Title

Essays on Public Economics

Author
Lobel, Felipe

Publication Date
2024

Peer reviewed|Thesis/dissertation

Essays on Public Economics

by

Felipe Lobel

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Emmanuel Saez, Co-chair
Professor Patrick Kline, Co-chair
Professor Alan Auerbach
Professor Benjamin Schoefer
Professor Gabriel Zucman

Spring 2024

Essays on Public Economics

Abstract

Essays on Public Economics

by

Felipe Lobel

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Emmanuel Saez, Co-chair

Professor Patrick Kline, Co-chair

This dissertation addresses empirically, theoretically and experimentally issues of business taxation. The first chapter studies firms' margins of response to a historically large payroll tax cut that affects a subset of Brazilian firms. Difference-in-differences estimates based on plausibly exogenous legal variation indicate that the payroll tax reduction causes an increase in employment, wages, and profits, while capital decreases. Responses are substantially more pronounced among small firms, and workers' earnings gains are concentrated at the top of the distribution. Reduced-form estimates reveal that consumers pay 65% of payroll taxes, firm owners 23%, and workers 12%. These results establish not only that payroll tax cuts primarily benefit consumers, but also exacerbate within-firm earnings inequality. This evidence cannot be reconciled within a competitive framework. I estimate a model that allows for product and labor market power to explain these findings.

The second chapter, in joint work with Thiago Scot and Pedro Zuniga, we study corporate responses to a minimum income tax, using the universe of corporate tax filings in Honduras. The policy design allows us to separately estimate cost misreporting under profit taxation and the elasticity of reported revenue. Large corporations overreport true costs when taxed on profits. Taxing revenue leads to a substantial decrease in reported revenues: we estimate an elasticity in the range 0.35-1. The elasticity of revenue is attenuated when third-party information on the revenue of firms is available, suggesting misreporting plays an important role. Our results inform trade-offs when broadening tax bases to curb evasion.

The third chapter of this study utilizes an experiment to examine the consequences of violating the stable unit treatment value assumption (SUTVA), a fundamental concern in the empirical evaluation of business tax policy. One significant issue is that changes to the tax

code often generate general equilibrium effects, potentially biasing empirical estimates of tax responses. To study this risk, I collaborated with David Holtz, Ruben Lobel, Inessa Liskovich, and Sinan Aral to design an experiment specifically aimed at quantifying the biases introduced by general equilibrium spillovers. Cluster randomization, i.e., the practice of randomizing treatment assignment at the level of "clusters" of similar individuals, is an established experiment design technique for countering interference bias in social networks, but it is unclear ex ante if it will be effective in marketplace settings. In this paper, we use a meta-experiment or "experiment over experiments" conducted on Airbnb to both provide empirical evidence of interference bias in online market settings and assess the viability of cluster randomization as a tool for reducing interference bias in marketplace TATE estimates. Results from our meta-experiment indicate that at least 19.76% of the TATE estimate produced by an individual-randomized evaluation of the platform fee increase we study is attributable to interference bias and eliminated through the use of cluster randomization. We also find suggestive, non-statistically significant evidence that interference bias in seller-side experiments is more severe in demand-constrained markets, and that the efficacy of cluster randomization at reducing interference bias increases with cluster quality.

To my parents Magda & Léo, and my fiancée Sam.

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply grateful to my advisor, Emmanuel Saez. Long before I arrived at Berkeley, I had admired Emmanuel's work for its scientific rigor, both theoretical and empirical, in addressing critical societal issues. I still vividly remember attending a Berkeley seminar for the first time and seeing him in person. I was profoundly struck by being in close proximity to one of the economists who has most influenced me. At that time, I could not have imagined that I would have the privilege of receiving close mentorship from Emmanuel during my graduate studies. His advice and exemplary brilliance and dedication have been incredibly instructive. He also demonstrates that it is possible to combine high productivity with courtesy and kindness.

I am also indebted to my co-advisor, Patrick Kline, who has significantly shaped my approach to economic research. Pat is among the most talented economists of our time, with a deep understanding of economic models and their connection to data analysis. I am thankful for his detailed feedback on my drafts and for our numerous stimulating discussions on projects at all stages of research. I am extremely grateful and sincerely hope that our paths will continue to intersect in the future.

Alan Auerbach has been an extraordinarily special advisor to me. As one of the founders of much of the business tax literature I study, Alan possesses a sharp and comprehensive understanding of Public Economics. I have been fortunate to receive his mentorship since my first year in the program. His constant support and generosity with his time have been invaluable to me, and I am profoundly thankful.

Benjamin Schoefer and Gabriel Zucman have also played crucial roles during my PhD journey. I greatly value their mentorship, and it is because of excellent advisors like them that Berkeley is the best place to study public and labor economics.

Above all, I am most thankful to my family. My parents, Magda and Léo, have given me unconditional love and worked tirelessly to raise my sister and me. My fiancée, Sam, has been patient and supportive throughout the ups and downs of graduate school. I love them dearly and deeply appreciate their support.

Chapter 1

Who Benefits from Payroll Tax Cuts? Market Power, Tax Incidence and Efficiency

1.1 Introduction

Who benefits from payroll tax cuts? This question has emerged as one of the most important topics in the public discourse as payroll taxes account for 30% of total tax collection, and the adoption of payroll tax cut programs is becoming widespread (OECD, 2019). Traditional Public Finance approaches this issue within a competitive framework, in which the answer arises from properties of aggregate labor demand and supply (Gruber, 1997). This study challenges the traditional view, by providing evidence that product and labor market power are also central in shaping tax incidence.

This paper investigates the implications of payroll tax cuts in the context of Brazil, which implemented a payroll tax reform in 2012. Due to arbitrary sector-specific legal requirements, tax rates were reduced by 20 p.p. for a small subset of firms. Eligible and ineligible firms are not only remarkably similar in levels but most importantly, in pre-reform trends. This resemblance between groups provides a compelling basis for comparison, which I implement in a difference-in-differences specification. To evaluate this policy variation, I rely on novel anonymized administrative tax microdata, which enables the tracking of firms and workers over time, both before and after the reform.

I find that the tax cut caused a 9% employment increase, a phenomenon even more pronounced among small firms. The competitive framework predicts that a firm-specific shock, which does not change workers' outside options, should not affect workers' earnings. However, I find that earnings increased by 2% on average, and by 3% three years after the reform. Even though these effects could, in theory, be influenced by compositional changes, I detect no evidence of such adjustments. These results provide clear evidence of labor market power. Interestingly, most gains are captured by individuals in the top percentiles of the earnings distribution, witnessing gains as high as 14%. This finding underscores that payroll tax cuts exacerbate within-firm earnings inequality.

Consistent with the unequal pass-through within firms, there are significant differences across occupations and races. Specifically, high-skilled workers benefit from a 6% pass-through, while low-skilled workers witness no gains from the same tax cut. I am not able to detect significant differences across gender. All of the earnings increase is concentrated among white workers. While racial disparities are a core concern in the social sciences, to the best of my knowledge this is the first study to empirically assess racial inequality in tax pass-through. The lack of prior evidence stems from the fact that most tax authorities, the US among them, do not record race information.

Given that rich administrative microdata were previously unavailable to researchers, the payroll tax literature has focused on employment and wage responses. This study broadens the analysis by incorporating understudied margins of adjustment such as capital, profits, and revenue. Interpreting the capital response is not straightforward, since substitution and scale forces operate in opposite directions. Consistent with an optimal behavior of substituting toward cheaper inputs, I find that a decrease in labor costs leads to a 3% reduction in capital usage. Likewise, the revenue response is influenced by a quantity increase and a price decrease. I find a 5% revenue rise, which, combined with the scale response identified by the inputs choice, helps to quantify the extent of tax incidence passed onto prices. Profits - a key metric for gauging firms' willingness to pay for a tax reduction - surged by 30% in response to the reform. This empirical result is particularly meaningful, as numerous previous incidence papers do not observe accounting profits and instead rely on structural assumptions (e.g., Suárez Serrato and Zidar, 2016a; Suarez Serrato and Zidar, 2023).

The identifying assumption is that conditional on fixed effects, eligibility is uncorrelated with time-varying unobserved determinants of outcomes. There are two threats to the validity of this design. The first relates to selection on eligibility, i.e., that Congress anticipated sector-specific trends when defining eligibility. I address this concern in several ways. I show not only that pre-trends aren't statistically indistinguishable from zero in any of the outcomes, but also that eligibility is balanced in baseline levels. These results are not surprising per se, as the political process that determined eligibility often assigned remarkably similar sectors to different eligibility statuses, as illustrated by the cases of hotels and motels. Additionally, as a robustness test, I recover determinants of eligibility using a logit model and apply the associated propensity scores in a matching difference-in-differences procedure, which alternatively relies on the conditional independence assumption (CIA). Results from both methods are similar.

The second threat relates to the manipulation of sectoral choice. Identification would be compromised if firms strategically select into eligible sectors after the reform has been announced. To address this concern, I first show in the data that firms rarely change sectors. This is consistent with the fact that to change sectors, a firm must provide extensive supporting documentation to multiple agencies to confirm a proper shift in its core activities. Even in the rare instances in which I do observe sector changes, there is not a trend of switching toward eligible sectors. As an additional robustness check, I restrict the sample to firms that have never changed sectors and the results remain the same.

Although employment increases after the reform, this effect could be driven by mere shifts from existing informal to formal jobs, both within and across firms. This margin of response is particularly relevant in the landscape of developing countries (Ulyssea, 2018a; Haanwinckel and Soares, 2021). Nevertheless, I conducted several tests indicating that informality does not play a major role in response to the payroll tax variation. In one of these tests, I leverage the panel structure of the data to show that the reform does not affect the share of formal new hires transitioning from non-employment or informality. This result is consistent with the fact that the informal sector in Brazil is predominantly characterized by self-employment and is prominently susceptible to fixed costs associated with licensing, legal liabilities, sanitary and security regulations (Maloney, 2004).

To interpret the empirical findings, I develop a simple model in which firms have labor market power, as in Manning, 2011; Card et al., 2018, and product market power as in Hamermesh, 1996; Criscuolo et al., 2019. The interplay between these two competitive frictions, often modeled separately, sheds light on a key aspect: employment and wage pass-through are determined not just by the slope of the labor supply and product demand curves, but also hinge on behavioral responses that guide shifts of the marginal revenue product of labor and product supply. Consistent with the model, I find that both employment and earnings effects are more pronounced in small firms – the ones estimated to have less market power. This pattern, which standard monopsony models in a perfectly competitive product market fail to explain, resonates with a broad range of empirical findings in the context of industrial policies (Bronzini and Iachini, 2014; Howell, 2017; Zwick and Mahon, 2017; Criscuolo et al., 2019).

The model delivers invertible mapping between relevant parameters and reduced form estimates. I estimate the labor supply elasticity faced by the firm ($\epsilon = 4.15$), capital-labor elasticity of substitution ($\sigma_{KL} = 1.72$), and output demand elasticity ($\eta = 1.43$). The labor supply elasticity implies a wage markdown of 0.81, suggesting that Brazilian firms capture 19% of the marginal revenue product of labor. This value aligns closely with estimates from other countries (Card et al., 2018), as well as with findings by Lagos, 2019 in Brazil. The capital-labor elasticity of substitution is similar to Karabarbounis and Neiman, 2014. Lastly, the output demand elasticity reveals the presence of product market power, with an estimated markup of 0.41, which seats toward the upper range of prior estimates, but still between the values found in Harasztosi and Lindner, 2019 and Curtis et al., 2021.

In terms of mechanisms, two-thirds of the employment response is due to firms boosting their scale and the remaining one-third to capital-labor substitution. A full-incidence analysis indicates that consumers pay 65% of payroll taxes, workers 12%, and firm owners 23%. To measure the deadweight loss of payroll taxation, the model connects reduced-form responses to changes in economic surplus and the net fiscal cost. On the margin, an additional dollar in tax cuts leads to a \$0.66 in efficiency gains. This relates to a marginal value of public funds (MVPF) of 1.66, which reflects the high distortionary costs of taxation in developing countries. This estimate falls in the upper range of the 0.5-2 interval reviewed by Hendren and Sprung-Keyser, 2020.

Literature and Contributions. The paper's main contribution is to thoroughly assess firms' margins of response to payroll taxation. The study provides theoretical insights into the role of market power in shaping tax incidence and efficiency. To the best of my knowledge, this is the first paper to incorporate consumers in payroll tax incidence analysis. Although the incidence to consumers is novel to Public Finance, my estimate aligns closely with the minimum wage literature (Harasztosi and Lindner, 2019).

This paper contributes to four strands of literature. First, it builds on a large body of work that finds mixed effects of payroll tax cuts on employment and wages (Gruber, 1997; Saez, Schoefer, and Seim, 2019; A. Kugler, M. Kugler, and Prada, 2017; Cruces, Galiani, and Kidyba, 2010; A. Kugler and M. Kugler, 2009; Saez, Matsaganis, and Tsakloglou, 2012). This study can reconcile the debate by adding a key element: market power. Ongoing work (Biro et al., 2022) also accounts for the role of labor (not product) market power in the tax incidence analysis. However, they analyze an age-specific lump-sum policy, while I study a firm-specific tax rate variation, which allows me to directly measure labor and product market power and alleviates pay equity confounding concerns (Dube, Giuliano, and Leonard, 2019; Breza, Kaur, and Shamdasani, 2018).

The Brazilian payroll tax variation of 20 p.p. is unprecedented. In the US, for example, research that leverages payroll tax variation relies on changes of less than 1 p.p (Guo, 2023).[1] To advance cutting-edge research that leverages employee-level data to offer a nuanced understanding of how corporate taxes affect various groups of workers (Ohrn, 2023; Carbonnier et al., 2022; Risch, 2024), I pair the sizable shock with rich employer-employee matched tax microdata. I find that responses to the tax cut vary not only among different types of workers but also across firms.

Second, my empirical findings document clear evidence that Brazilian firms retain labor market power, which is in line with a burgeoning strand of frontier research (Card et al., 2018; Berger, Herkenhoff, and Mongey, 2022; Lagos, 2019; Jäger and Heining, 2022; Kline et al., 2019; Garin and Silvério, 2019; Benmelech, Bergman, and H. Kim, 2022; Burdett and Mortensen, 1998). I build on this body of work by quantifying the channels through which imperfect competition shapes firms' responses to industrial policies, which in turn impacts the incidence and efficiency of Government subsidies. Differently from Berger, Herkenhoff, and Mongey, 2022, this paper integrates labor and product market power, taking the model directly to heterogeneous firm-level empirical responses. As argued by Manning, 2021, few papers aim to directly estimate the labor supply curve faced by the firm, mostly because it is challenging for researchers to disentangle market from firm-level shocks. The frontier has adopted two alternatives: a model-based, and an experimental-based approach (Dal Bo, Finan, and Rossi, 2013; Dube, Jacobs, et al., 2020;

[1] Studies on the Brazilian payroll tax reform (Dallava, 2014, Baumgartner, Corbi, and Narita, 2022, Scherer, 2015) rely on aggregated sector data, and do not analyze business outcomes. Anonymized firm-level tax data allows for observation of the two margins of imperfect compliance: eligible firms that do not take-up, and those that are treated in outside sectors due to the product criteria.

Belot, Kircher, and Muller, 2019). I contribute to this strand by providing well-identified quasi-experimental evidence, leveraging the uniqueness of the Brazilian reform.

Third, this study also advances the literature by estimating elasticities of substitution between capital and labor (Karabarbounis and Neiman, 2014; D. R. Raval, 2019; Chirinko, Fazzari, and Meyer, 2011; Caballero et al., 1995; Oberfield and D. Raval, 2021). In a meta-analysis, Gechert et al., 2022 criticize prior work because of the use of cross-country variation and omission of the first-order condition for capital. Papers that have addressed these concerns, as I do, using local variation and optimality conditions for both inputs (Harasztosi and Lindner, 2019; Curtis et al., 2021) have suffered from not accounting for labor market power.

Finally, an important industrial policy literature studies Government subsidies for R&D (Bronzini and Iachini, 2014; Howell, 2017); equipment (Zwick and Mahon, 2017); and investment (Criscuolo et al., 2019). This body of work has found that subsidies are more effective for boosting employment in small businesses. This paper is the first to document this pattern for payroll tax incentives. In addition, it posits that market power is a key ingredient in rationalizing the mechanism behind the notable responses of small firms in this literature.

The rest of the paper is organized as follows. Section 1.2 presents institutional background and data. Section 1.3 presents the empirical analysis, including data-driven evidence of market power. Section 2.3 develops the model. Section 1.5 identifies and estimates the model. Section 1.6 estimates the incidence and excess burden associated with the Brazilian payroll tax system. Section 2.7 concludes.

1.2 Institutional Background and Data

This section describes the institutional background of the payroll tax system in Brazil and provides details on the payroll tax reform implemented in 2012. The section then describes the main datasets used to measure the effects of payroll tax variation on various outcomes.

Brazilian Payroll Tax System and the 2012 Reform

Similar to most OECD countries, Brazilian payroll taxes are designed to fund social security programs, such as retirement pensions and unemployment insurance. Tax rates are also similar to those of other OECD countries (see Figure 1.3 for cross-country comparison). In contrast to the tax reforms studied in the past, the Brazilian payroll tax cut program offers unique advantages from an empirical perspective. First, it alleviates pay equity concerns as the policy is targeted at the firm, rather than worker-level. Second, the Brazilian reform offered a large first-stage, evidenced by a 20 p.p. payroll tax reduction. Third, only a few firms were affected, minimizing general equilibrium effects. Fourth, the reform lasted for many years, allowing for short- and long-term decomposition.

Institutional Setting. The Brazilian payroll tax schedule has three components, and all are collected from firms. The main component is a 20% flat tax over the total wage bill, which is affected by the reform. Second, there is an accident risk insurance component that varies between 1% and 3%. The last layer is an 8% to 11% tax on wages, which is employee-specific and can vary among workers in the same firm. These tax components are deposited in a social security fund that pools resources from all workers in the country. This means that the public social security system does not provide individual savings accounts in which resources could be traceable and mapped to workers' specific benefits.

