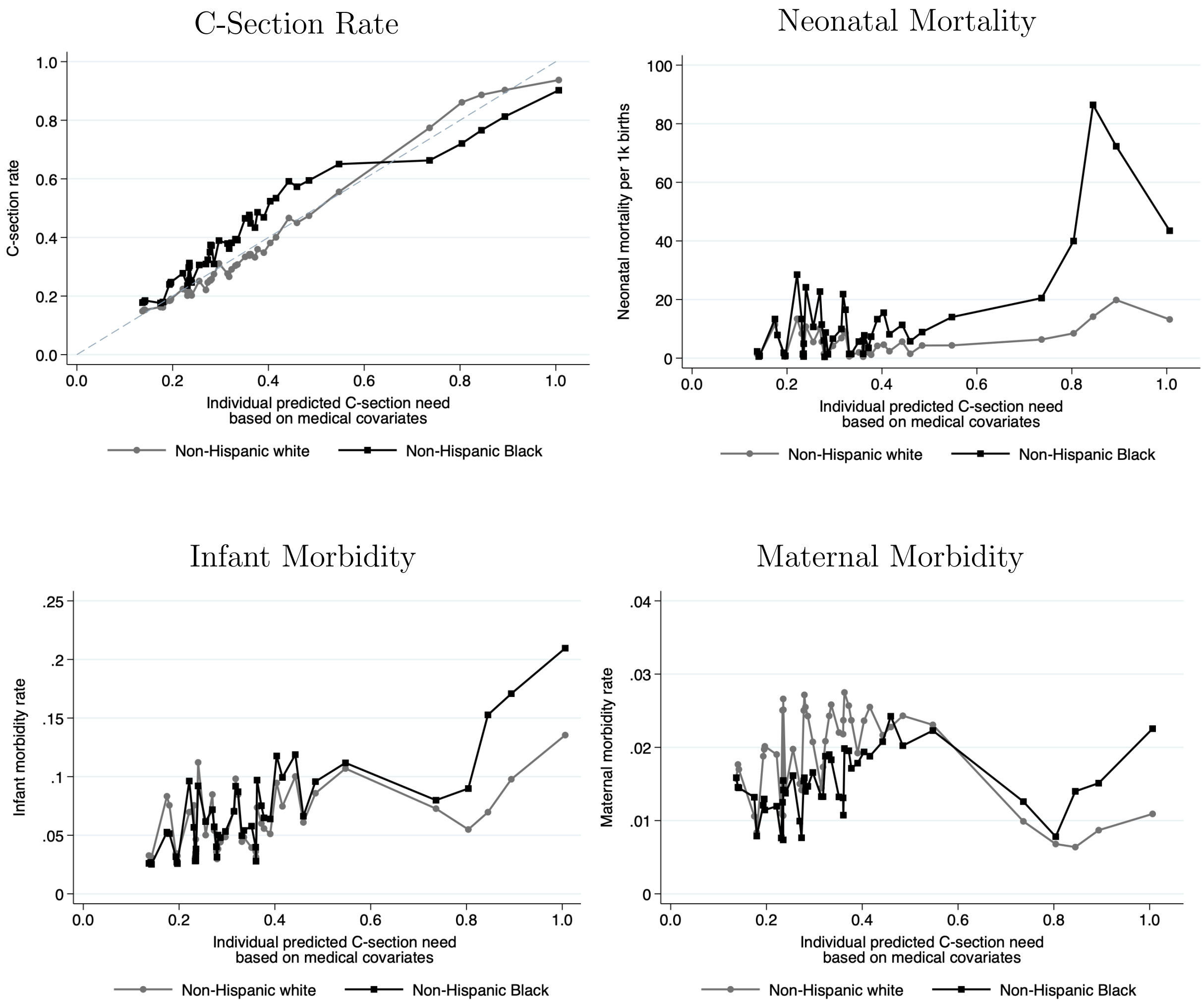
Notes: This figure shows (raw) neonatal mortality rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section need is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects.

Figure A.33: Maternal and Infant Morbidity by Predicted C-Section Need and County C-Section Rate
(Based on Medical Covariates Only)



Notes: This figure shows (raw) maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births that take place in counties with low, medium, and high adjusted C-section rates (terciles weighted by number of singleton first births). Predicted C-section need is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation.