Policy Motivation. The official goal announced for the tax reform was to increase the competitiveness of Brazilian firms. The Government at the time had the tradition of engaging in industrial policies that subsidized specific corporations and sectors. To uncover the Government's rationale for favoring certain firms over others, I conducted extensive empirical investigations. I tested (and rejected) the hypothesis that becoming eligible for tax benefits was associated with more contributions to political campaigns. Section 1.3 leverages an additional analysis that relies on propensity scores to predict eligibility. Overall, the suggestive evidence indicates that the process of defining eligibility was a complex political decision, which did not seem to anticipate sector-specific trends. It is important to underscore that the research design does not assume random eligibility assignment. Instead, it posits that in the absence of the tax reform, eligible and ineligible sectors would have followed a similar trajectory. Section 1.3 presents a set of tests that provide details on the eligibility rules and test trends and balance across the eligibility status.

Eligibility. The policy established sector- and product-specific eligibility criteria for the payroll tax exemption. Product eligibility was defined based on Mercosur Common Nomenclature (NCM). Most of the product-eligible firms are in the manufacturing industry, but treatment due to NCM criteria is not restricted to the manufacturing sector. Indeed, all sectors in the Brazilian economy contain firms treated based on NCM product criteria.[2] Treatment due to the NCM eligibility criterion only allows for a partial payroll tax waiver, according to the share of eligible products in the firms' gross income.

Within broadly defined industries, the reform did not grant eligibility to all sectors. For example, the media industry is eligible for the open television sector, but it is not for cable television. In the lodging industry, hotels are eligible but motels are not. Similarly, Table 1.1 provides a non-exhaustive list of numerous additional examples of similar sectors granted distinct eligibility status. This finely detailed level of eligibility assignment across similar sectors provides a compelling basis for comparison, which I implement in a difference-in-differences framework. It also mitigates confounding concerns from concurrent policies, such as those under the umbrella of "Plano Brasil Maior", which did not target the same sectors at such a granular level. In the empirical analysis, I add industry-year fixed effects to leverage variation within broadly defined industries, which further alleviates concerns related to other industry-specific shocks.

[2]This can be precisely observed in the anonymized micro tax data.

Timing. The first tax bill outlining the policies and eligible sectors was passed in December 2011 and implemented a few months later, April 2012. The reform was initially outlined in an executive bill that skipped prior discussion in Congress. This type of payroll tax cut had never been implemented previously in Brazil, so the policy was not expected by employers and employees. The policy is still valid today, and there is no expectation of being eliminated soon. Several other tax bills added more sectors to the reform in 2013 and 2014.[3]

Tax Variation. On December 14, 2011 Congress enacted the payroll tax cut reform that waived the main component of payroll taxation for a small share of sectors and products. Treated firms faced a uniform decrease in payroll tax rates, from 30 p.p. to 10 p.p. of the total wage bill, without any cap for high-income earners. To provide slight compensation to the Government budget in the face of this large drop in tax collection, targeted firms were forced to pay a small 1% to 2.5% tax on the gross revenue. Importantly, the reform did not affect individuals' perception of the solvency of their retirement plans, because the Federal Treasury committed to cover any deficits caused to the social security system.

Within-Sector Variation. In several of the granular eligible sectors, several firms were not affected by the reform. We need to start by remembering that 45% of firms in Brazil are informal (Ulyssea, 2018b) and do not pay payroll taxes. Also, firms in the "Simples" tax regime are also not subject to payroll taxes,and therefore not affected by the reform even if they are in eligible sectors.[4] Finally, among firms that satisfy all of the eligibility requirements, a substantial share of those do not take-up the benefit. Section 2.2 focuses on understanding imperfect take-up behavior.

Overall, less than 2% of formal firms in the country are impacted by the reform. Even within granularly defined local labor markets, less than 3% of firms are affected. To highlight the modest macro relevance, Table 1.2 shows that at the peak of its implementation in 2014, the payroll tax cut program covered only a relatively small share of Brazilian sectors (9%), firms (1.7%), and workers (6%).[5] Section 1.3 provides several spillover tests that support the view that the reform should be seen as a firm- rather than a market-level shock.

Data and Descriptive Statistics

By combining tax and labor administrative data on the universe of formal firms operating in Brazil between 2008 and 2017, I constructed two samples. One at the firm- and the other

[3]IT, Call Center and Hotels were added in 2012. Retail, Construction and Maintenance were added in 2013. And a final wave in 2014 added Transportation, Infra-structure and Media sectors.

[4]This alternative tax system was created in 1996 and had two main goals: to simplify tax rules and reduce the tax burden on small corporations.

[5]The fact that "Simples" firms are not eligible and there is imperfect take-up in eligible sectors contributes to the share of firms being smaller than the share of sectors. The fact that larger firms are more likely to take-up contributes to the share of workers being larger than the share of treated firms.

at the worker-level. The final dataset is anonymized and arranged in a panel structure. Below, I describe each data source.

Labor Market Data. For labor market data I use *Relação Anual de Informações Sociais (RAIS)*, which is the matched employer-employee data set administered by the Ministry of Labor. This data are compiled annually and contain information on all formal job spells in the country. It uniquely identifies firms and workers based on tax codes (PIS and CNPJ, respectively), which do not change over time. The data include firms' characteristics such as sector, age and location. It also covers detailed workers' information, such as occupation, earnings, race, gender, industry, and municipality, as well as hiring and termination dates. The main shortcoming is the lack of information on informal and non-employed workers. To access information on the informal sector, I rely on the 2010 Census, which is administered by the Brazilian Census Bureau (IBGE). The Census measures formalization rates in each of the 5,300 Brazilian municipalities.

Anonymized Tax Records. To conduct a comprehensive analysis of the tax reform, this study relied on detailed anonymized data from the Brazilian federal tax authority (RFB). These data include information on the universe of corporate tax returns, including payroll and revenue taxes, gross revenue, capital, and profits.[6] The data structure is a panel at firm-by-year level, ranging from 2008 to 2017. A firm is defined based on the 8-digit tax code, known as "CNPJ", which aggregates all establishments by firms. This is the relevant unit of analysis because tax planning across establishments tends to be consolidated at firm-level. In any case, 95% of firms are single establishment and 99% of firms are single sector.

Firm Sample. To appropriately study the payroll tax reform in Brazil using administrative data, I imposed a few sample restrictions. I focus on firms that throughout the analysis have never participated in the *Simples Nacional*, which is a special tax tier not subject to payroll taxes. This restriction is crucial, because firms switching in and out of the *Simples* regime would exhibit gaps in their observed payroll tax data.

The sample provides a broad representation of the Brazilian economy and covers 19 out of 21 industries in the Brazilian economy. The construction industry is not included because the reform applied to construction firms on a site-specific basis, rather than at firm-level. Without access to detailed construction site-level data, I cannot accurately determine the proportion of treated sites within a firm, the number of workers employed at specific sites, or assess the precise effect of the policy on construction payroll tax liability. Also, construction was at the epicenter of the "Car Wash" operation, a massive corruption scandal uncovered during the decade this study examines. Investigations revealed that economic transactions within the construction industry were heavily influenced by illicit business arrangements, which led to the bankruptcy of major construction players.

[6]Due to confidentiality constraints, these data was not shared with the researcher. The anonymized tax data were handled solely by the tax authority on official computers, and all results have been reviewed to preserve full confidentiality.

The retail industry is not included in the sample because I am not able to control for changes in the value-added tax system (VAT) known as ICMS. This tax is predominantly concentrated in the retail industry, in which over 85% of the tax collected stems from VAT (Naritomi, 2019). While payroll taxes are administered at the federal level in Brazil, states are responsible for VAT. During the period of analysis, states engaged in multiple VAT tax reforms. These include sector and product-specific exemptions and rate changes, as well as variations in withholding policies and auditing programs. The main sample is not winsorized or balanced, but the results are robust to these procedures (see Appendix 1.14).

Worker Sample. To maintain consistency between the firm- and worker-level analysis, I apply the same sample restrictions previously discussed to ensure an equivalent set of employers in both data sets. I follow the displacement literature (Jacobson, LaLonde, and Sullivan, 1993; Lachowska, Mas, and Woodbury, 2020) and impose a tenure restriction in order to focus on workers who have been employed for at least three years in the pre-reform period. In this sample, workers are assigned to treatment based on their pre-reform employer, regardless of the firms they end up working for.

Take-up. Figure 1.2 shows that there is a substantial share of eligible firms that do not take-up the benefit. This phenomenon is generalized across all cohorts of eligibility from the beginning of the program. It may be puzzling that numerous eligible firms are not taking advantage of the generous Government subsidies. To interpret this observational fact, it is important to bear in mind that the increase in revenue tax would surpass the payroll tax decrease for only 1% of eligible firms. Thus, the substantial imperfect take-up cannot be rationalized through the lens of a perfect tax optimization choice. A few facts help to rationalize the imperfect take-up. First, the tax bills did not impose punishment for non-compliers, possibly because, from a legal point of view, eligibility was seen as beneficial to firms. Based on the Brazilian tax code, it is implausible that prosecutors would sue firms that do not opt into a supposedly beneficial tax system. Second, enrollment in the program was not automatic, as in the Swedish case studied by Saez, Schoefer, and Seim, 2019.

In Brazil, firms have to self-report eligibility on Government-provided software to enable tax exemptions, based on separate tax forms. Figure 1.1 lists the set of information requested on the tax platform in order to enroll in the payroll tax cut program. Even though enrollment implied a net tax cut, empirical findings in other countries (Henrik J Kleven and Waseem, 2013a; Janet, David, John, et al., 2006; Zwick, 2021; Moffitt, 2007) suggest that the operational process can lead to non-responsiveness, even in dominated tax regions.[7] In line with this view, Figure 1.1 shows that take-up is monotonically increasing with firm size. This pattern is consistent with the fact that larger firms are more

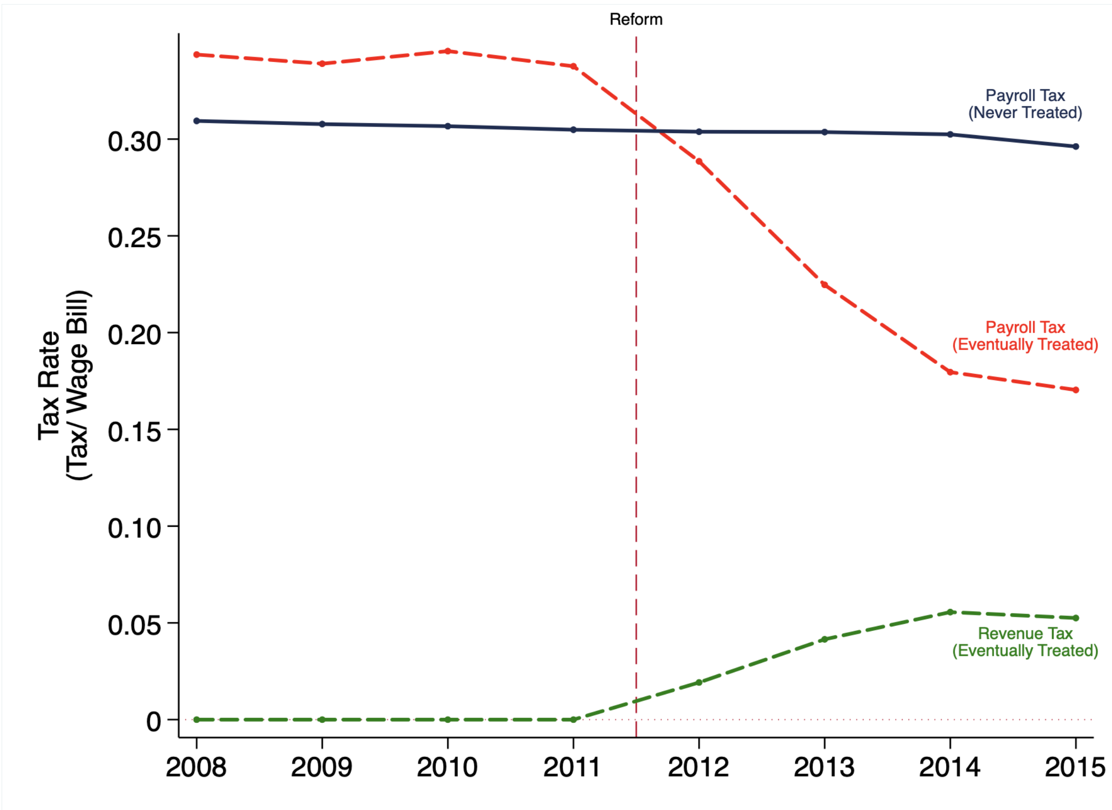
[7]This is related to an extensive body of work dedicated to understanding rational attention (Hoopes, Reck, and Joel Slemrod, 2015) and other frictions that can rationalize low participation rates in public programs (Currie et al., 2001; Heckman and Smith, 2004). Similarly, several papers study the role of tax salience (Chetty, Looney, and Kroft, 2009; Chetty, John N Friedman, and Saez, 2013; Finkelstein, 2009)

likely to have accounting support, be aware of tax benefits, and be able to pay for filling costs.

Payroll Tax Cuts. Figure 1.1 compares payroll and revenue tax rates for firms that were treated at some point in time (eventually treated) vis-à-vis firms that never received the tax benefits. The group of never-treated firms includes, for example, eligible firms that did not take-up the benefit. Revenue taxes are divided by the total wage bill, so all tax rates are comparable. Reassuringly, tax rates from the raw data in Figure 1.1 align well with statutory rates. The figure reports unprecedented payroll tax reductions. For context, studies that leverage payroll tax variation in the US rely on changes of less than 1 p.p (Guo, 2023). Also, it is important to note that the payroll tax drop is considerably greater than the revenue tax increase, which reinforces the interpretation of an overall tax cut as opposed to a tax substitution.[8]

[8]The graph reports averages, but even when we look at outliers in the labor share, only in 1% of cases would it not be advantageous taking up the benefit.

Figure 1.1: Tax Variation



Note: This figure presents the evolution of tax rates for eventually treated vs never treated ones. The blue line depicts that payroll tax rates for never treated firms are stable over time. The dashed red line represents the payroll tax rates for treated firms. The dashed green line presents the revenue tax rates that are substituted in once treatment takes place. Revenue tax rates are computed as a function of the total wage bill in order to facilitate comparisons.

1.3 Empirical Analysis

The payroll tax cut causes a sharp expansion in employment, with small but significant effects on long-term pre-tax wages. In this section, I present details on the main results, including heterogeneity analysis across firm size and workers' characteristics.

Identification Strategy

The main empirical strategy is a fuzzy event study instrumented by sector eligibility. The design explores the staggered implementation of the program and the fact that the vast majority of firms are never eligible or treated. The IV is necessary to adjust for two margins of imperfect compliance: imperfect take-up in eligible sectors and take-up in ineligible sectors due to NCM product eligibility criteria. The first analysis relies on the firm-level sample, where I estimate the following structural equation:

$$Y_{jt} = \sum_{k=-4,\neq -1}^{3} \beta_k D_{jt}^k + X_{jt}'\gamma + \alpha_j + \xi_{I(j),t} + \epsilon_{jt} \tag{1.1}$$

where, X_{jt} are a set of controls for workforce composition (e.g., education, gender, race, age, and its square), $\xi_{I(j),t}$ is industry (broader than sector) interacted with year fixed effect, α_j is the firm fixed effect, and k indexes the time relative to treatment. For each time t relative to treatment, there is one respective first-stage equation. Thus, in total there are K first-stage equations given by

$$D_{jt}^k = \sum_{l=-4,\neq -1}^{3} \pi_{kl} \times \mathbb{I}(t = e_{s(j)} + l) \times L_{s(j)} + \alpha_j + \xi_{I(j),t} + X_{jt}'\delta_k + \eta_{jt},$$
$$\forall k \in [-4,-2] \cup [0,3] \tag{1.2}$$

where, $e_{s(j)}$ is the event date, in which firm j's sector becomes eligible, $L_{s(j)}$ indicates whether firm j's sector is eventually eligible, and the remaining coefficients are the same as described before. Because eligibility is defined at the sector level (mostly at 7-digit), I conservatively cluster at 5-digit industry-by-state level (Bertrand, Duflo, and Mullainathan, 2004 ; Cameron and Miller, 2015). Appendix 1.9 provides more details on the empirical model, underlying assumptions, and reduced-form equations.

I also estimate an IV difference-in-differences model, in which all periods after the policy implementation are pooled into a single post-period indicator. The first stage and structural equations are outlined in equations (1.3) and (1.4), respectively:

$$D_{jt} = \pi L_{s(j)t} + \alpha_j + \gamma_t + \xi_{I(j),t} + X_{jt} + u_{jt} \tag{1.3}$$

where, D_{jt} indicates that firm j is treated in year t, $L_{s(j)t}$ indicates that firm j's sector became eligible before period t, and the remaining coefficients are the same as before. The first-stage coefficient π increases as the take-up rate on treated sectors increases, and deflates as more treatments occur in non-treated sectors due to the NCM criteria. The associated reduced form is expressed in equation (1.4):

$$Y_{jt} = \delta L_{s(j)t} + \alpha_j + \gamma_t + \xi_{I(j),t} + X_{jt} + u_{jt} \tag{1.4}$$

Validity of Design. Identification relies on the assumption that conditional on fixed effects, eligibility is uncorrelated with time-varying unobserved determinants of employment and wage growth. This implies that in the absence of the reform, outcomes for eligible and ineligible firms would follow similar trends. There are two main threats to this design. The first is the potential for Congress to anticipate sector-specific trends when determining eligibility rules. The second threat stems from the possibility that firms might strategically move to eligible sectors after the reform is announced. Section 1.3 provides several tests that mitigate these concerns.

Spillover Analysis

Theoretical predictions regarding the effects of a tax change hinge on whether the shock impacts the entire market or is specific to particular firms. The fact that a very small share (1.5%) of formal firms and workers benefited from the tax cut is indicative (but not conclusive) that the reform should be seen as a firm-specific variation. To further examine this, I follow literature that has considered job-switching patterns to define local labor markets (Felix, 2021). This analysis shows that 67% of Brazilian job switchers stay in the same occupation and region rather than the same industry. That said, I define the local labor market in occupation x region cells.[9] To evaluate spillovers within the local labor market, I provide a evidence that the Brazilian reform was a firm-, rather than market-level shock.

First, I provide purely descriptive evidence that even at the local labor market level, the share of treated firms is small. Figure 1.2 provides intuition that within a labor market there are eligible and ineligible sectors. Within eligible sectors, there are unaffected firms that are either in the informal sector or in ineligible tax tier (*Simples*) or decided not to take-up the benefit. Table 1.1 walks through this logic and shows that conditional on having an eligible sector in the local labor market (LLM), less than 3% of firms in the LLM are affected.

Second, I run a spillover test using firms from the *Simples* tax regime (ineligible tax tier). These firms are ineligible for the payroll tax benefit, but they can operate in eligible industries. If the reform were to create a market-level shock, we should expect to see a negative employment effect in these firms compared with other *Simples* firms in ineligible sectors. Figure 1.3 shows that this is not the case. It reports a small and not statistically different than zero spillover effect. It also shows that *Simples* firms in eligible vs ineligible sectors follow similar trends in the pre-reform period. In contrast, it shows that among *non-Simples* firms, there are substantial effects of being in an eligible sector.

Finally, if there were a spillover effect we would expect to see relatively more wage pass-through in intensively treated local labor markets.[10] The reasoning is that spillovers affect workers' outside options, and therefore generate greater wage hikes. Figure 1.4 shows that the wage effect for workers in high versus low intensively treated markets is not statistically different from each other. This evidence suggests that the driving force

[9] This definition uses the 2-digit occupation code from CBO, and the micro-region defined by the Brazilian Census Bureau (IBGE).

[10] I define treatment intensity by the share of treated firms, but the results are qualitatively the same if I define intensively treated markets based on the average or total amount of subsidy.

underlying the workers' earnings increase is not the market spillover, supporting the view that the Brazilian payroll tax reform should be interpreted as a firm-specific shock.

Validity of the Empirical Design

The identifying assumption is that eligibility, conditional on fixed effects, is uncorrelated with time-varying, unobserved factors that influence employment and wage growth. The validity of this assumption would be violated if Congress anticipated sector-specific trends in its definition of eligibility rules. Another issue could arise if firms strategically chose sectors after the reform was announced. In this section, I conduct multiple tests to address both of these concerns related to selection on eligibility and sector choice manipulation.

Selection on Eligibility

Trends. The concern about selection on eligibility is whether firms that were granted eligibility status might have exhibited different trends relative to those that weren't. To address this, it's common practice to evaluate pre-existing trends. Figures 1.2 and 1.4 depict event study coefficients that reflect reassuring pre-reform results that are not statistically different from zero. This suggests that in the absence of the tax reform, the outcomes for both eligible and ineligible firms (and workers) would have followed parallel trends in the post-period if the tax reform had not been enacted.

Baseline Levels. Besides parallel trends, firms across various eligibility statuses also demonstrated a balance in levels among several characteristics in the pre-reform period, as shown in Table 1.4. This is not surprising per se, as we saw that eligible and non-eligible sectors are fairly similar, as evidenced by the example of hotels and motels. The one characteristic that did not exhibit a balanced distribution across groups was gender, a variable that I will control for in all specifications. Important to highlight that even though the two groups present balanced baseline levels, this is not required for identification. The empirical strategy will not require random treatment assignment, but rather that the two groups would have evolved similarly had the reform not happened.

Alternative Identification. The reason specific sectors were chosen was not disclosure, nor was there an objective criterion to determine eligibility. From an econometric standpoint and with respect to potential identification concerns, it has been established that sector choice did not seem to anticipate sector trends. To further investigate underlying criteria that determined eligibility, I fit a logit model on baseline firms' observable characteristics. I then use propensity scores to break ties in a procedure, which matches firms based on pre-reform deciles on average employment, workers' earnings, firm age, net revenue, and profits. Using this matched sample, I conducted a difference-in-differences analysis as a robustness check. Interestingly, in this alternative empirical strategy, the identification assumption hinges on the Conditional Independence Assumption (CIA), which is validated by the balance tables in Appendix 1.14. Importantly, this strategy does not make any assumptions about the political process that determines eligibility.

Results from both the primary empirical strategy and the alternative matching approach are qualitatively similar. Detailed analysis and the corresponding results are provided in Appendix 1.14. To further substantiate the matching approach, I conducted additional robustness checks. I randomly assigned a placebo treatment and applied the same matching process to these placebo-treated firms. As anticipated, the placebo tests yielded zero effects on employment and wages, and thereby providing evidence that the results are not influenced by any inconsistencies in the matching algorithm.

Manipulation on Sectoral Choice

Sector Immobility. Given the seemingly arbitrary nature of eligibility assignment, one might wonder whether firms could manipulate their sector classification to move to eligible sectors after the announcement of the reform. In this scenario, the concern is that firms that expect employment growth could self-select into treatment, and thereby compromise the causal interpretation. Fortunately, our panel data allow us to track firms and assess whether they changed sectors upon the reform's implementation. The data show only a small number of firms changing sectors, and among these there is no trend of switching to eligible sectors. This low manipulation response aligns with the bureaucratic challenges of changing sectors.

Bureaucratic Process. Firms in regular tax tiers (the object of this study) face a long and costly process to change sectors. To do so, they need first to change their operating agreement, which requires proof that they are operating in a new industry. Subsequently, they must request new operational licenses from multiple administrative bodies, including city, state, and federal tax authorities. Also they must obtain clearance from local tax authorities and civil registry offices. Failure at any of these steps can result in sanctions and fines.

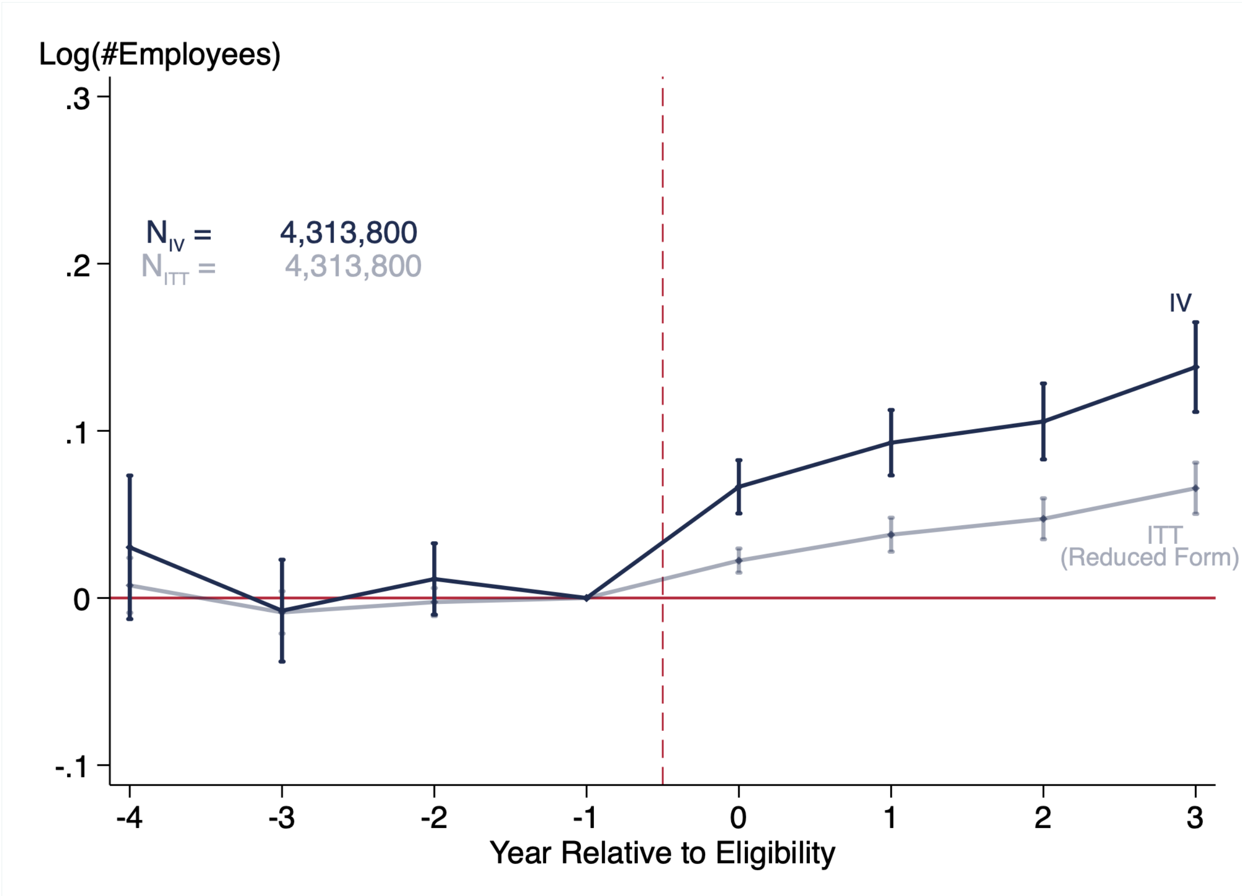
Additional Robustness Tests. To further ensure that sector changes are not driving the results, I conducted several additional robustness checks. First, I assigned firms to eligibility based on their pre-reform sectors, and results remained qualitatively the same. Similarly, when I restricted the sample to firms that never changed sectors, results were unchanged. Taken together, these tests indicate that sector manipulation is not an active margin of response, which reinforces the causal interpretation of the results.

Main Results

In this subsection, I report the causal effects of the payroll tax reform on a comprehensive set of outcomes measured at the firm-level. The findings indicate that after a firm-specific payroll tax cut, employment and wages rise, which provides evidence of labor market power. Reduced-form estimates also reveal that the reform causes an increase in revenue, profits, and a decrease in capital. These results, along with heterogeneous break downs, are key for the structural estimation and provide direct evidence on the incidence of payroll taxation.

Employment. The difference-in-differences estimation, from Equation (1.4), reveals that the payroll tax program causes a highly significant 9% increase (se = 0.03) in the number of employees. The event study analysis validates the parallel trends assumption and shows that there is an immediate employment response that is sustained and slightly increases over time, as shown in Figure 1.2. Based on the policy-induced labor cost variation (see Table 1.1, Column (1)), the corresponding empirical elasticity of employment with respect to the labor cost ($\epsilon^{L}_{1+\tau}$) is -0.71. These results remain qualitatively similar within the balanced sample of firms (Appendix 1.14), which suggests that the employment effect is not governed by the dynamics of firm entry and exit. However, the average effect masks substantial heterogeneity, which I exploit next.

Figure 1.2: Event Study Estimates on Employment



Note: This figure presents event study estimates for employment. The event is the year in which the firm enters treatment for the first time. I normalize results with respect to one year prior to the event. The analysis spans four years prior to entering the payroll tax cut program and three years after. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Firm Size. The richness of the data allows me to evaluate heterogeneity responses based on firm size, which is measured in the pre-reform period. Table 1.1, Columns (2)-(4), reveal statistically significant differences in employment responses between small and large firms. These findings are consistent with the heterogeneity per market concentration, measured by the market share at the local labor market-level.[11] Figure 1.7 shows

[11] Based on job switcher patterns, and following Felix, 2021, local labor markets are defined at the commuting zone x 2-digits occupation level

that both employment and wage effects are greater for firms that exhibit less market concentration.

The firm size analysis is implemented by separately fitting the same specification for a sample of small, medium, and large firms. One advantage of this approach is that comparing small firms in treatment and control groups eliminates the *mean reversal* channel that could potentially explain the heterogeneous pattern. Important to mention that this heterogeneity is not mechanically driven by differences in the first stage. Figure 1.8 demonstrates that the reform impacts labor costs uniformly across the entire firm size distribution. Another potential explanation for the firm size heterogeneity could be financial constraints, which I discuss next.

Table 1.1: Firm-Level Estimates

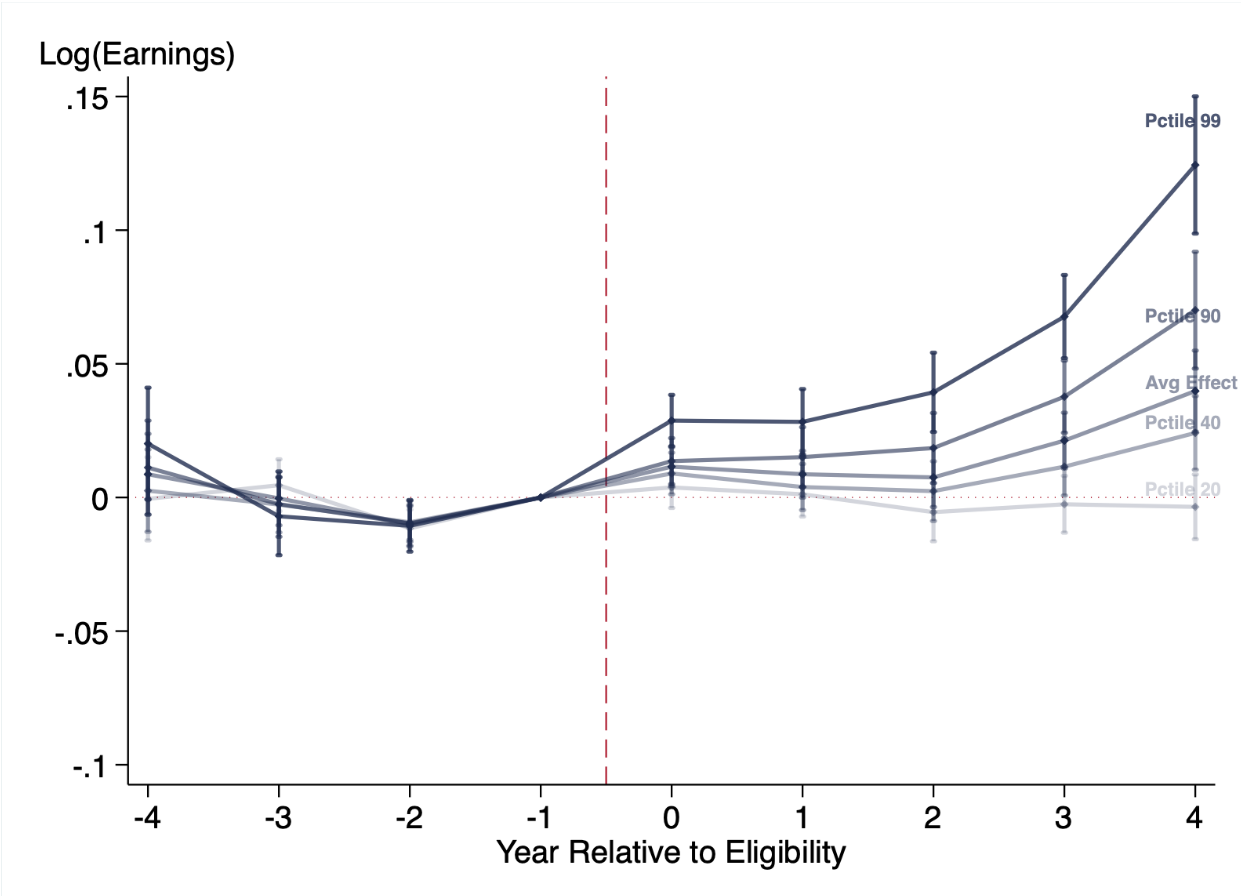
	Log(1+τ)	**Log(#Employees)**	**Log(#Employees)**		
	All Sample (1)	All Sample (2)	Small (3)	Medium (4)	Large (5)
Panel A: IV					
Diff-in-Diff	-.132*** (.003)	.092*** (.029)	.245*** (.047)	.134*** (.031)	.05* (.029)
Long Diff	-.118*** (.005)	.13*** (.028)	.303*** (.046)	.191*** (.039)	.102** (.043)
Panel B: ITT					
Diff-in-Diff	-.063*** (.002)	.044*** (.014)	.096*** (.017)	.07*** (.017)	.03* (.017)
Long Diff	-.066*** (.003)	.063*** (.016)	.133*** (.02)	.101*** (.021)	.054** (.026)
Controls	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓
Sector x Year FE	✓	✓	✓	✓	✓
# Clusters	10,489	10,679	8,548	7,040	6,082
N	4,095,696	4,234,882	2,613,652	685,405	457,937

Note: This table presents IV and reduced-form (ITT) estimates for the firm-level sample. Difference-in-differences coefficient is estimated in equations (1.3) and (1.4), where there is only one post-period. The long difference comes from the period t=+3 in the event study design. Panel A reports IV coefficients, which adjust for imperfect compliance. Panel B reports reduced form coefficients, which are interpreted as the intention to treat (ITT) coefficients. Column (1) reports policy-induced labor cost variation, which provides evidence on the first stage. The remaining columns have log of employment as the dependent variable. Column (2) presents the average effect in the whole sample. Columns (3-5) present heterogeneity based on pre-reform firm size. Firms are categorized as small if they have fewer than 9 workers in the pre-period, medium if they have 10-49, and large if they have more than 50. Standard errors are conservatively clustered at 5-digit industry-by-state level.

Earnings. On average, I find that the reform causes an earnings increase of 2%. In a competitive labor market, firm-specific shocks that do not alter the outside option of workers should generate employment, but no earnings response. The combination of both employment and earnings positive effects serves as evidence of labor market power. In addition to this key insight, several other findings emerge when looking at the effect on workers' earnings.

First, there is a positive average earnings increase that gradually builds over time. More strikingly, however, is the inequality in the wage pass-through across the within-firm distribution. Figure 1.3 illustrates this by showing that the earnings effect across different percentiles of the within-firm distribution displays a monotonic pattern. The effect is null at the bottom and rises to more than 4% at the firm's 99^{th} percentile, which is consistent with Kline et al., 2019; Carbonnier et al., 2022; Ohrn, 2023; and Risch, 2024. The absence of pre-trends across the entire range of the earnings distribution further corroborates these findings. These results shed light on an important consequence of payroll tax cuts: it exacerbates within-firm wage inequality. As the Government reduces payroll tax rates, wages for those who already had higher earnings increase relatively more. To accommodate the more pronounced earnings pass-through to high-skill workers, Appendix 1.13 extends the model to allow for skill-specific labor supply.

Figure 1.3: Earnings Effect Within Firm Wage Distribution



Note: This figure presents event study estimates for wages at different percentiles of the within-firm wage distribution. The event is the year in which the firm enters treatment for the first time. I normalize results with respect to one year prior to the event. The analysis spans three years prior to entering the payroll tax cut program and three years after. Standard errors are conservatively clustered at 5-digit industry-by-state level.