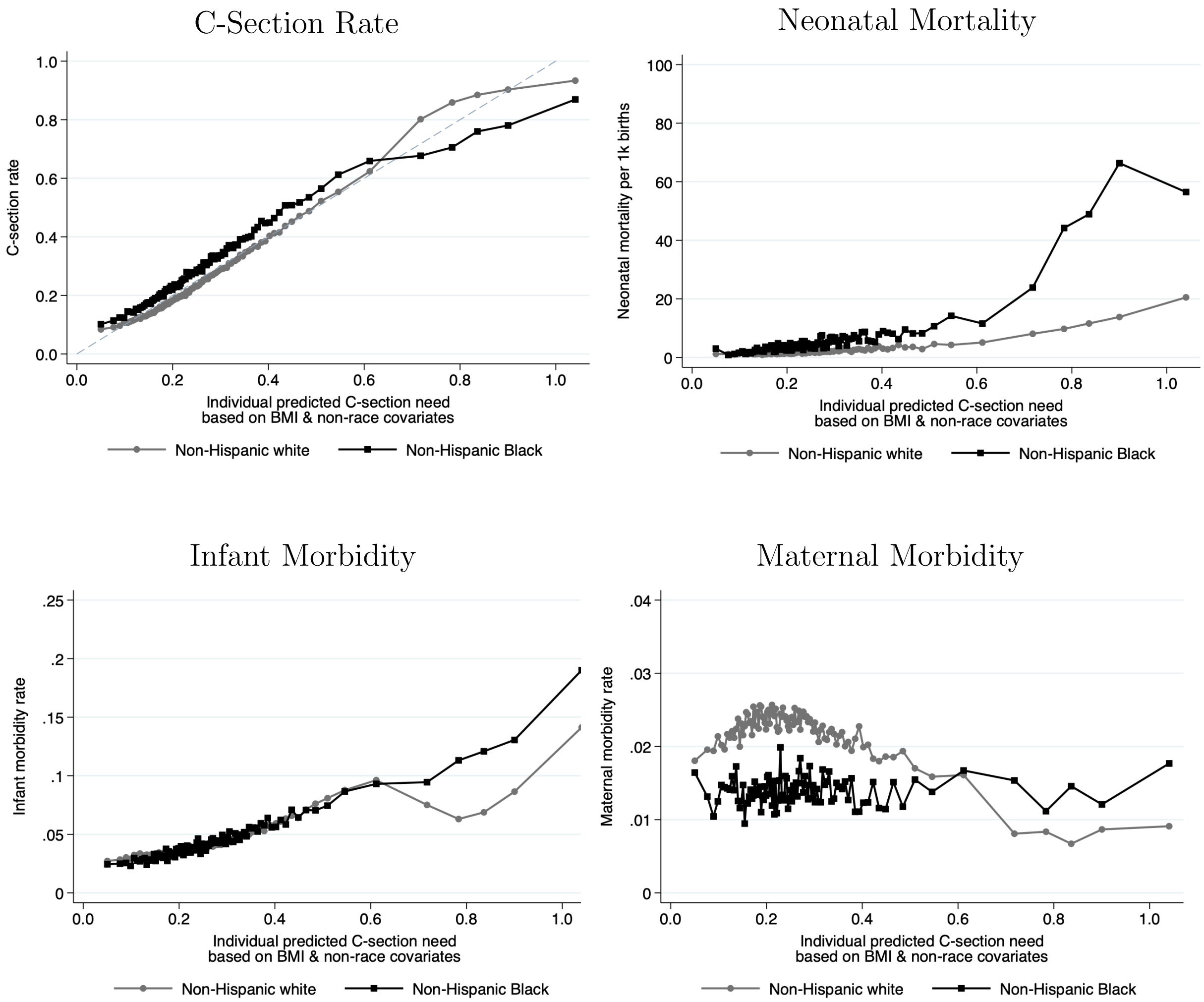
Figure A.34: C-Section Rates, Neonatal Mortality, and Maternal and Infant Morbidity by Predicted C-Section Need and Race, 2015-2017
(Based on Medical Covariates Only)



Notes: This figure shows (raw) c-section, neonatal mortality, and maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births with non-Hispanic white mothers and non-Hispanic Black mothers. Predicted C-section need is derived from a regression based on medical covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation.

A.3.4 Using Body Mass Index

Figure A.35: C-Section Rates, Neonatal Mortality, and Maternal and Infant Morbidity by Predicted C-Section Need and Race, 2015-2017
(Based on Non-Race plus BMI Covariates)



Notes: This figure shows (raw) c-section, neonatal mortality, and maternal and infant morbidity rates for singleton first births in each percentile of predicted C-section need. This is shown separately for births with non-Hispanic white mothers and non-Hispanic Black mothers. Predicted C-section need is derived from a regression based on non-race plus BMI covariates and county fixed effects, but the prediction excludes the county fixed effects. Maternal morbidity is the presence of any of the following: febrile, excessive bleeding, seizure, transfusion, perineal laceration, ruptured uterus, unplanned hysterectomy, admission to ICU. Infant morbidity is the presence of any of the following: meconium, injury, seizure, ventilation. Note each marker in the figure represents a percentile (i.e., there are 100 markers for each curve).