Other Margins of Adjustment So far I have empirically established that upon a sizable tax cut, firms present substantial employment growth and moderate increases in workers' earnings. However, these two response dimensions alone do not provide a comprehensive understanding of the tax incidence. For instance, it is not immediately clear whether the increase in employment arises from scale or substitution. These alternative channels influence prices and output differently, leading to contrasting incidence implications. For example, upon expansion there is an additional supply of goods that can ultimately benefit consumers. However, firms can substitute capital by labor keeping production constant. In that case, there would be no incidence to consumers, but there is still additional demand for labor directing some benefits to workers. Moreover, to infer the incidence on firm owners requires examining not only employment and earnings pass-through, but also the effect on profits.

To measure these responses, we rely on firm-level anonymized balance sheet information from tax records. Unfortunately, not all formal firms have to report balance sheet information. Only firms in the "Real Profit" tax tier are obligated to file that information.[12] Table 1.2 displays the results. Notably, within this tax tier, the treatment effects on labor market outcomes closely resemble those estimated for the entire universe of formal firms (refer to Columns (2) and (4)). Figure 1.5 further validates the parallel trends assumption, which demonstrates that pre-reform coefficients are not statistically different from zero for all outcomes.

The payroll tax reform presents an unambiguous incentive for firms to expand employment from the perspectives of both scale and substitution. Nevertheless, the usage of capital is subject to two counteracting forces. On one hand, reduced labor costs stimulate production, and thus positively affecting capital demand. On the other hand, lower labor costs generate incentives to substitute labor for capital. Column (3) quantifies the net effect on capital, and the relative importance of these two channels. It shows that as payroll taxes decrease, firms expand employment and shift away from capital.

The identification of scale effects, based on the choice of inputs, is a crucial step in determining price responses. This process is formally executed in Section 1.5, in which I estimate the demand elasticity. Combining scale and revenue pins down prices, which is instrumental in assessing the incidence passed onto consumers. The perfect fit between the model and data, plus the overidentification tests, enhance the reliability of the structural estimations. Another advantage of observing many margins of responses is that it allows me to evaluate the coherence among multiple channels and models' predictions, which I turn to in the next section.

Relying solely on employment responses requires the use of structural assumptions to map input choices to profit outcomes - an approach often employed in the tax incidence literature (e.g., Suárez Serrato and Zidar, 2016a; Suarez Serrato and Zidar, 2023). The richness of the data used in this study enables direct observation of the tax cut captured by firm owners in the form of accounting profits. Given that profits can plausibly be negative or zero, I opted not to rely on logs for this specific outcome. Instead, I leverage the analysis in levels, where I divide the point estimates by the average profit in pre-reform years.[13] Reassuringly, results are robust to using the inverse hyperbolic sine transformation.

[12]"Real Profit" is a tax tier for larger firms with annual gross revenue above BRL $78 million - approximately USD 15 million at the current exchange rate.

[13]The analysis in levels become more sensitive to outliers. For this reason, I follow the standard winsorization procedure at the 5% and 95% level (Yagan, 2015; Kline et al., 2019).

Table 1.2: Firms' Margins of Adjustment

	(1) Log Labor Cost $(1+\tau)$	(2) Log Employment	(3) Log Capital	(4) Log Earnings	(5) Log Revenue	(6) Ebit (β/μ)
Panel A: Diff-in-Diff						
Baseline	-.1321***	.1019***	-.0967***	.0251***	-.073***	.0249
	(.0032)	(.0206)	(.0335)	(.0063)	(.0232)	(.1028)
Small Firms	-.1352***	.2691***	-.0858	.0563***	-.0216	-.0241
	(.0067)	(.0366)	(.0784)	(.016)	(.0516)	(.481)
Large Firms	-.1308***	.0573**	-.0953***	.0152**	-.0831***	.0154
	(.0038)	(.0239)	(.0361)	(.0059)	(.0251)	(.0989)
Panel B: Long-Diff (t + 3)						
Baseline	-.1247***	.1279***	-.0337	.0309***	.0495	.3037*
	(.0059)	(.0311)	(.0496)	(.0094)	(.0371)	(.1633)
Small Firms	-.1275***	.3461***	.0085	.0602**	.1864**	.5632
	(.0224)	(.0587)	(.1133)	(.0252)	(.0875)	(.797)
Large Firms	-.1253***	.0969***	-.0312	.0228***	.0223	.2988*
	(.0047)	(.0352)	(.0549)	(.0086)	(.0391)	(.1559)
Controls	✓	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓
Sector x Year FE	✓	✓	✓	✓	✓	✓
N	449,679	450,387	345,217	450,387	374,774	265,889

Note: This table reports difference-in-differences and event study coefficients instrumented by sector eligibility. Each column reports different margins of adjustment, such as labor cost, employment, capital, earnings, gross revenue, and profits. Results are presented for the baseline sample and separately per firm size, which is defined with respect to the median in the pre-reform period. Panel A reports difference-in-differences coefficients. Panel B reports long-diff coefficients, which are the event study coefficients for t+3. Standard errors are reported in parentheses.

Informality

Given the identified causal employment response to the tax cut, it is worth considering whether the increase in employment is due to the formalization of existing employees or the addition of new ones. I present several pieces of evidence that informality is not driving the results.

Transition. The panel structure of the data allows to track previous employment spells for workers who held formal jobs in the past. Essentially, the data enable me to determine for each new hire whether they transitioned from non-employment/informality or another formal job. If the positive employment effect were due to hiring existing informal workers, we would expect to see a sharp increase in the proportion of new hires transitioning from non-employment or informality in treated firms after the reform. However, as Figure 1.5 indicates, the proportion of new hires coming from non-employment and informality remains constant over time and across treatment status, which suggests that formalization is not a significant margin of response.

Regional Variation. Another approach to the informality question is to leverage regional variation in informality rates. Brazil's large and diverse developing economy has local labor markets that range from those resembling developed economies to those similar to African countries. Two years before the payroll tax reform, the Brazilian Census Bureau conducted a national Census survey that provided rich regional informality data at the municipality-level. As Figure 1.6 shows, there is a wide range of informality rates across Brazil's 5,300 municipalities.

I exploit this variation to distinguish the effects of a payroll tax reform in settings with different degrees of exposure to informality. I divide regions into two groups: those below and above the median in terms of formalization rate. If the main employment response to the tax cut was driven by the mere formalization of informal workers, it would be reasonable to expect larger employment effects in regions with high informality. However, my findings indicate the opposite (Table 1.5). One might still be concerned that the labor cost variation induced by the policy in low- and high-informality areas can be different. I show that the first stage is uniform across informality status, which reinforces that formalization is not driving the results.

Workers' Education and Capital Response. As Ulyssea, 2018a notes, informal employment is concentrated among firms with a lower average education. The labor data provide information on workers' educational level, which enables me to compute average education per firm. I show that responses are concentrated in firms with higher shares of qualified workers i.e., firms less likely to hire informally. This serves as additional evidence that the employment effect is not driven by informality (Table 1.5). Finally, if the observed employment effect resulted merely from informality, it would represent a nominal shift with no substantial economic consequence. Yet, as highlighted in Section 1.3, the reform prompts a shift from capital utilization, which suggests that employment responses are real.

Discussion. The empirical evidence that suggests informality is not a primary driver of employment responses aligns with several factors highlighted in previous research. First, informality in Brazil is primarily driven by self-employment rather than informal employment in a formal firm (PNAD, 2012). This implies that informal workers are more similar to entrepreneurs than employees, and their formalization decision is more sensitive to other fixed costs such as licenses to operate, costs related to opening and maintaining a firm, other corporate taxes, legal liabilities, sanitary and security regulations (Maloney, 2004). Second, even though there is a reduction in labor cost, the worker's decision to formalize extends beyond a simple cost-benefit analysis (see Perry, 2007 for discussion).

Additional Results

To interpret the main results, additional analysis is necessary to clarify driving forces that underlie the reduced-form estimates. For example, the earnings effect could be driven by a change in workforce composition rather than earnings pass-through. Similarly, firm size heterogeneity could be driven by liquidity constraints. In this subsection, I provide additional evidence to address these alternative interpretations.

Liquidity Constraints. The observed heterogeneity in firm size is consistent with previous studies (Bronzini and Iachini, 2014; Howell, 2017; Zwick and Mahon, 2017; Criscuolo et al., 2019; Saez, Schoefer, and Seim, 2019), which found that tax subsidies yield greater employment responses by small firms in different contexts. One view is that the payroll tax cut serves to alleviate the financial constraints of the firm, subsequently leading to an increase in employment. I use the rich microdata to empirically test this mechanism.

I use anonymized firm-level balance sheet data to define liquidity constraint as the ratio of short-term assets to short-term liabilities. Cash and short-term bills, are examples of current liabilities. I divide the sample into firms that fall below and above the median based on the pre-reform measure of liquidity constraint. Employment effects for both groups turn out to be strikingly similar. The results reported in Table 1.6 suggest that liquidity constraints do not rationalize the employment responses.

Composition. Interpretation of the earnings results in terms of pass-through could be compromised if, as a result of the policy, firms change the composition of their labor force. However, Table 1.3 demonstrates that the tax reform does not significantly impact the composition of employed workers across various dimensions. The only exception is gender, where the reform induces a marginal but statistically significant effect of 1 p.p in the share of male workers, which is an economically irrelevant effect given the baseline share of 60% (column 3). Columns (1) and (2) show that the effects on the share of workers with high school and college degrees are indistinguishable from zero. Columns (4) and (5) present evidence that the reform did not affect the share of employed white workers or the average employees' age.

Columns (6) and (7) follow the composition analysis of Kline et al., 2019. Column (6) investigates whether the reform influences firms to hire workers from different parts of the earnings distribution and finds no effect on the quality of new hires proxied by their pre-hiring earnings. Column (7) reports the impact on a quality index, computed using a Mincer regression of log earnings on a quartic in age fully interacted with gender and race, estimated annually with firm fixed effects as additional controls. Table 1.3 suggests no evidence of skill upgrading in response to the reform.

Finally, I explore whether the tax cut affects the types of occupations firms employ. I exploit the detailed CBO occupational codes[14], which contain 2,300 occupations. After ranking occupations based on pre-reform earnings and grouping them into percentiles, I determine each firm's average occupation percentile. Table 1.7 reveals a sharp zero effect of the reform on firms' average occupation percentile. This empirical fact implies that the tax reform expands employment within, rather than between, occupations. This underscores the fact that there is no major shift in worker composition or technology-induced labor demand.

Table 1.3: Effect on Labor Composition

	(1) Share High School +	(2) Share College +	(3) Share Male	(4) Share White	(5) Avg Worker's Age	(6) Log Earnings New Hires (bf hired)	(7) Log Quality
Post × Treatment	.0091	.0099	.0132***	.0005	-.1343	-.0014	-.0005
	(.0085)	(.0061)	(.0034)	(.0045)	(.1497)	(.0116)	(.005)
Mean	.52	.11	.59	.67	39.72	7	1
Controls	×	×	×	×	×	✓	✓
Firm FE	✓	✓	✓	✓	✓	✓	✓
Sector x Year FE	✓	✓	✓	✓	✓	✓	✓
# Clusters	7,925	7,925	7,930	7,924	7,930	6,924	7,561
N	2,494,842	2,494,842	2,521,030	2,491,523	2,521,030	604,988	1,739,827

Note: This table reports difference-in-differences coefficients to assess the effect of the reform on the firm's labor composition. The empirical specification is the same as presented in equations 1.3 and 1.4. The regression is estimated in the balanced sample of firms to isolate any noise due to firm entry and exit. The goal is to depict the firm-level compositional effect. In Columns (1)-(5), controls are not included to avoid over-controlling given that in these regressions the outcome is part of the set of variables typically used as controls. Column (6) reports the effects on new hires' previous earnings. Column (7) depicts the effects on a measure for worker's quality based on a Mincer regression of log earnings on a quartic in age fully interacted with gender and race, estimated annually with firm fixed effects as additional controls. Standard errors are reported in parentheses.

Worker-level Analysis

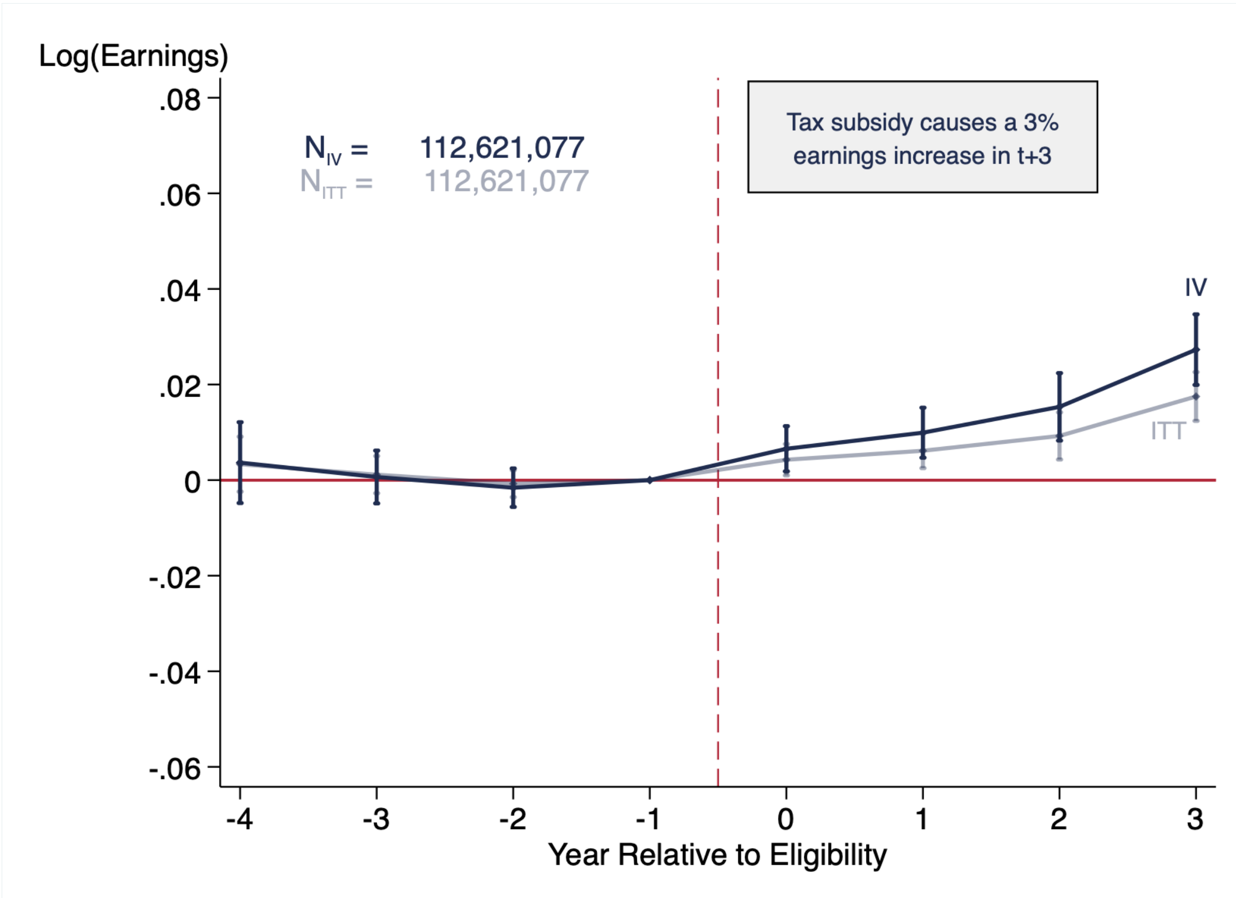
An alternative approach to evaluating earnings pass-through involves tracking workers as opposed to firms. This strategy offers two main advantages. First, it enables the assessment of whether firm-level earnings response is the result of pass-through or shifts in labor force composition. A zero wage pass-through could be consistent with positive firm-level earnings response in the instances of upscaling the labor force. Second, it yields insights into how tax variation impacts workers' career paths, particularly for

[14]Classificação Brasileira de Ocupação (CBO) is the legal norm for classifying occupations in the Brazilian labor market. It was established on decree approval 397/2002.

various types of workers. To conduct the worker-level analysis, I fit the empirical specification outlined in Section 1.3 to the worker sample, which allows for the inclusion of worker fixed effects.

Workers' Earnings. Figure 1.6 depicts a pronounced drop in the gross earnings paid by firms, which is mostly attributed to the mechanical reduction in payroll tax rates. Consistent with the positive earnings response measured at the firm-level, Figure 1.4 reveals that workers' take-home payments increased by 2%. The effect intensifies to 3% three years after the tax cut. This result reinforces the notion that the positive earnings response is rationalized by pass-through rather than compositional changes.

Figure 1.4: Event Study Estimates of Workers' Earnings



Note: This figure presents event study estimates for average earnings (net of payroll taxes) for stable workers. I normalize results with respect to one year before the treatment event. The analysis spans four years before the payroll tax cut program and three years after. Blue markers report the IV estimates and gray markers are the intention-to-treat. Standard errors are conservatively clustered at 5-digit industry-by-state level.

Occupation. Consistent with the earnings inequality result found within firms, I show that workers in high-skill and managing positions benefit relatively more from the reform. To implement this analysis, I rely on the CBO to split employees into two occupation groups. Managers, directors, and qualified technical positions are in the top bucket and comprise 15% of the sample, and the remaining 85% of lower positions are evaluated separately. Figure 1.7 shows that the pass-through to highly skilled workers is 6%, and is almost zero to low-skilled workers. Appendix 1.13 provides an extension of the model that includes two types of labor and is able to rationalize the findings in a setting in which low-skill workers have higher labor supply elasticity. This is consistent with a more concentrated local labor market for high-skill labor, and suggests that a low-skill labor market operates closer to perfect competition.

Racial Wage Gap. The payroll tax program does not distinguish workers based on background characteristics such as race. It offers a flat 20 p.p. cut that remains constant across all income levels, which suggests no explicit intention to disadvantage workers of a specific race. However, if race correlates with occupation or other factors that determine unequal pass-through, the tax system can inadvertently widen the racial wage gap. A unique feature of the Brazilian data is its ability to identify workers and their race. I use the policy-induced tax variation to find that white workers benefit significantly more from the reform than non-white workers. This intriguing result holds even after controlling for firm fixed effects, which suggests that the racially unequal pass-through is not attributed to firm sorting. I also evaluated heterogeneous pass-through according to gender but found zero statistical difference. Figure 1.8 summarizes the analysis across different types of workers.

Unintended Discrimination. This paper introduces another critical aspect to the public debate. Despite the fact that racial discrimination is a pressing social issue in modern society, it has not been incorporated in the tax literature.[15] This might be because modern tax codes do not contain any explicit elements of racial discrimination that could be classified as either statistical or taste-based discrimination. However, taxes can exacerbate racial inequality through indirect channels that are substantiated in existing frictions. This paper provides novel evidence that behavioral responses to tax changes can lead to unintended consequences for racial inequality.

1.4 Model

The empirical evidence provided so far (particularly in Table 1.2) emphasizes the importance of the product market in shaping responses to payroll taxation. The presence of imperfect product competition allows the transmission of cost shocks to consumers, a phenomenon the payroll tax literature has not yet studied. In its original form, the Marshall-Hicks framework acknowledges that firms can set prices above marginal cost. However, it assumes that labor markets operate in perfect competition, which is in stark contrast to the positive earnings effect documented in this paper.

In this section, I extend the conventional pass-through framework (Criscuolo et al., 2019; Harasztosi and Lindner, 2019) to incorporate imperfect competition in both product and labor markets. By combining these two competitive friction, which are often modeled separately, I can provide insights consistent with the heterogeneous firm-level responses. The actual degree of market power in the economy is an empirical question uncovered by

[15] A few exceptions are Brown, 2022 and Holtzblatt et al., 2023, who study racial inequality in the context of couples' taxation in the US. Elzayn et al., 2023 study racial discrimination in tax audits.

firms' response to the tax shock. The model yields key identifying equations that directly connect the reduced-form estimates to structural parameters of easy interpretation. Using the combination of model and data, I can quantify mechanisms of response and measures of tax incidence and efficiency.

Setup

Motivated by the firm-specific nature of the reform studied in this paper, the model considers a partial equilibrium framework in which firms operate as monopolists in the product market and monopsonists in the labor market. The model has a single period, in which firms choose their input mix and output level. After selling production, the firm concludes its operations. Firms are endowed with a CES technology with constant returns, which uses capital and labor as inputs.

$$f(L,K) = (s_L L^\rho + s_K K^\rho)^{\frac{1}{\rho}}$$

where the aggregate L is the total efficiency units of labor at the firm and s_g are the inputs' cost share ($g \in \{L, K\}$). The capital market operates in perfect competition, which means that the marginal revenue product of capital equals its cost. However, the labor market operates in imperfect competition, and labor supply elasticity ϵ dictates the firm's ability to mark wages below the marginal revenue product of labor. Firms face an upward-sloping labor supply curve, and cannot discriminate wages across incumbents and new hires.

$$w_j = A_j L_j^{\frac{1}{\epsilon}}$$

The wage-setting rule suggests that if wages rise due to a firm-specific shock, both incumbents and new hires experience equal benefits — an observation supported by the data. From a theoretical standpoint, the static labor supply curve can be micro-founded by an analogy to Industrial Organization's discrete choice models, which are employed to estimate demand with differentiated goods. In the labor market context, the "differentiation" arises from workers' preference for particular employers. This argument is formalized in Appendix 1.10. As in Card et al., 2018; and Haanwinckel, 2023, I assume that firms ignore their contribution to the tightness of the labor market — an approximation that is appropriate when firms have small market share.

The output market operates in monopolistic competition, with firms determining quantity based on a constant price elasticity denoted as η (Hamermesh, 1996; Criscuolo et al., 2019). Specifically, firms face the inverse product demand described by $P_j = Q_j^{\frac{-1}{\eta}}$. The subscript j indexes a specific firm, but for ease of notation this subscript will be omitted in the rest of the paper. The degree of monopolistic power is dictated by the parameter η, which is flexible to accommodate any market structure, including perfect competition. Given the output choice, firms solve a cost minimization problem to decide on the input mix. The Government can manipulate labor cost $(1+\tau)$ through perturbations in the payroll tax rate (τ). The percentage variation in labor cost induced by the Brazilian policy is denoted by ϕ_1.

Firm's Problem

Profit Maximization The firm chooses output to maximize profits, according to the following program:

$$\max_{Q} \underbrace{Q^{1-\frac{1}{\eta}}}_{\text{Revenue}} - \underbrace{A(1+\tau)L^{1+\frac{1}{\epsilon}} - rK}_{\text{Cost, C(Q)}}$$

At the optimum, firms choose quantity that equates marginal cost with marginal revenue:

$$\underbrace{\left(\frac{\eta-1}{\eta}\right)Q^{\frac{-1}{\eta}}}_{\text{Mg Revenue}} = \underbrace{\frac{\partial C(\tau,Q)}{\partial Q}}_{\text{Mg Cost}} \tag{1.5}$$

In contrast to a perfectly competitive environment, the marginal cost is no longer a linear function of the output level (see proof of Lemma 1, in Appendix 1.10). The intuition is that there is an increasing cost to expand plant size due to inframarginal wages. As a result, imperfect labor competition limits the pass-through to employment. Mathematically, equation (1.5) reveals how output level influences labor demand by raising the marginal cost of scale expansion. This relationship is increasing in the firm's market power (decreasing in η). The employment effect is further determined in the cost minimization program, which I turn to next.

Cost Minimization Once the output quantity is fixed, firms decide on the input mix that minimizes cost. Formally,

$$\begin{aligned} \min_{K,L} \quad & A(1+\tau)L^{\frac{1}{\epsilon}+1} + rK \\ \text{s.t.} \quad & f(K,L) \geq Q \end{aligned}$$

At the optimum, the labor choice equates the marginal cost of labor to the marginal revenue product of labor:

$$\underbrace{\overbrace{\left(\frac{\epsilon+1}{\epsilon}\right)}^{\text{inverse mark down}} A(1+\tau)L^{\frac{1}{\epsilon}}}_{\text{MCL}} = \underbrace{\lambda f^{1-\rho} s_L L^{\rho-1}}_{\text{MRPL}}$$

Note that the marginal cost of labor is decreasing on the level of labor market competition, which guides the steepness of the labor supply curve. Putting together the optimal input choice and applying the envelope theorem, I can compute the cost function. The monopsony power in the labor market breaks the linear relationship between average and marginal cost:

$$\underbrace{\frac{\partial C}{\partial Q}}_{\text{Mg Cost}} = \underbrace{\frac{C}{Q}}_{\text{Avg Cost}} + \overbrace{\underbrace{\frac{C_L}{\epsilon}\frac{1}{Q}}_{\text{Avg Incumbent Rent}}}^{\textbf{Monopsony}}$$

I denote this new term as the average incumbent's rent because it is related to the wage increase perceived by inframarginal workers when the firm increases plant size. In particular, the rent converges to zero as we move to perfect competition ($\epsilon \to \infty$), similar to traditional models (Hamermesh, 1996). The nonlinearity in the cost function will be key to understanding pass-through responses to payroll tax reforms.

Pass-Through

Thus far, I have presented the framework for firms' decisions in both output and input markets. This section develops intuition on the interaction between these decisions and the policy-induced payroll tax variation. In particular, it sheds light on the role of market power in shaping the pass-through, which ultimately drives the incidence and efficiency of the payroll tax system.

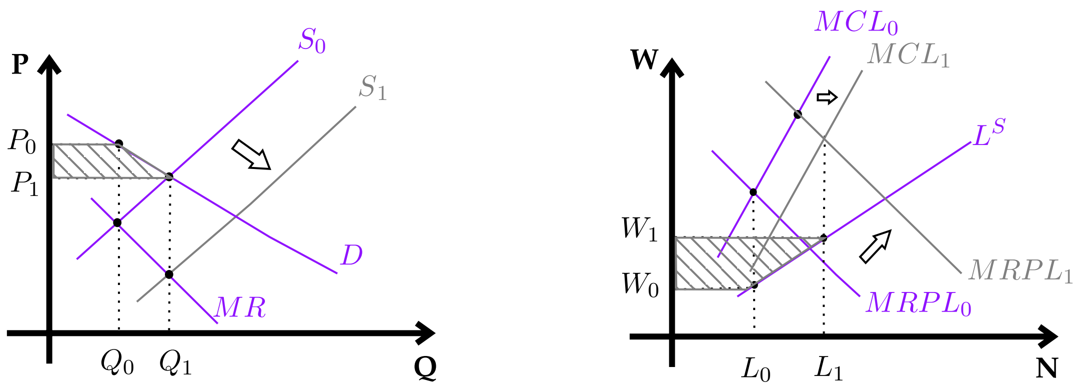
Output Market

In the output market, a payroll tax reduction shifts the supply of goods. The consequences for output depend on two factors: (i) the behavioral response, which determines the magnitude of the shift in product supply/ labor demand, and (ii) the slope of the demand curve. Figure 1.5 illustrates how the price effect increases with market power. To quantify this effect, I totally differentiate the pass-through equations to compute price elasticity with respect to labor cost:

$$\epsilon^P_{1+\tau} = \frac{-1}{\eta}\epsilon^Q_{1+\tau}$$

The price elasticity depends on the scale response, and product market power, which is determined by the constant elasticity η. Due to monopsony power in the labor market, the scale effect cannot be evaluated solely on the basis of product market. Remember that now the inframarginal rent affects the plant size decision. At the optimum, the effects of tax policy on marginal revenue should equal the effects on marginal cost:

Figure 1.5: Conceptual Framework



(a) Product Market (b) Labor Market

Note: This figure illustrates pass-through in the product and labor markets from a firm-specific payroll tax cut. The left graph shows the intuition for the case of product monopolistic competition. Compared with the perfectly competitive case, there is a smaller quantity (or scale) effect due to the price-setting power. On the right, the graph depicts the intuition for the monopsonistic case. In this framework, the employment effect is not as large as in perfect labor competition, but as the tax reform expands the labor demand, it provokes a wage increase.

$$\underbrace{-\frac{1}{\eta}\epsilon^{Q}_{1+\tau}}_{\substack{\text{Effect on mg revenue} \\ \text{(η drives slope of D)}}} = \underbrace{\overbrace{\epsilon^{\lambda}_{1+\tau}}^{\text{Direct}} + \overbrace{\epsilon^{\lambda}_{Q}\epsilon^{Q}_{1+\tau}}^{\text{Inframarginal}}}_{\substack{\text{Effect on mg cost} \\ \text{(shift in supply curve)}}} \tag{1.6}$$

Equation (1.6) sheds light on two mechanisms through which imperfect competition affects firm responses. First, competition in the output market flattens the demand curve ($\frac{-1}{\eta}$), which enhances the scale effect. Second, competition increases the pass-through to the marginal cost, which amplifyies the shift in the supply curve and thereby the scale effect. It is important to highlight the fact that the elasticities expressed in equation (1.6) are endogenous to the tax system. Appendix 1.10 further develops this formula to establish a closed-form solution for the pass-through as a function of primitives, which are expressed in equations (1.10)-(1.13).

Labor Market

Labor market forces determine the tax pass-through according to the effects on the marginal cost of labor and marginal revenue product of labor. Figure 1.5 provides intuition on the interaction between imperfect labor competition and a firm-specific shock. Equations

(1.7) and (1.8) quantify the elasticity of the marginal cost of labor and the elasticity of the marginal revenue product of labor with respect to labor cost in the case of monopsonistic labor markets.

$$\frac{\partial \log MCL}{\partial \log(1+\tau)} = \overbrace{1}^{\text{Direct effect on MCL}} + \overbrace{\frac{\epsilon^{L}_{1+\tau}}{\epsilon}}^{\text{Inframarginal on MCL}} \tag{1.7}$$

$$\frac{\partial \log MRPL}{\partial \log(1+\tau)} = \overbrace{\epsilon^{\lambda}_{1+\tau} + \epsilon^{\lambda}_{Q}\epsilon^{Q}_{1+\tau}}^{\text{Direct + inframarginal on mg rev}} + \overbrace{(1-\rho)(\epsilon^{Q}_{1+\tau} - \epsilon^{L}_{1+\tau})}^{\text{Effect on MPL}} \tag{1.8}$$

Marginal Cost of Labor. The pass-through to the marginal cost of labor (MCL) is comprised of two components. As in a perfectly competitive labor market, the first component perfectly correlates MCL with variations in labor cost. The second component is unique to monopsonistic firms and can be decomposed into two channels: (i) the behavioral response, which governs the shift in the marginal cost of labor, and (ii) the slope of the marginal cost of labor. Market power affects these two channels in opposite directions. While it dampens behavioral responses, it amplifies the steepness of the marginal cost of labor.

Marginal Revenue Product of Labor. As equation (1.8) suggests, the effect of the tax policy on the marginal revenue product of labor depends on the pass-through to the marginal product of labor (MPL) and marginal revenue. Pass-through to the marginal product of labor is negatively related to the substitution across inputs ($\sigma_{KL} = \frac{1}{1-\rho}$), positively related to the scale effect, and negatively related to the employment effect. Pass-through to marginal revenue depends on the direct and inframarginal effects of the firm-specific labor cost variation. Note that marginal revenue depends on the labor cost $(1+\tau)$ and the output level. Therefore, when the firm reacts to a labor cost reduction by increasing plant size, the scale effect inflates costs and offsets part of the initial cost reduction. Equation (1.9) relies on evelope arguments to quantify these responses, and the associated effect to the marginal cost.

$$\overbrace{\frac{\partial \lambda(Q,\tau)}{\partial(1+\tau)}}^{\substack{\text{Effect on}\\ \text{mg cost}}} = \overbrace{\frac{AL^{1+\frac{1}{\epsilon}}}{Q}}^{\substack{\text{Effect on}\\ \text{avg cost}}} + \overbrace{\frac{AL^{1+\frac{1}{\epsilon}}}{Q\epsilon}}^{\substack{\text{Direct effect on}\\ \text{incumbent rent}}} + \overbrace{\frac{A(1+\tau)}{\epsilon}\left(\frac{\epsilon+1}{\epsilon}\right)\frac{L^{\frac{1}{\epsilon}}}{Q}\frac{\partial L}{\partial(1+\tau)}}^{\substack{\text{Indirect effect on incumbent}\\ \text{rent from L response}}} \tag{1.9}$$

The interaction between labor market power and pass-through to marginal cost is unambiguous. The higher the market power, the higher the direct pass-through to the incumbent's rent; it also amplifies the indirect effect on the incumbent's rent due to labor responses.

1.5 Structural Estimation

This section connects model and data. This structural mapping allows me to credibly estimate parameters of interest, and understand mechanisms.

Identification and Interpretation

To operationalize the structural estimation, I derive the model's predictions for the Brazilian payroll tax reform. These responses form a system of equations that depend on parameters: the labor supply elasticity faced by the firm (ϵ); capital-labor elasticity of substitution (σ); output demand elasiticity (η). I present direct connection between the structural parameters and reduced-form estimates.

Pass-through Formulae. Following the derivation outlined in Section 2.3 (and detailed in Appendix 1.10), I compute closed-form solutions for the tax pass-through to employment, capital, earnings, and revenue. To embrace all elements of the Brazilian tax reform, I also take into account the revenue tax variation, which turns out to have muted effects due to the small rate variation on the revenue side and the small share of firms subject to this tax.[16] Once I account for product and labor market power, the effects of the Brazilian tax reform on employment, capital, revenue, and earnings can be expressed as a function of observables and the three parameters to be estimated (ϵ, η, ρ):

$$\beta_L = \left(\frac{\epsilon\sigma}{\sigma+\epsilon}\right)\left[\left(\frac{(\epsilon+2\epsilon^L_{1+\tau})(\sigma-\eta)}{\sigma\epsilon}\right)\left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)-1\right]\phi_1 \tag{1.10}$$

$$\beta_K = \left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\left(\frac{\epsilon+2\epsilon^L_{1+\tau}}{\epsilon}\right)\left(\sigma-\eta\right)\phi_1 \tag{1.11}$$

$$\beta_{Rev} = (1-\eta)\left[\left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{\epsilon+2\epsilon^L_{1+\tau}}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\right]\phi_1 \tag{1.12}$$

$$\beta_W = \frac{\epsilon^L_{1+\tau}}{\epsilon}\phi_1 \tag{1.13}$$

where s_L is the labor share, $\epsilon^L_{1+\tau}$ is the empirically estimated elasticity of employment with respect to the labor cost, and ϕ_1 measures the first stage associated with the policy — i.e., the percentage variation in tax rates induced by the reform. Using anonymized tax data, I precisely estimate ϕ_1. The pass-through formulae developed here are more general than the ones employed in recent studies that assume perfect labor competition. My framework can accommodate perfect labor competition as a particular case, in which ϵ goes to infinity. Taking the limit of pass-through equations (1.10)-(1.11), I recover the

[16]Since the revenue tax has negligible effects, I will omit them in the main text. Careful derivation of the revenue tax perturbation can be found in Appendix 1.10.

same expressions derived in a standard Marshall-Hicks analysis and estimated by Curtis et al., 2021; Criscuolo et al., 2019; and Harasztosi and Lindner, 2019. In the standard competitive case, substitution and scale effects are separable, as illustrated below.

$$\lim_{\epsilon\to\infty}\beta_L = \Big(\underbrace{-s_K\sigma}_{\text{substitution}} - \underbrace{s_L\eta}_{\text{scale}}\Big)\phi_1 \qquad \lim_{\epsilon\to\infty}\beta_K = s_L\Big(\underbrace{\sigma}_{\text{substitution}} - \underbrace{\eta}_{\text{scale}}\Big)\phi_1$$

Identification. Manipulating the pass-through expressions (1.10-1.13), I find closed-form solutions for the structural parameters. Equation (1.14) directly maps the labor supply elasticity faced by the firm to the reduced-form elasticities estimated in the data. The intuition is that the ratio of the employment and earnings effect identifies the slope of the labor supply curve faced by firms:

$$\epsilon = \frac{\beta_L}{\beta_W} \tag{1.14}$$

From the capital and labor responses, σ_{KL} is identified:

$$\sigma = \frac{\beta_K - \beta_L}{\beta_W + \phi_1} \tag{1.15}$$

The parameter σ is derived from contrasting the capital and labor responses. Equation (1.15) depicts the intuition that as β_K decreases relative to β_L, this is an indication that firms are substituting capital for labor. Also, it is interesting to note that when β_W goes to zero, σ boils down to the standard expression from previous studies that assumed perfect labor market competition. Another angle to read equation (1.15) is that ignoring labor market power would generate a biased estimate for the capital-labor elasticity of substitution. Finally, I can identify the output demand elasticity using the capital and revenue responses:

$$\eta = \frac{\sigma\beta_R - \beta_K}{\beta_R - \beta_K} = \frac{-\beta_Q}{\beta_P} \tag{1.16}$$

The economics behind equation (1.16) are that η can be identified based on the ratio between scale and price responses to the tax reform. This ratio determines the slope of the demand curve in the product market.