Bibliography

Abaluck, Jason, Leila Agha, Chris Kabrhel, Ali Raja, Arjun Venkatesh, et al. 2016. "The Determinants of Productivity in Medical Testing: Intensity and Allocation of Care." *American Economic Review*, 106(12): 3730–3764.

Abaluck, Jason, Leila Agha, David C Chan Jr, Daniel Singer, and Diana Zhu. 2020. "Fixing misallocation with guidelines: Awareness vs. adherence." National Bureau of Economic Research.

AHRQ. 2021. "MEPS-IC Data Tools."

Aizenman, Joshua, Yothin Jinjarak, Hien Nguyen, and Ilan Noy. 2021. "The Political Economy of the COVID-19 Fiscal Stimulus Packages of 2020." NBER Working Paper 29360.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics*, 84(3): 488–500.

Alesina, Alberto, Nouriel Roubini, and Gerald D. Cohen. 1997. *Political Cycles and the Macroeconomy.* MIT Press.

Almond, Douglas, and Janet Currie. 2011. "Killing me softly: The fetal origins hypothesis." *Journal of Economic Perspectives*, 25(3): 153–72.

Alt, James E., and Robert C. Lowry. 1994. "Divided government, fiscal institutions, and budget deficits: Evidence from the states." *American Political Science Review*, 811–828.

Anand, Priyanka. 2017. "Health Insurance Costs and Employee Compensation: Evidence from the National Compensation Survey." *Health Economics*, 26(12): 1601–1616.

Ansolabehere, Stephen, John M. De Figueiredo, and James M. Snyder Jr. 2003. "Why is there so little money in US politics?" *Journal of Economic Perspectives*, 17(1): 105–130.

Asatryan, Zareh, Thushyanthan Baskaran, and Friedrich Heinemann. 2017. "The effect of direct democracy on the level and structure of local taxes." *Regional Science and Urban Economics*, 65: 38–55.

Asatryan, Zareh, Thushyanthan Baskaran, Theocharis Grigoriadis, and Friedrich Heinemann. 2017. "Direct Democracy and Local Public Finances under Cooperative Federalism." *The Scandinavian Journal of Economics*, 119(3): 801–820. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjoe.12169.

Ashworth, John, Benny Geys, and Bruno Heyndels. 2006. "Determinants of tax innovation: The case of environmental taxes in Flemish municipalities." *European Journal of Political Economy*, 22(1): 223–247.

Baicker, Katherine. 2005. "The spillover effects of state spending." *Journal of Public Economics*, 89(2): 529–544.

Baicker, Katherine, and Amitabh Chandra. 2006. "The Labor Market Effects of Rising Health Insurance Premiums." *Journal of Labor Economics*, 24(3): 609–634.

Baicker, Katherine, Jeffrey Clemens, and Monica Singhal. 2012. "The rise of the states: US fiscal decentralization in the postwar period." *Journal of Public Economics*, 96(11-12): 1079–1091.

Baicker, Katherine, Kasey S Buckles, and Amitabh Chandra. 2006. "Geographic variation in the appropriate use of cesarean delivery: do higher usage rates reflect medically inappropriate use of this procedure?" *Health Affairs*, 25(Suppl1): W355–W367.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *The Quarterly Journal of Economics*, 130(4): 1623–1667.

Bailey, James, and Nathan Blascak. 2016. "The effect of state health insurance benefit mandates on premiums and employee contributions." *Applied Economics Letters*, 23(14): 1042–1046.

Baker, Scott R., Nicholas Bloom, Brandice Canes-Wrone, Steven J. Davis, and Jonathan Rodden. 2014. "Why Has US Policy Uncertainty Risen since 1960?" *American Economic Review*, 104(5): 56–60.

Baker, Scott R., Pawel Janas, and Lorenz Kueng. 2020. "Correlation in State and Local Tax Changes." *Available at SSRN 3648513*.

Battaglia, Emily. 2022. "The effect of hospital closures on maternal and infant health."

Bernecker, Andreas. 2016. "Divided we reform? Evidence from US welfare policies." *Journal of Public Economics*, 142: 24–38.

Besley, Timothy, and Anne Case. 1995*a*. "Does Electoral Accountability Affect Economic Policy Choices? Evidence from Gubernatorial Term Limits." *The Quarterly Journal of Economics*, 110(3): 769–798.