Estimation Methods. I rely on the Classical Minimum Distance (CMD) approach to estimate structural parameters. The CMD minimizes the squared difference between the model and data, weighting it by the inverse variance-covariance matrix, $\hat{W}^{-1}$. Formally, the method solves $\min_\beta[\xi(\hat{\beta}) - \xi(\beta)]'\hat{W}^{-1}[\xi(\hat{\beta}) - \xi(\beta)]$, where $\xi(\beta)$ is the vector of model predictions = $[\epsilon^L_{1+\tau}, \epsilon^K_{1+\tau}, \epsilon^W_{1+\tau}, \epsilon^R_{1+\tau}]$, and $\xi(\hat{\beta})$ is the vector of reduced-form estimates = $[\hat{\epsilon}^L_{1+\tau}, \hat{\epsilon}^K_{1+\tau}, \hat{\epsilon}^W_{1+\tau}, \hat{\epsilon}^R_{1+\tau}]'$. Standard errors are computed based on a parametric bootstrap. I also compute these parameters using a seemingly unrelated regression (SUR) to jointly estimate equations (1.14)-(1.16). I use the Delta method to estimate standard errors for each structural parameter.

Parameter Estimates. Table 1.8 presents estimates for the three parameters of interest. Column (1) presents baseline results for all firms, and columns (2) and (3) break the estimates down based on firm size. There are several reasons to break the estimates down this way. First, they are highly correlated with measures of market concentration we can observe, such as market share in the local labor market. Second, it is policy informative, given that firm size is a characteristic that is easy to target in policy design. Third, a large literature finds that small firms react more to industrial policies, which contributes to general interest in understanding small firms' behavior.

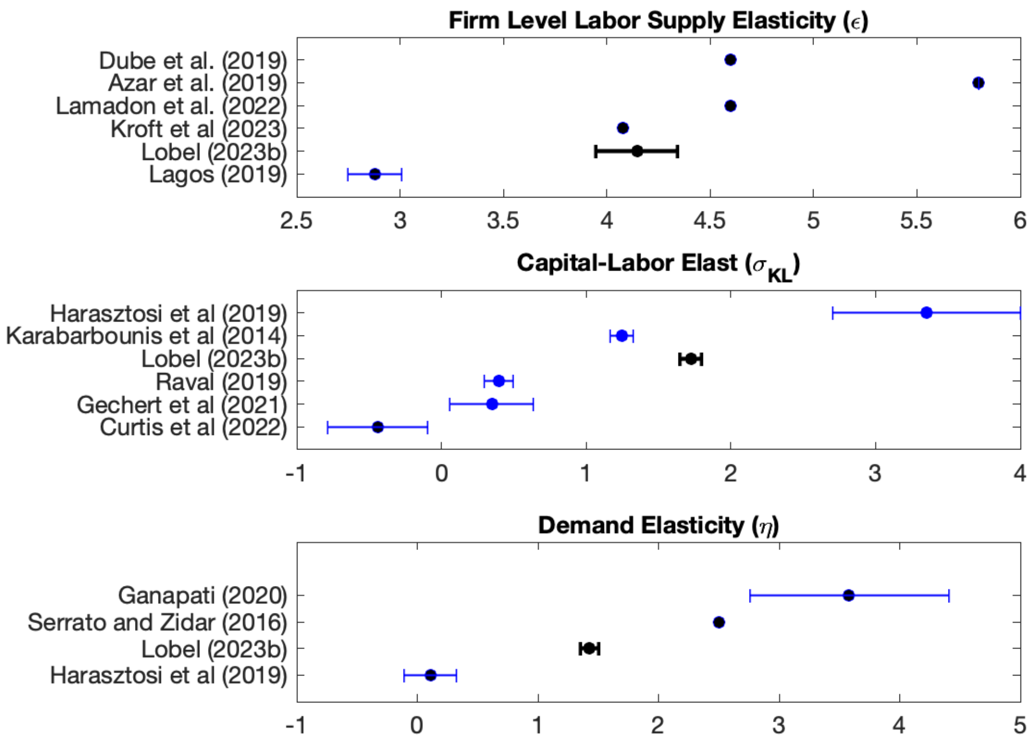
Elasticity of Substitution. The labor-capital elasticity of substitution ($\sigma_{KL} = \frac{1}{1-\rho}$) is equal to 1.72 (se 0.08) at baseline. This result is similar to that of Karabarbounis and Neiman, 2014 and implies that capital and labor are substitutable, which supports the view that lowering the cost of capital may increase income inequality (Piketty and Zucman, 2014).[17] Interestingly, I find that capital and labor are more substitute in small firms (5.01, se 0.34) than in large firms (1.25, se 0.08). This result is valuable, because most of the literature on capital-labor elasticities focuses on large firms in manufacturing. In contrast, my study encompasses a wide range of firm sizes and sectors. Greater substitutability identified in smaller firms can reconcile my estimates with recent literature, focused on large manufacturing firms, which finds that capital and labor are complementary.

Labor Supply. The labor supply elasticity faced by the firm ϵ is 4.15 (se 0.20), which is remarkably close to recent estimates: 4.08 (Kroft et al., 2020); 4.0 (Card et al., 2018); 4.6 (Lamadon, Mogstad, and Setzler, 2022). I am also not that far from the 2.88 estimate that Lagos, 2019 found for Brazilian firms. Figure 1.6 summarizes the literature and points out several studies that report labor supply elasticity between 3 and 5. My baseline estimate implies a wage markdown of 0.81 ($\mu = \frac{\epsilon}{1+\epsilon}$), which suggests that Brazilian firms capture 19% of the marginal revenue product of labor. Columns (2) and (3) report that the labor supply elasticity for small and large firms is 5.75 (se 0.33) and 4.25 (se 0.28), respectively. This result is consistent with the increasing and monotonic relationship between labor market power and firm size demonstrated by Yeh, Macaluso, and Hershbein, 2022.

Demand Elasticity. Output demand elasticity with respect to price is 1.43 (se 0.07) a value greater than one, which aligns with the conventional notion that monopolies operate on the elastic side of the demand curve. If a firm independently decides to raise prices, the quantity loss outweighs the revenue gains from higher prices. Heterogeneous responses to the tax variation reveal that large firms have substantially more market power in the product market. The output demand elasticity for small and large firms is 5.21 (4.21) and 1.10 (se 0.06), respectively. These elasticities are key to examining the theoretical implications of market power on the scale response to a tax cut — a phenomenon that will play a central role in the subsequent discussion of underlying mechanisms.

[17] Other recent studies have found that capital and labor are complements (D. R. Raval, 2019).

Figure 1.6: Literature Benchmark



Note: This figure places my estimates with respect to estimates in the literature. Parameters are outlined on the x-axis and their respective estimates are on the y-axis. The top panel refers to the labor supply elasticity faced by the firm. The middle panel reports the capital-labor elasticity of substitution. Finally, the bottom panel depicts the output elasticity with respect to price.

Overidentification Test. To assess the validity of the model, I compare the reduced-form estimates with their corresponding model predictions, based on equations (1.10)-(1.13) and the estimated structural parameters. Note that there are four moments and only three unknown parameters (ϵ, σ, η), which enables me to test for overidentifying restrictions. The p-value for the J-test, reported in the last row of Table 1.8, indicates that the restriction is not rejected. This finding provides evidence that the model fits the data well and is appropriately specified.

Mechanisms. We further leverage the data to investigate whether the employment increase is due to scale, or substitution responses. Scale refers to an expansion on output production, which increases the use of both inputs. Substitution refers to the extent that employment increases as a replacement for capital. Appendix 1.10 demonstrates how scale is identified. Equation 1.17 shows that the scale response can be quantified based of reduced-form coefficients and structural parameters estimated in the data.

$$\beta_Q = -\left(\frac{\epsilon + 1}{\epsilon}\right)\left(1 + \frac{\epsilon^L_{1+\tau}}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L} + \frac{1}{\epsilon}}\right)\left(\frac{\eta(\epsilon + 2\epsilon^L_{1+\tau})}{\epsilon + \epsilon^L_{1+\tau}}\right)\phi_1 \tag{1.17}$$

Table 1.4 reports that employment increases by 12%[18], with the scale margin accounting for 52% of this effect. The interplay between scale and substitution has unique implications for analyzing tax incidence and efficiency. A more prominent scaling effect leads to greater price reductions, which ultimately benefit consumers.

[18]The empirical moments refer to the long-diff regressions, i.e., responses measured in t+3. This is the preferred timing because some responses such as workers' earnings take time to manifest. However, using the pooled difference-in-differences coefficients do not cause abrupt changes to the structural estimates.

Table 1.4: Scale vs Substitution

	(1)
Mechanisms	**Baseline**
Price effect, β_P	-0.05
Scale effect, β_Q	0.07
Scale / Employment, β_Q/β_L	0.52
Empirical Estimates	
Employment effect, β_L	0.12
Capital effect, β_K	-0.03
Earnings effect, β_W	0.03
Revenue effect, β_R	0.05
Cost Shares	
Labor	0.80
Capital	0.20
J-test	
Overid test (pvalue)	0.74
Observations	
N	450,387

Notes: This table presents the parameters estimated, according to the method presented in Section 1.5. In the "Mechanisms" section, the table reports effects on prices (β_P), scale (β_Q), and the share of employment effect that is explained by the scale response $\frac{\beta_Q}{\beta_L}$. In the empirical section, the table displays coefficients estimated in Section 1.3, and used for the structural estimation, as well as the cost shares. At the bottom, the table displays the p-value associated with the J-test for overidentification.

1.6 Incidence and Efficiency Gains

In this section, I establish the incidence of payroll taxes on workers, firm owners, and consumers. The computation of tax incidence lays the groundwork for a welfare measure, which delivers a measure for the deadweight loss associated with payroll taxation. This section leverages empirical estimates to provide two key insights. First, a novel payroll tax examination that accounts for the role of consumers in the tax pass-through. Second, a credible design to precisely measure the distortionary costs arising from payroll taxes in Brazil.

Incidence Framework

Government. The tax base is determined by the total wage bill. Therefore, when payroll tax rates drop, there is a mechanical effect on tax collection,

$$dM = Bd\tau = B(\tau_1 - \tau_0)$$

where τ_0 is the payroll tax rate in the pre-reform period, and τ_1 is the post-reform rate. Nonetheless, the empirical analysis in Section 1.3 reveals substantial employment and wages responses to tax variation, which partially offset the mechanical tax loss. The resulting behavioral effect on tax revenue is given by:

$$dH = \tau_0 dB = \tau_0 B \left(\frac{\epsilon + 1}{\epsilon} \right) \beta_L$$

Putting all together, the impact of the reform on total tax collection is the mechanical effect net of behavioral adjustments:

$$dR = dM + dH = B \left[d\tau + \tau_0 \left(\frac{\epsilon + 1}{\epsilon} \right) \beta_L \right]$$

This equation offers two direct interpretations. First, a greater employment response implies less tax revenue loss. Second, for a given employment response, labor market power exacerbates wage pass-through, which results in reduced tax revenue loss. For each dollar that is effectively lost in tax collection, it is possible to identify the associated gains. To ensure comparability with existing literature, I rely on a money metric approach for welfare measurement.

Firm owners. As in Suárez Serrato and Zidar, 2016b; Fuest, Peichl, and Siegloch, 2018b, the incidence of the reform to firm owners is quantified based on the share of tax dollars captured by firms in the form of profits. The difference is that in this paper I directly observe profits, as opposed to relying on structural assumptions. To compute the surplus appropriated by firm owners I use the reduced-form coefficients:

$$d\pi = \epsilon^{\pi}_{1+\tau} B \frac{s_\pi}{s_L} \phi_1$$

where, $\epsilon^{\pi}_{1+\tau}$ is the elasticity of profits with respect to the labor cost, while s_L and s_π represent the labor and profit shares, respectively. I rearrange terms to write the effect on firm owners as a function of the total wage bill. The benefit of this approach is that it allows all individual welfare measures to be referenced to the same base, which appropriately weights the welfare attributed to each stakeholder.

Workers. In a monopsonistic labor market, the tax impact on worker surplus is illustrated by the tax-induced variation in area above the labor supply curves, and below the wage times the number of workers. The change in worker surplus can be computed by,

$$dB = w_1 L_1 - \int_0^{L_1} AL^{\frac{1}{\epsilon}} dL - \left(w_0 L_0 - \int_0^{L_0} AL^{\frac{1}{\epsilon}} dL \right) = B\beta_W$$

where w_0, L_0, w_1, L_1 refer to the wage level and employment before and after the reform, respectively. The intuition is that the incidence borne by workers is dictated by the wage effect. Thus, in a perfectly competitive labor market — where all jobs offer equivalent compensation for a given skill set — the incidence to workers is null. This is because under perfect competition, employment at a specific firm provides no additional benefits, since workers have equally attractive opportunities elsewhere.

Consumers. Analogously, the tax impact on a monopolistic product market equilibrium illuminates the welfare effects to consumers surplus, which is computed by the variation in the area between the demand curve and the price times the quantity. The change in consumerers' surplus can be computed by:

$$dC = \int_0^{Q_1} Q^{\frac{-1}{\eta}} dQ - P_1 Q_1 - \left(\int_0^{Q_0} Q^{\frac{-1}{\eta}} dQ - P_0 Q_0 \right) = \frac{B}{s_L} \frac{\beta_R}{\eta - 1}$$

The intution is that the effect on consumers is driven by the output price reduction. The reform mitigates labor costs, and a portion of this cost reduction is transferred to the output price, thereby benefiting consumers. Despite prices not being directly observed, they can be inferred from the revenue effect and the demand elasticity η, which is estimated based on the perfect fit between the model and data. Intuitively, the combination of observed effects on revenues and production inputs allows us to back out the price response.

Directly from the data we can also compute the impact on consumers by relying on a residual method. The benefits of a payroll tax cut must be distributed somewhere. If only a small portion is allocated to profits and workers' earnings, then majority of the benefits must be directed to consumers. An advantage of this approach is that it relies solely on empirical estimates. The incidence estimates show that the precise and residual approaches are consistent with each other.

Efficiency Gains

The efficiency gain induced by a discrete payroll tax cut can be computed following the steps outlined in Appendix 1.11. Equation 1.18 precisely measures, from the empirical responses, the deadweight loss associated with the payroll tax reform.

$$\Delta W = B\Bigg[\underbrace{\beta_w}_{\text{worker, dw}} + \underbrace{\frac{\beta_\pi s_\pi}{s_L}}_{\text{firm owner, } d\pi} + \underbrace{\frac{\beta_R}{s_L(\eta-1)}}_{\text{consumer, dp}} + \underbrace{\overbrace{(\tau-\tau_0)}^{<0 \text{ (tax cut)}} + \tau_0\frac{\beta_L(\epsilon+1)}{\epsilon}}_{\text{Governement, dT}}\Bigg] \tag{1.18}$$

Relatedly, the efficiency gains can be measured through the "Marginal Value of Public Funds" (MVPF) metric, applied across a variety of contexts to evaluate the willingness to pay in relation to the net fiscal cost (Mayshar, 1990; J. Slemrod and Yitzhaki, 2001; Henrik Jacobsen Kleven and Kreiner, 2006; Hendren, 2016; Bailey et al., 2020).

Estimates. Upon establishing the theoretical incidence and efficiency, I proceed with the structural estimation, as shown in Table 1.5. Panel B reveals that consumers bear 65% of payroll taxes, while firm owners and workers bear 23% and 12%, respectively. Risch, 2024 reports a similar incidence to workers, yet observes no change in employment following a tax increase on S-Corp's owners. A possible explanation for the distinct employment effect is that reductions in payroll taxes decrease the cost of labor, generating incentives for businesses to expand plant size and substitute capital for labor. The aggregate welfare gains experienced by these stakeholders surpass the decrease in Government revenue, which results in a MVPF of 1.66.

Discussion. The analysis conducted herein yields insightful findings for the tax incidence literature. A key takeaway is that payroll taxes are predominantly paid by consumers. This novel insight, although not yet thoroughly explored in the tax literature, aligns remarkably with minimum wage incidence studies (Harasztosi and Lindner, 2019). Furthermore, the efficiency gain from a tax cut is inversely proportional to the distortionary effects of the tax. Essentially, a higher efficiency gain signifies prior to the tax cut there was higher deadweight loss from taxation. The substantial welfare gain calculated for Brazil underscores the prevailing notion that taxes exert particularly distortionary effects in developing economies. This view is supported by the MVPF calculation, which falls in the upper end of the 0.5-2 range reported by Hendren and Sprung-Keyser, 2020.