Besley, Timothy, and Anne Case. 1995*b*. "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition." *The American Economic Review*, 85(1): 25–45. Publisher: American Economic Association.

Besley, Timothy J., and Harvey S. Rosen. 1998. "Vertical externalities in tax setting: evidence from gasoline and cigarettes." *Journal of Public Economics*, 70(3): 383–398.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Option." *The Quarterly Journal of Economics*, 132(3): 1319–1372.

Bierbrauer, Felix, Pierre Boyer, Andrew Lonsdale, and Andreas Peichl. 2021. "Tax Reforms and Political Feasibility."

Binder, Sarah A. 2004. "Stalemate: Causes and Consequences of Legislative Gridlock." Brookings Institution Press.

BLS. 2020. "Medical care premiums in the United States: Fact Sheet."

Boadway, Robin, and Jean-François Tremblay. 2012. "Reassessment of the Tiebout model." *Journal of Public Economics*, 96(11-12): 1063–1078.

Bordignon, Massimo, Floriana Cerniglia, and Federico Revelli. 2003. "In search of yardstick competition: a spatial analysis of Italian municipality property tax setting." *Journal of Urban Economics*, 54(2): 199–217.

Bozzano, Monica, Paola Profeta, Riccardo Puglisi, and Simona Scabrosetti. 2021. "Women' Voice on Redistribution: From Gender Norms to Taxation." Social Science Research Network SSRN Scholarly Paper ID 3819099, Rochester, NY.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics*, 132(3): 1261–1318.

Buettner, Thiess. 2003. "Tax base effects and fiscal externalities of local capital taxation: evidence from a panel of German jurisdictions." *Journal of Urban Economics*, 54(1): 110–128.

Bundorf, M. Kate. 2002. "Employee demand for health insurance and employer health plan choices." *Journal of Health Economics*, 21(1): 65–88.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Card, David, Alessandra Fenizia, and David Silver. 2023. "The Health Impacts of Hospital Delivery Practices." *American Economic Journal: Economic Policy*, 15(2): 42–81.

Case, Anne C., Harvey S. Rosen, and James R. Hines. 1993. "Budget spillovers and fiscal policy interdependence: Evidence from the states." *Journal of Public Economics*, 52(3): 285–307.

Castanheira, Micael, Gaëtan Nicodème, and Paola Profeta. 2012. "On the political economics of tax reforms: survey and empirical assessment." *International Tax and Public Finance*, 19(4): 598–624.

Cebul, Randall D., James B. Rebitzer, Lowell J. Taylor, and Mark E. Votruba. 2011. "Unhealthy Insurance Markets: Search Frictions and the Cost and Quality of Health Insurance." *American Economic Review*, 101(5): 1842–1871.

Chan, David C, Matthew Gentzkow, and Chuan Yu. 2022. "Selection with variation in diagnostic skill: Evidence from radiologists." *The Quarterly Journal of Economics*, 137(2): 729–783.

Chandra, Amitabh, and Douglas O Staiger. 2007. "Productivity spillovers in health care: evidence from the treatment of heart attacks." *Journal of Political Economy*, 115(1): 103–140.

Chandra, Amitabh, and Douglas O Staiger. 2020. "Identifying Sources of Inefficiency in Healthcare." *The Quarterly Journal of Economics*, 135(2): 785–843.

Chandra, Amitabh, Evan Flack, and Ziad Obermeyer. 2021. "The Health Costs of Cost-Sharing." NBER Working Paper 28439.

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight. 2010. "Patient Cost-Sharing and Hospitalization Offsets in the Elderly." *American Economic Review*, 100(1): 193–213.

Chang, Andrew C., Linda Cohen, Amihai Glazer, and Urbashee Paul. 2020. "Politicians avoid tax increases around elections." *Available at SSRN*.

Chirinko, Robert S., and Daniel J. Wilson. 2017. "Tax competition among U.S. states: Racing to the bottom or riding on a seesaw?" *Journal of Public Economics*, 155: 147–163.

Choudhry, Niteesh K., Jerry Avorn, Robert J. Glynn, Elliott M. Antman, Sebastian Schneeweiss, Michele Toscano, Lonny Reisman, Joaquim Fernandes, Claire Spettell, Joy L. Lee, Raisa Levin, Troyen Brennan, and William H. Shrank. 2011. "Full Coverage for Preventive Medications after Myocardial Infarction." *New England Journal of Medicine*, 365(22): 2088–2097.

Chu, Susan Y, SY Kim, Christopher H Schmid, Patricia M Dietz, WM Callaghan, J Lau, and KM Curtis. 2007. "Maternal obesity and risk of cesarean delivery: a meta-analysis." *Obesity reviews*, 8(5): 385–394.

Clemens, Jeffrey, and David M. Cutler. 2014. "Who pays for public employee health costs?" *Journal of Health Economics*, 38: 65–76.

CMS. 2020. "HHS Notice of Benefit and Payment Parameters for 2021 Final Rule Fact Sheet."

Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The diffusion of an innovation among physicians." *Sociometry*, 20(4): 253–270.

Congressional Research Service. 2020. "Unfunded Mandates Reform Act: History, Impact, and Issues." CRS Report R40957.

Corry, Maureen P, Suzanne F Delbanco, and Harold D Miller. 2013. "The cost of having a baby in the United States." *Truven Health Analytics, Greenwood Village, CO, USA*.

Currie, Janet, and W Bentley MacLeod. 2017. "Diagnosing expertise: Human capital, decision making, and performance among physicians." *Journal of Labor Economics*, 35(1): 1–43.

Currie, Janet, W Bentley MacLeod, and Jessica Van Parys. 2016. "Provider practice style and patient health outcomes: The case of heart attacks." *Journal of Health Economics*, 47: 64–80.

Cutler, David, Angus Deaton, and Adriana Lleras-Muney. 2006. "The determinants of mortality." *Journal of Economic Perspectives*, 20(3): 97–120.

Cutler, David M, and Ellen Meara. 2000. "The technology of birth: Is it worth it?" Vol. 3, De Gruyter.

Dalton, Christina M., and Sara B. Holland. 2019. "Why Do Firms Use Insurance to Fund Worker Health Benefits? The Role of Corporate Finance." *Journal of Risk and Insurance*, 86(1): 183–212.

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

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2022. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." SSRN Working Paper 3731856.

Deryugina, Tatyana, and David Molitor. 2020. "Does when you die depend on where you live? Evidence from Hurricane Katrina." *American Economic Review*, 110(11): 3602–3633.

Deryugina, Tatyana, and David Molitor. 2021. "The causal effects of place on health and longevity." *Journal of Economic Perspectives*, 35(4): 147–70.

Deskins, John, and Brian Hill. 2010. "Have State Tax Interdependencies Changed Over Time?" *Public Finance Review*, 38(2): 244–270. Publisher: SAGE Publications Inc.

Devereux, Michael P., Ben Lockwood, and Michela Redoano. 2007. "Horizontal and vertical indirect tax competition: Theory and some evidence from the USA." *Journal of Public Economics*, 91(3-4): 451–479.

Dillender, Marcus, Carolyn J. Heinrich, and Susan Houseman. 2022. "Effects of the Affordable Care Act on Part-Time Employment Early Evidence." *Journal of Human Resources*, 57(4): 1394–1423.

DOL. 2021. "Annual Report on Self-Insured Group Health Plans."

Egger, Peter, and Marko Koethenbuerger. 2010. "Government Spending and Legislative Organization: Quasi-experimental Evidence from Germany." *American Economic Journal: Applied Economics*, 2(4): 200–212.

Einav, Liran, and Amy Finkelstein. 2011. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives*, 25(1): 115–138.

Epstein, Andrew J, and Sean Nicholson. 2009. "The formation and evolution of physician treatment styles: an application to cesarean sections." *Journal of health economics*, 28(6): 1126–1140.

Ericson, Keith Marzilli, and Justin Sydnor. 2017. "The Questionable Value of Having a Choice of Levels of Health Insurance Coverage." *Journal of Economic Perspectives*, 31(4): 51–72.

Erler, H. Abbie. 2007. "Legislative term limits and state spending." *Public Choice*, 133(3-4): 479–494.

Fang, Hanming, and Alessandro Gavazza. 2011. "Dynamic Inefficiencies in an Employment-Based Health Insurance System: Theory and Evidence." *American Economic Review*, 101(7): 3047–3077.

Ferede, Ergete, Bev Dahlby, and Ebenezer Adjei. 2015. "Determinants of statutory tax rate changes by the Canadian provinces." *Economics of Governance*, 16(1): 27–51.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of geographic variation in health care: Evidence from patient migration." *The Quarterly Journal of Economics*, 131(4): 1681–1726.

Fischer, Stefanie J, Heather Royer, and Corey D White. 2023. "Health Care Centralization: The Health Impacts of Obstetric Unit Closures in the US." *NBER Working Paper No. 30141.*

Foremny, Dirk, and Nadine Riedel. 2014. "Business taxes and the electoral cycle." *Journal of Public Economics*, 115: 48–61.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech." *Econometrica*, 87(4): 1307–1340. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA16566.