1.7 Conclusion

In this paper, I study an unprecedentedly large payroll tax reduction that affected a small subset of firms in Brazil. I use firm-level microdata and an empirical strategy that leverages exogenous variation in the eligibility rules to estimate firm and worker-level responses to the tax cut. While capital decreases, a payroll tax reduction causes an increase in employment, wages, revenue, and profits. Firms respond to the reduction in labor cost by substituting capital for labor, and by prominently increasing production. This expansion in production then pushes output prices down, leading revenue to respond less than inputs. These results shed light on a novel and important insight for tax policy: payroll taxes are primarily absorbed by consumers. Furthermore, the analysis reveals that

Table 1.5: Structural Parameters and Incidence Estimation

	(1) Identified by	(2) Estimate
Panel A. Parameters Estimate		
Labor Supply Elasticity, ϵ	$\frac{\beta_L}{\beta_W}$	4.15
K-L Elasticity of Substituion, σ	$\frac{\beta_K - \beta_L}{\beta_W + \phi_1}$	1.72
Demand Elasticity, η	$\frac{\sigma\beta_R - \beta_K}{\beta_R - \beta_K}$	1.43
Panel B. Incidence		
Worker, dB	β_W	0.12
Firm Owner, $d\pi$	$\frac{\beta_\pi s_\pi}{s_L}$	0.23
Consumer, dp	$\frac{\beta_R}{s_L(\eta - 1)}$	0.65
Government, dT	$\Delta\tau + \tau_0\beta_L\frac{\epsilon + 1}{\epsilon}$	-0.60
Welfare, dW	$\frac{dw + d\pi + dp + dT}{dT}$	0.66
MVPF	$\frac{dw + d\pi + dp}{dT}$	1.66

Note: This table bridges reduced form and structural estimation. Panel A identifies and estimates structural parameters. Panel B identifies and estimates payroll tax incidence to workers, firm owners, and consumers. Panel B also reports efficiency measures such as the MVPF.

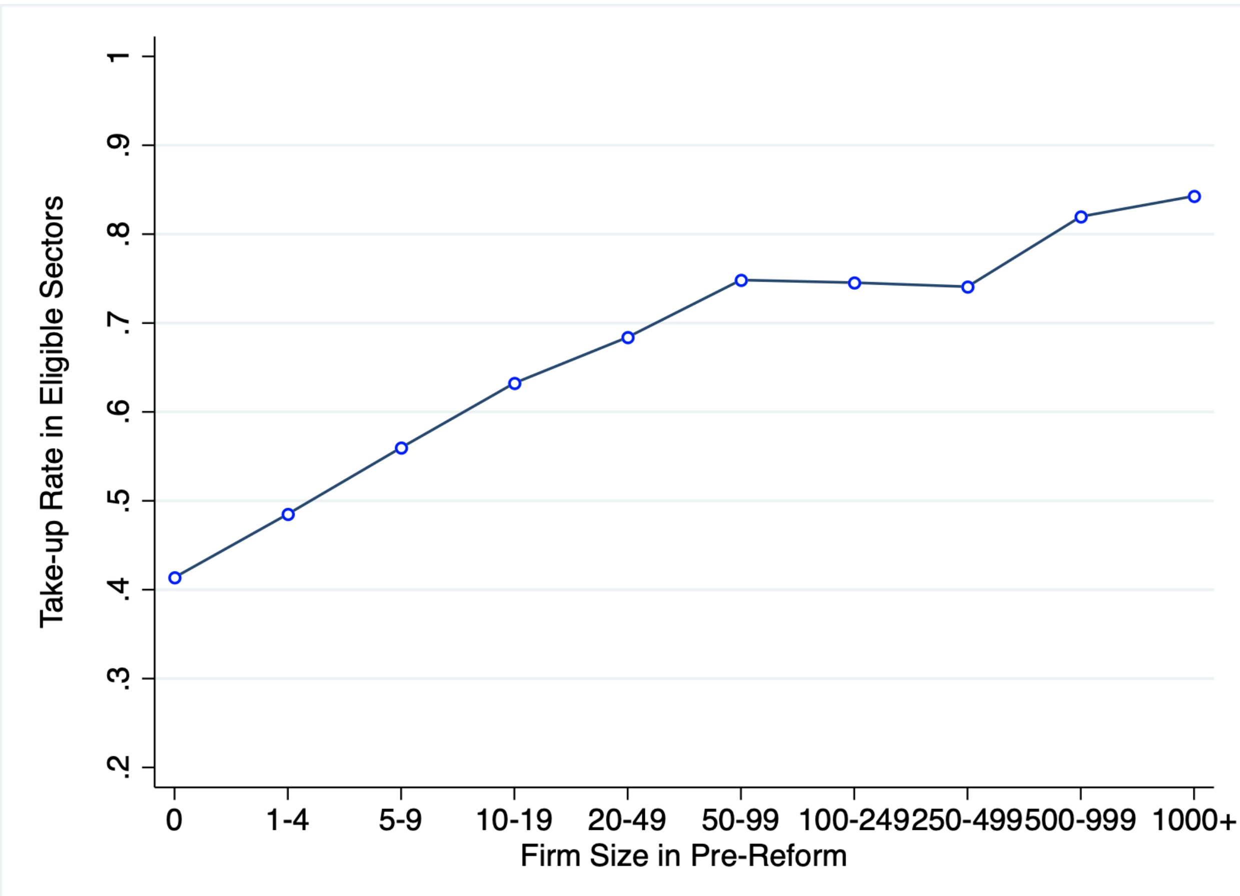
responses vary by type of firms and workers. Skilled workers capture a larger share of the tax benefits compared to low-skilled employees, while small businesses benefit more than larger corporations.

The combination of firm-specific shock and positive workers' earnings effect provides compelling evidence against perfectly competitive models. I use direct evidence of firm-level pass-through to underscore the role of (product and labor) market power in mediating tax incidence and efficiency. Imperfect competition can account not only for the incidence to consumers but also for heterogeneous firm responses. In macro-level policy that affects the entire economy, market power dynamics also come into play but are more challenging to identify under the general equilibrium effects. The insights from this study illuminate the role of firms in dictating the consequences of tax policy.

One key takeaway for industrial policies - characterized as subsidies targeted to specific businesses - is that responses vary depending on the type of firms. Notably, firms with market power tend to reduce inputs and output to a lesser extent in response to tax increases. With that in mind, it may be advantageous to target industrial policies to small businesses, or integrate tax policy with market power regulation. For macro tax policies, market power forces remain influential, highlighting that the "tax does not stay where it lands". Specifically, labor taxes are primarily borne by consumers. This lesson is relevant not only for understanding the distributional consequences of payroll taxes but also other labor policies such as the minimum wage.

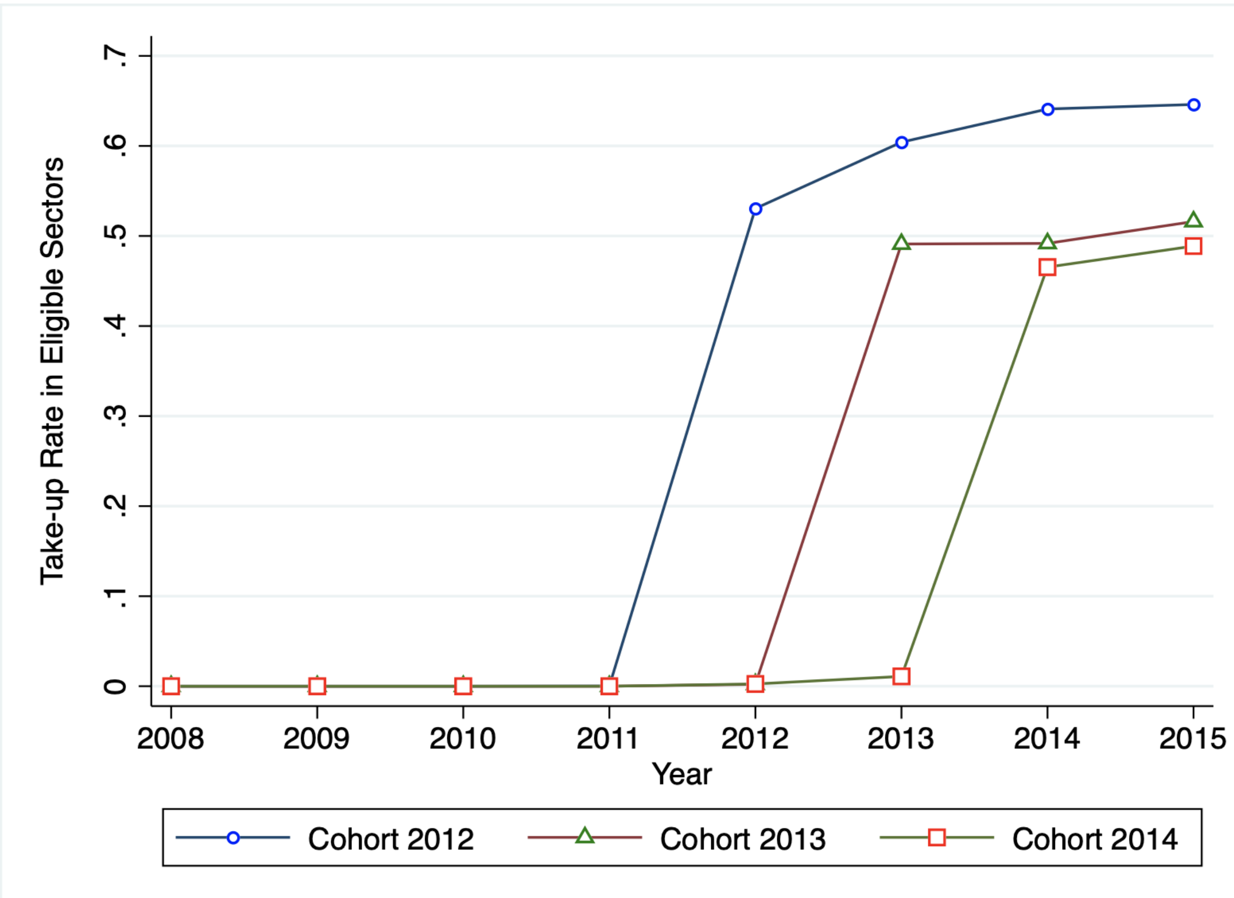
1.8 Figures and Tables

Figure 1.1: Take-up per Firm Size



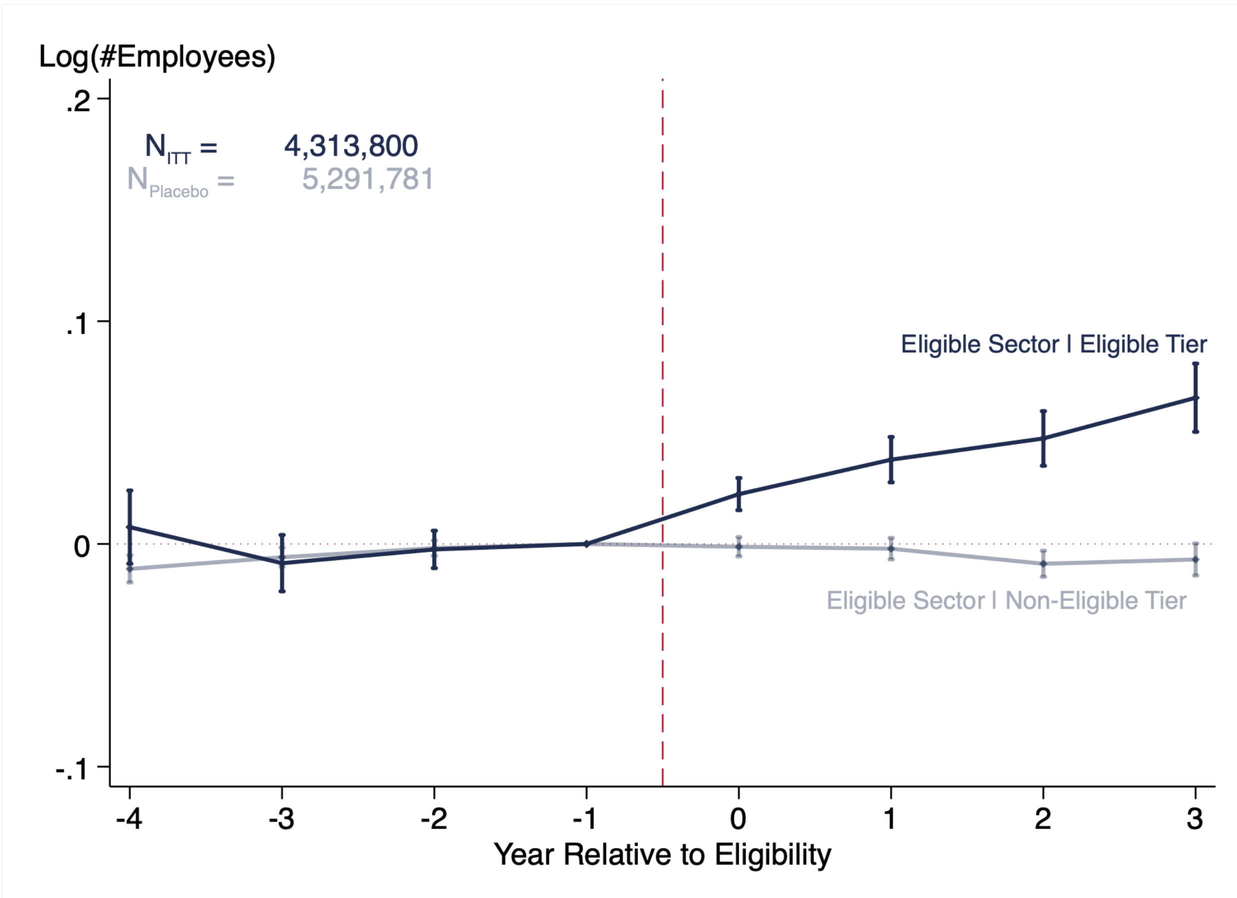
Note: This figure plots cohort-specific take-up rates among eligible firms. Eligibility is based on the firm's 7-digit sector and its observed tax tier. Firms in the "Simples" tax regime are not eligible for the reform, even if they belong to eligible sectors. Firms that have ever participated in the "Simples" regime are not included in this analysis. The figure is computed in the year 2015 after all cohorts have gained eligibility. Firm size buckets are constructed in the pre-reform years. As can be seen, eligibility is monotonically increasing with firm size.

Figure 1.2: Take-up



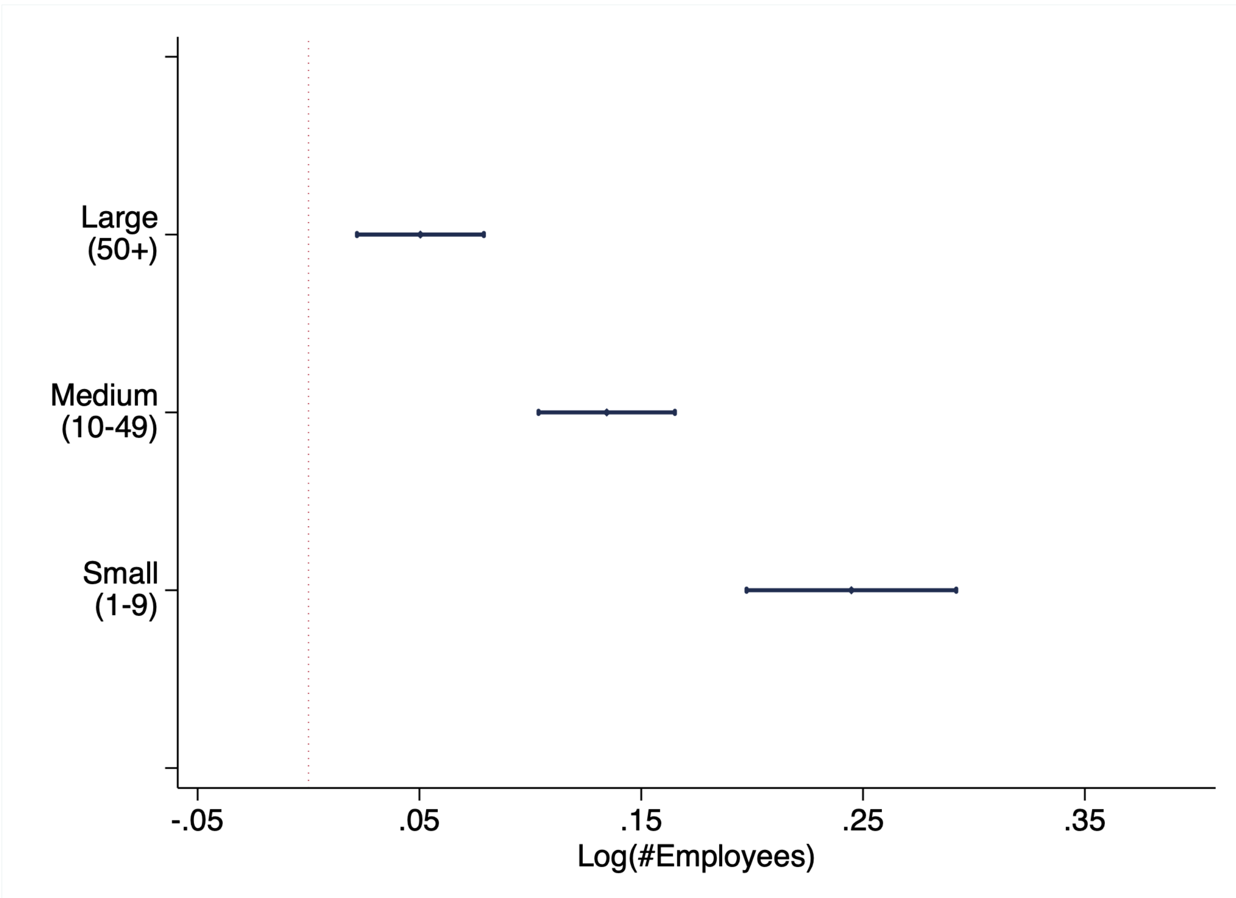
Note: This figure plots cohort-specific take-up rates among eligible firms. Eligibility is based on the firm's 7-digit sector and its observed tax tier. Firms in the "Simples" tax regime are not eligible for the reform, even if they belong to eligible sectors. Firms that have ever participated in the "Simples" regime are not included in this analysis. As expected, take-up rates are zero in the years prior to the implementation of the reform to each cohort.