Geys, Benny. 2006. "Looking across borders: A test of spatial policy interdependence using local government efficiency ratings." *Journal of Urban Economics*, 60(3): 443–462.

Gilligan, Thomas W., and John G. Matsusaka. 2001. "Fiscal policy, legislature size, and political parties: Evidence from state and local governments in the first half of the 20th century." *National Tax Journal*, 57–82.

Glover, J Alison. 1938. "The incidence of tonsillectomy in school children."

Goldstein, Gerald S., and Mark V. Pauly. 1976. "Group Health Insurance as a Local Public Good." In *The Role of Health Insurance in the Health Services Sector*. 73–114. NBER.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

Goodspeed, Timothy J. 2000. "Tax structure in a federation." *Journal of Public Economics*, 75(3): 493–506.

Goodspeed, Timothy J. 2002. "Tax competition and tax structure in open federal economies: Evidence from OECD countries with implications for the European Union." *European Economic Review*, 46(2): 357–374.

Griffith, Kevin N, Yevgeniy Feyman, Samantha G Auty, Erika L Crable, and Timothy W Levengood. 2021. "County-level data on US opioid distributions, demographics, healthcare supply, and healthcare access." *Data in Brief*, 35: 106779.

Gross, Tal, Timothy J. Layton, and Daniel Prinz. 2022. "The Liquidity Sensitivity of Healthcare Consumption: Evidence from Social Security Payments." *American Economic Review: Insights*, 4(2): 175–190.

Gruber, Jonathan. 1994*a*. "The Incidence of Mandated Maternity Benefits." *The American Economic Review*, 84(3): 622–641.

Gruber, Jonathan. 1994*b*. "State-mandated benefits and employer-provided health insurance." *Journal of Public Economics*, 55(3): 433–464.

Gupta, Sanjeev, and João Tovar Jalles. 2020. "On the political economy determinants of tax reforms: evidence from developing countries." Mimeo.

Hall, Mark A., and Michael J. McCue. 2021. "Expanding Consumer Health Insurance Options by Easing the Boundaries Between Individual and Small-Group Markets." Commonwealth Fund.

Hannah, Mary E, Walter J Hannah, Sheila A Hewson, Ellen D Hodnett, Saroj Saigal, Andrew R Willan, and Term Breech Trial Collaborative. 2000. "Planned caesarean section versus planned vaginal birth for breech presentation at term: a randomised multicentre trial." *The Lancet*, 356(9239): 1375–1383.

Hyde, Matthew J, Alison Mostyn, Neena Modi, and Paul R Kemp. 2012. "The health implications of birth by Caesarean section." *Biological Reviews*, 87(1): 229–243.

Inman, Robert P., and Michael A. Fitts. 1990. "Political Institutions and Fiscal Policy: Evidence from the U.S. Historical Record." *Journal of Law, Economics, and Organization*, 6(Special Issue): 79–132.

Jensen, Gail A., and Jon R. Gabel. 1992. "State Mandated Benefits and the Small Firm's Decision to Offer Insurance." *Journal of Regulatory Economics*, 4(4): 379–404.

Jensen, Gail A., and Michael A. Morrisey. 1999. "Employer-Sponsored Health Insurance and Mandated Benefit Laws." *The Milbank Quarterly*, 77(4): 425–459.

Jensen, Gail A., Kevin D. Cotter, and Michael A. Morrisey. 1995. "State Insurance Regulation and Employers' Decisions to Self-Insure." *The Journal of Risk and Insurance*, 62(2): 185–213.

Johnson, Erin M, and M Marit Rehavi. 2016. "Physicians treating physicians: Information and incentives in childbirth." *American Economic Journal: Economic Policy*, 8(1): 115–41.

Kakpo, Eliakim. 2019. "On the Political Economy of State Corporate Tax Reforms in the U.S." *Research in Applied Economics*, 11(3): 25–48. Number: 3.

Katsimi, Margarita, and Vassilis Sarantides. 2012. "Do elections affect the composition of fiscal policy in developed, established democracies?" *Public Choice*, 151(1): 325–362.

Keag, Oonagh E, Jane E Norman, and Sarah J Stock. 2018. "Long-term risks and benefits associated with cesarean delivery for mother, baby, and subsequent pregnancies: Systematic review and meta-analysis." *PLoS medicine*, 15(1): e1002494.

Kennedy-Moulton, Kate, Sarah Miller, Petra Persson, Maya Rossin-Slater, Laura Wherry, and Gloria Aldana. 2022. "Maternal and Infant Health Inequality: New Evidence from Linked Administrative Data." *National Bureau of Economic Research Working Paper No. 30693*.

KFF. 2019. "Health Insurance Coverage of the Total Population."

KFF. 2021. "Employee Health Benefits Annual Survey."

Kling, J. R., S. Mullainathan, E. Shafir, L. C. Vermeulen, and M. V. Wrobel. 2012. "Comparison Friction: Experimental Evidence from Medicare Drug Plans." *The Quarterly Journal of Economics*, 127(1): 199–235.

Laudicina, Susan, Joan Gardner, and Kim Holland. 2013. "State Legislative Healthcare and Insurance Issues." Blue Cross Blue Shield Association.

Liu, Chenyuan, and Justin Sydnor. 2022. "Dominated Options in Health Insurance Plans." *American Economic Journal: Economic Policy*, 14(1): 277–300.

Lubotsky, Darren, and Craig A. Olson. 2015. "Premium copayments and the trade-off between wages and employer-provided health insurance." *Journal of Health Economics*, 44: 63–79.

MacDorman, Marian F, and Eugene Declercq. 2018. "The failure of United States maternal mortality reporting and its impact on women's lives." *Birth (Berkeley, Calif.)*, 45(2): 105.

Marone, Victoria R., and Adrienne Sabety. 2022. "When Should There Be Vertical Choice in Health Insurance Markets?" *American Economic Review*, 112(1): 304–342.

Mathur, Aparna. 2010. "Health insurance and job creation by the self-employed." *Small Business Economics*, 35(3): 299–317.

Matsusaka, John G. 1995. "Fiscal Effects of the Voter Initiative: Evidence from the Last 30 Years." *Journal of Political Economy*, 103(3): 587–623.

Matsusaka, John G. 2000. "Fiscal effects of the voter initiative in the first half of the twentieth century." *The Journal of Law and Economics*, 43(2): 619–650.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2016. *Polarized America: The dance of ideology and unequal riches.* . Second Edition ed., MIT Press.

McCubbins, Mathew D. 1991. "Party Governance and U. S. Budget Deficits: Divided Government and Fiscal Stalemate." In *Politics and Economics in the Eighties*. 83–122. University of Chicago Press.

McCuskey, Elizabeth Y. 2022. "State Cost-Control Reforms and ERISA Preemption." Commonwealth Fund.

Meiselbach, Mark K., Matthew D. Eisenberg, Ge Bai, Aditi Sen, and Gerard F. Anderson. 2022. "Labor Market Concentration and Worker Contributions to Health Insurance Premiums." *Medical Care Research and Review*, 79(2): 198–206.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2014. "Resolving Debt Overhang: Political Constraints in the Aftermath of Financial Crises." *American Economic Journal: Macroeconomics*, 6(2): 1–28.

Mikesell, John L. 1978. "Election periods and state tax policy cycles." *Public Choice*, 33(3): 99–106. Publisher: Springer.

Molitor, David. 2018. "The evolution of physician practice styles: evidence from cardiologist migration." *American Economic Journal: Economic Policy*, 10(1): 326–56.

Moran, John R., Michael E. Chernew, and Richard A. Hirth. 2001. "Preference diversity and the breadth of employee health insurance options." *Health Services Research*, 36(5): 911–934.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing physician error: A machine learning approach to low-value health care." *The Quarterly Journal of Economics*, 137(2): 679–727.

Mulligan, Casey B. 2020. "The Employer Penalty, Voluntary Compliance, and the Size Distribution of Firms: Evidence from a Survey of Small Businesses." *Tax Policy and the Economy*, 34: 139–171.

NASBO. 2014. "Capital Budgeting in the States." National Association of State Budget Officers.

NCSL. 2010. "Fiscal Brief: State Balanced Budged Provisions." National Conference of State Legislatures.

Nelson, Michael A. 2000. "Electoral Cycles and the Politics of State Tax Policy." *Public Finance Review*, 28(6): 540–560.

New York State DFS. 2019. "Governor Cuomo Announces New Actions Ensuring Health Insurers Provide Expanded Coverage for HIV-Prevention Drug Prep and Consumers Get Health Insurance Protections Guaranteed Under The Law."

Office of Legislative Research. 2019. "Health Insurance Benefit Mandate Reimbursement Process."

Ortaliza, Jared, Matthew McGough, Emma Wager, Gary Claxton, and Krutika Amin. 2021. "How do health expenditures vary across the population?" Peterson-KFF Health System Tracker.

Oster, Emily. 2018. "Expert behavior change in response to best practice changes: Evidence from obstetrics."

Park, Christina H. 2000. "Prevalence of Employer Self-Insured Health Benefits: National and State Variation." *Medical Care Research and Review*, 57(3): 340–360.

Podulka, Jennifer, Elizabeth Stranges, and Claudia Steiner. 2011. "Hospitalizations related to childbirth, 2008." *Heathcare Cost and Utilization Project Statistical Brief 110*.

Poterba, James M. 1994. "State Responses to Fiscal Crises: The Effects of Budgetary Institutions and Politics." *Journal of Political Economy*, 102(4): 799–821.

Rhode, Paul W., and Koleman S. Strumpf. 2003. "Assessing the Importance of Tiebout Sorting: Local Heterogeneity from 1850 to 1990." *American Economic Review*, 93(5): 1648–1677.

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

Rork, Jonathan C. 2003. "Coveting Thy Neighbors' Taxation." *National Tax Journal*, 56(4): 775–787.

Rosenberg, Jacob. 1992. "Rationality and the political business cycle: The case of local government." *Public Choice*, 73(1): 71–81.

Rose, Shanna. 2006. "Do fiscal rules dampen the political business cycle?" *Public choice*, 128(3-4): 407–431. Publisher: Springer.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics*, 90(4): 629.

Roubini, Nouriel, and Jeffrey D. Sachs. 1989. "Political and economic determinants of budget deficits in the industrial democracies." *European Economic Review*, 33(5): 903–933.

Seegert, Nathan. 2016. "Tax Revenue Volatility." *Available at SSRN 2789889*.

Shah, Neel. 2017. "Keynote to American Ob-Gyns: C-section rates linked to hospital complexities." https://www.ariadnelabs.org/resources/articles/keynote-to-american-ob-gyns-c-section-rates-linked-to-hospital-complexities/, [Accessed 30-May-2023].

Silver, David. 2021. "Haste or Waste? Peer pressure and productivity in the emergency department." *The Review of Economic Studies*, 88(3): 1385–1417.

Singh, Manasvini. 2021. "Heuristics in the delivery room." *Science*, 374(6565): 324–329.

Skinner, Jonathan. 2011. "Causes and consequences of regional variations in health care." In *Handbook of Health Economics*. Vol. 2, 45–93. Elsevier.

Slattery, Cailin, Alisa Tazhitdinova, and Sarah Robinson. 2023. "Corporate political spending and state tax policy: Evidence from Citizens United." *Journal of Public Economics*, 221: 104859.

Slattery, Cailin, and Owen Zidar. 2020. "Evaluating state and local business incentives." *Journal of Economic Perspectives*, 34(2): 90–118.

Sloan, Frank A., and Christopher J. Conover. 1998. "Effects of State Reforms on Health Insurance Coverage of Adults." *Inquiry*, 35(3): 280–293.

Sommers, Benjamin D. 2005. "Who Really Pays for Health Insurance? The Incidence of Employer-Provided Health Insurance with Sticky Nominal Wages." *International Journal of Health Care Finance and Economics*, 5(1): 89–118.

Suarez Serrato, Juan Carlos, and Owen Zidar. 2018. "The structure of state corporate taxation and its impact on state tax revenues and economic activity." *Journal of Public Economics*, 167: 158–176.

Summers, Lawrence H. 1989. "Some Simple Economics of Mandated Benefits." *The American Economic Review Papers & Proceedings*, 79(2): 177–183.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

Tazhitdinova, Alisa. 2020. "Are changes of organizational form costly? Income shifting and business entry responses to taxes." *Journal of Public Economics*, 186: 104187.

Tiebout, Charles M. 1956. "A pure theory of local expenditures." *Journal of Political Economy*, 64(5): 416–424.

U.S. Census Bureau. 2019. "Statistics of U.S. Businesses: Annual Data Tables by Establishment Industry."

Valdes, Elise G. 2021. "Examining cesarean delivery rates by race: a population-based analysis using the Robson Ten-Group Classification System." *Journal of Racial and Ethnic Health Disparities*, 8(4): 844–851.

Veiga, Linda Gonçalves, and Francisco José Veiga. 2007. "Political business cycles at the municipal level." *Public Choice*, 131(1): 45–64.

Wennberg, John, and Alan Gittelsohn. 1973. "Small area variations in health care delivery: a population-based health information system can guide planning and regulatory decision-making." *Science*, 182(4117): 1102–1108.

Yang, Yanchao, and Maggie Mullen. 2022. "Black-White Disparities of Cesarean Section Rate in the US: Trends and Determinants." *Available at SSRN 4204876.*

Zucman, Gabriel. 2014. "Taxing across Borders: Tracking Personal Wealth and Corporate Profits." *Journal of Economic Perspectives*, 28(4): 121–148.