Figure 1.3: Spillover Test



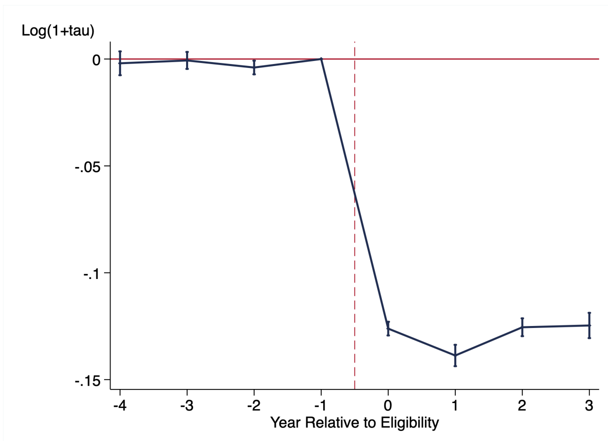
Note: The gray line plots event study coefficients that show non-statistically significant spillover effect to firms in eligible sectors, but ineligible tax tiers. The gray line is estimated on a sample that is restricted to firms in non-eligible tax tiers ("Simples" regime) and depicts a comparison between firms in eligible and non-eligible sectors. To avoid concerns about tier changes, this analysis is restricted to firms that have never changed tiers. The blue line is estimated on a sample that is restricted to firms in eligible tax tiers ("non-Simples" regime). It reports the intention to treat (ITT), i.e., compares eligible firms in eligible vs non-eligible sectors. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Figure 1.4: Employment by Firm Size

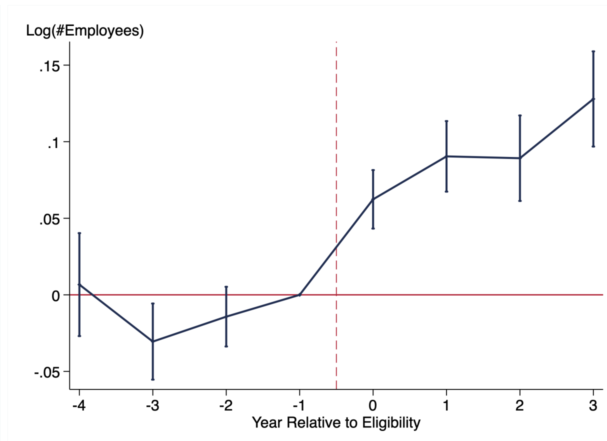


Note: This figure presents the event study estimates for the firm-level estimates, for three firm size groups (small, medium, and large firms). Size categories are defined in the pre-reform period. Firms are classified as small if they had less than nine employees, medium if they had between 10 and 49 workers, and large if they had more than 50 workers. The blue marks plot the employment difference-in-differences coefficient from the IV specification. Standard errors are reported and conservatively clustered at the 5-digit industry-by-state level.

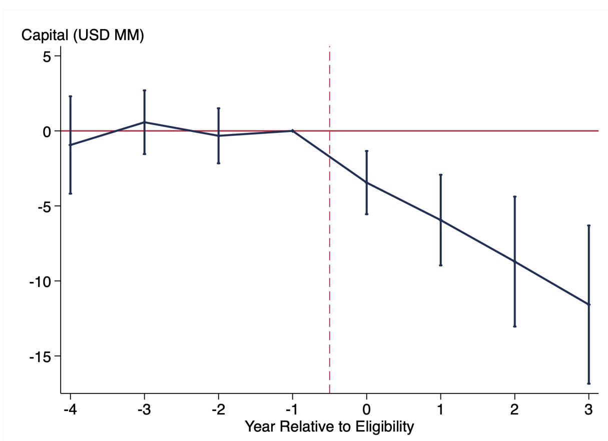
Figure 1.5: Firms' Margins of Adjustment



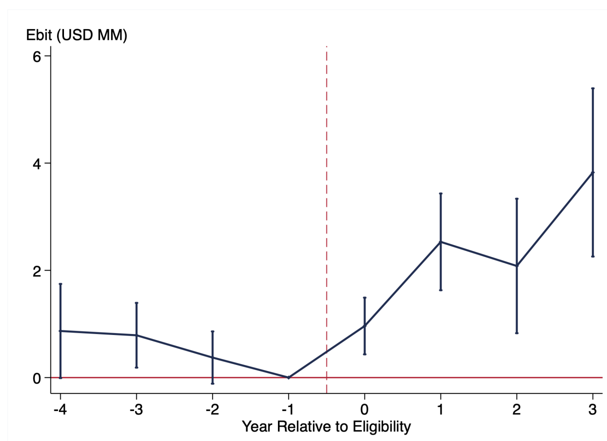
(a) Effects on payroll tax rates



(b) Effects on employment



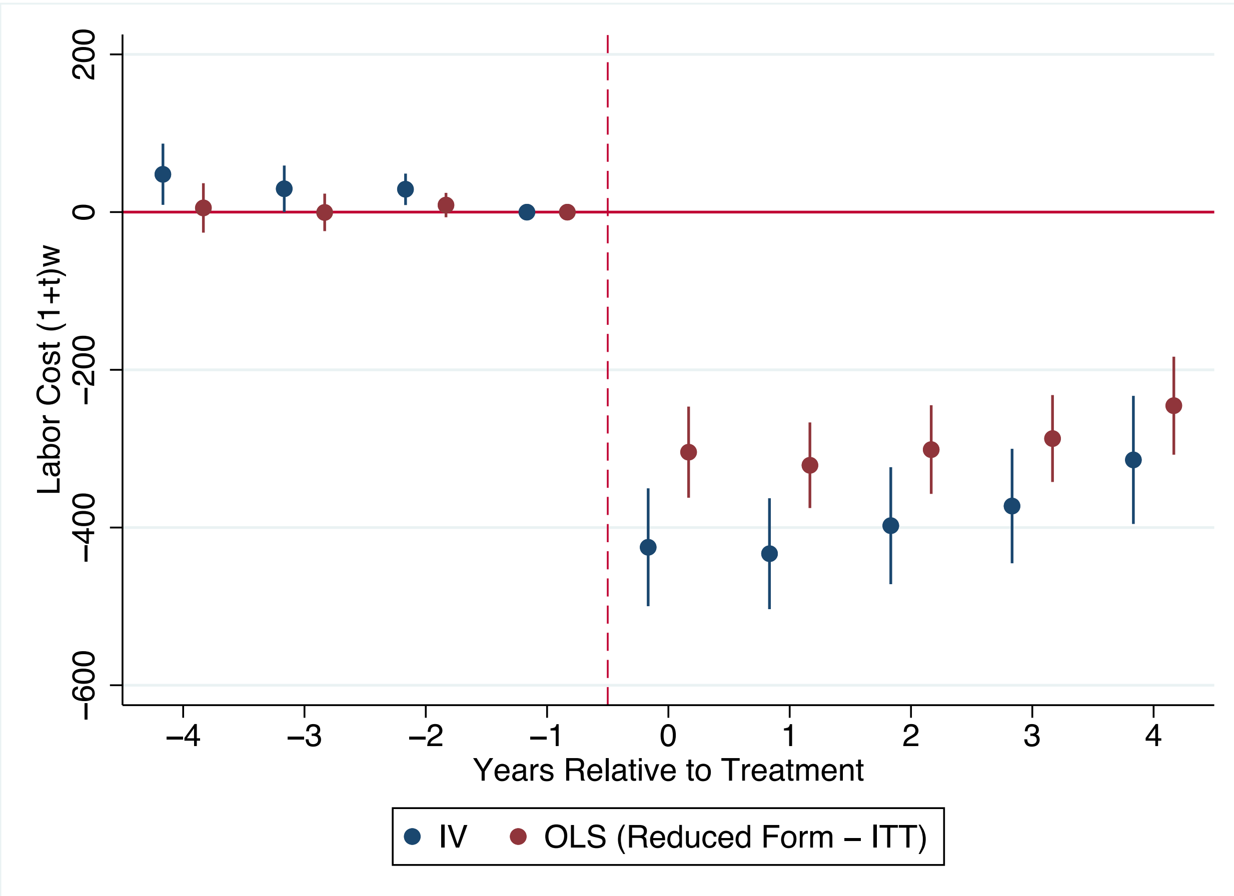
(c) Effects on capital



(d) Effects on profits

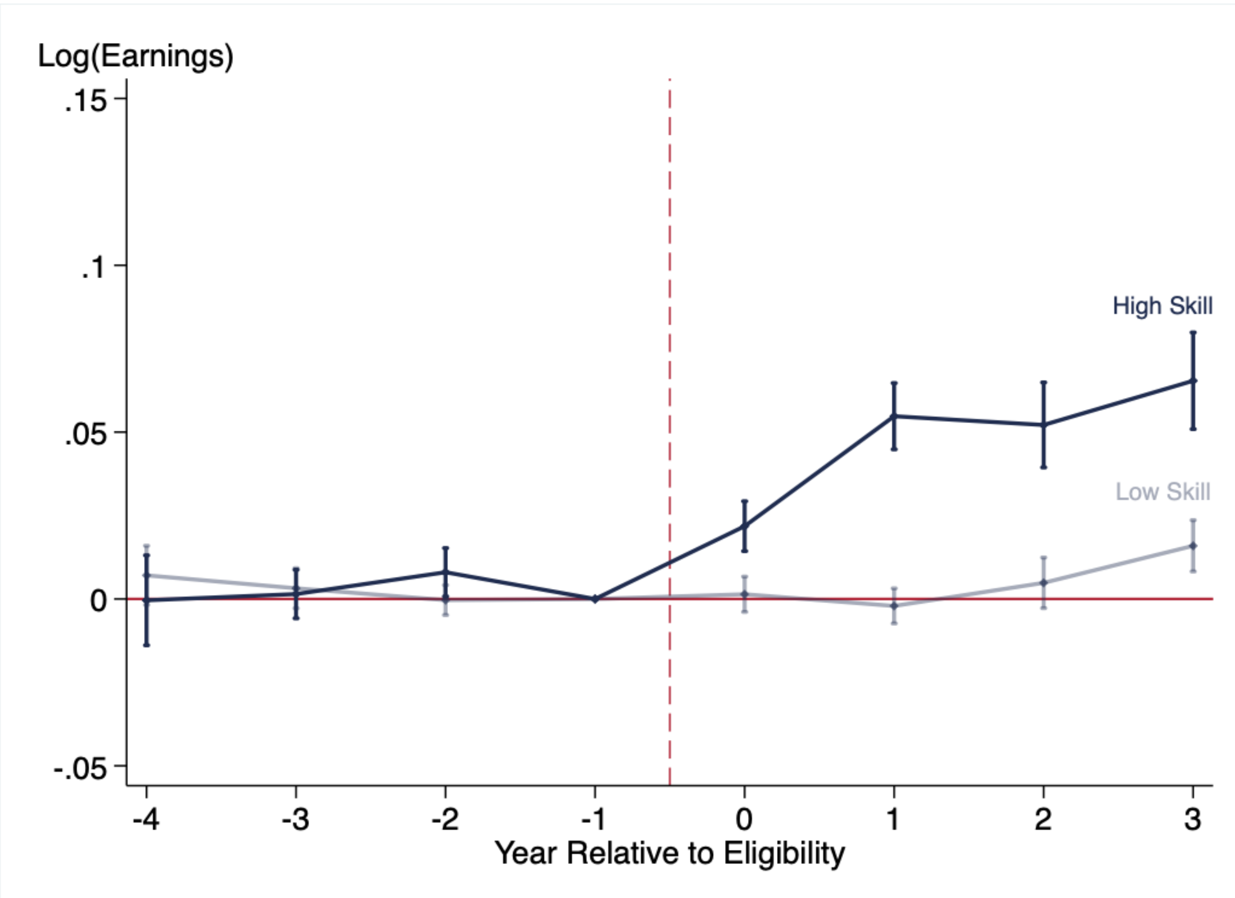
Note: This figure plots event study coefficients for multiple of the firms' margins of adjustment after the payroll tax cut. First, at the top left plot it shows the first stage, i.e., the reform induced a reduction in payroll tax liability. On the top right plot, it depicts the employment increase that has already been documented. The two bottom graphs shed light on other business outcomes, such as capital and profit.

Figure 1.6: Worker Level: Gross Earnings Effect



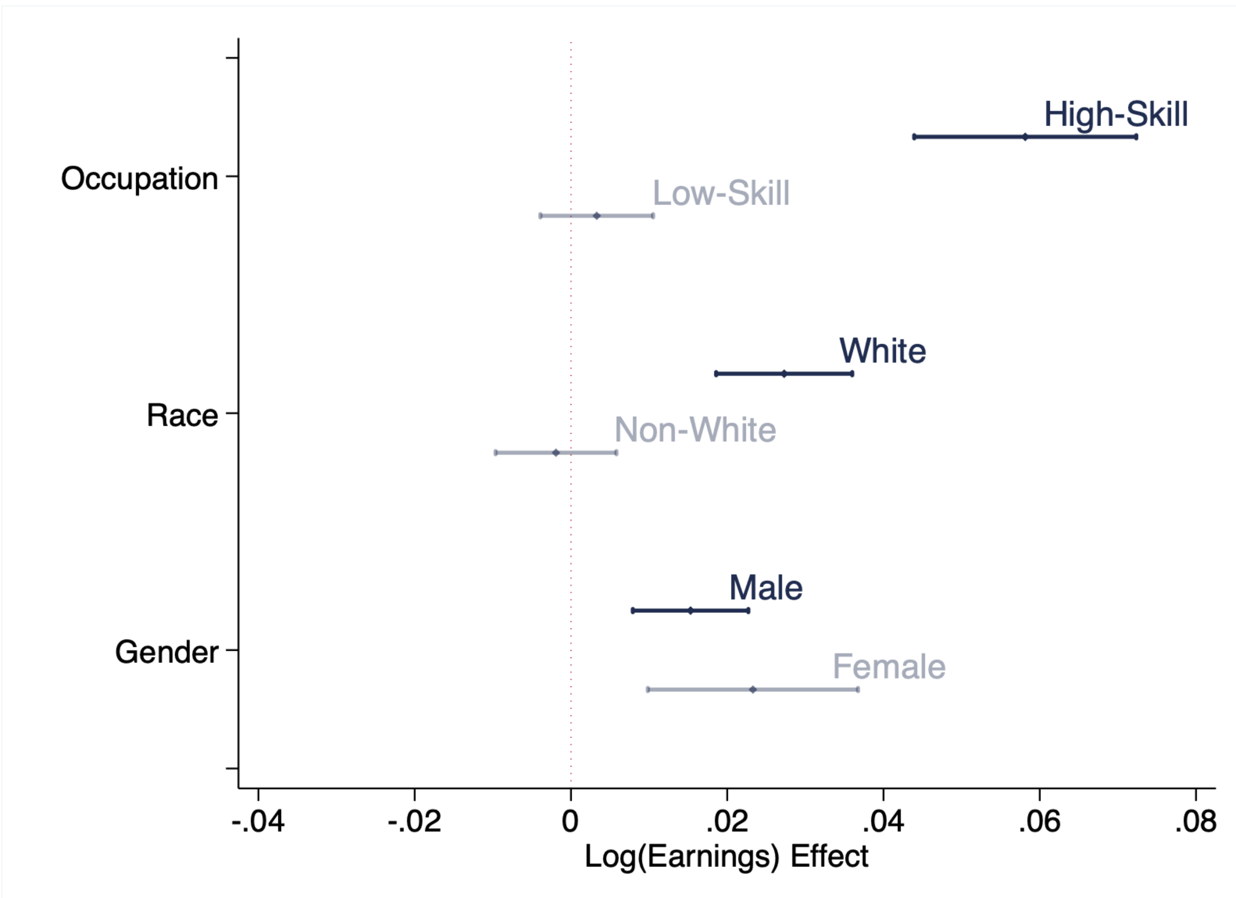
Note: This figure presents the event study estimates for average gross earnings (including payroll taxes) paid workers that were employed for at least three years in the same firm during the pre-reform period. The labor cost is computed using firm-level tax data, and worker-level earnings data. I apply the firm payroll tax rate in year t, to all employees in that firm in year t. I normalize the results with respect to one year prior to the treatment event. The analysis spans four years prior to the payroll tax cut program and three years after. The plot shows an average decrease of $400 on the gross earnings, which has an approximate average of $2,300 during the pre-reform period. The blue markers depict IV coefficients, and the red markers intention-to-treat. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Figure 1.7: Worker Level: Earnings per Occupation



Note: This figure presents the event study estimates for the log of pre-tax earnings per occupation group, at the worker-level based on pre-reform occcupations. Leaders are directors, managers and qualified technical positions according to the CBO classification. While leaders experience high pass-through to earnings of 6%, low-skilled occupation didn't see any significant earnings increase. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Figure 1.8: Heterogeneities by Worker Type



Note: This figure presents the IV difference-in-differences coefficient for the earnings effect at the worker-level sample, across many characteristics of interest, such as, occupation, gender and race. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.1: Eligible vs Non-Eligible Sectors

Eligible	Not Eligible
Hotels	Motels
Open television	Cable television
Public bus transportation	School bus and taxi
Eletronic games manufacturing	Toys and other recreative games manufacturing
Internet portals and content providers	News agencies
Trains	Touristic trains
Newspaper, magazine and book printing	Other periodic printing
Maintenance aircraft and vessels	Maintenance aircraft and other transportation modes

Note: This table presents a list of sectors that are displayed in the tax bills as eligible to the payroll tax cut, and compares it with another list of similar sectors that are not included in the tax reform. This is not an exhaustive list, but highlights (i) the similarity across eligibility groups; and (ii) that eligibility was defined at a very granular level.

Table 1.2: Macro Relevance of the Reform

	2012	2013	2014
# Sectors	10	81	124
Share	0.0076	0.0617	0.0944
# Firms	20,865	33,705	49,253
Share	0.0079	0.0121	0.0170
# Workers	2,950,925	5,028,078	6,113,091
Share	0.0304	0.0513	0.0618

Note: This table shows the comprehensiveness of the policy rollout over the years that new sectors gained eligibility (2012-2014). In the first part of the table it shows the number of 7-digit sectors eligible for the tax reform, and their representativeness computed as the share of existing sectors in the Brazilian economy. The second part of the table shows the number of formal firms in the final sample that were treated in each year. To adjust for informal firms that do not appear in my sample, I multiply the share by 0.55, which is the average formalization rate in Brazil, according to PNAD (official survey administered by the Brazilian Census Bureau, IBGE). In the last rows, the table reports the number of workers employed in treated firms. I compute the share of treated workers by dividing # of workers by the universe of Brazilian workers according to PNAD-C.

Table 1.3: Worker Level Estimates

Worker Level	**Log(Earnings)**	**Log(Earnings)**	
	All Sample (1)	Blue Collar (2)	White Collar (3)
Panel B: IV			
Diff-in-Diff	.018** (.007)	.003 (.007)	.058*** (.014)
Long Diff	.027*** (.007)	.016** (.008)	.064*** (.014)
Panel A: OLS			
Diff-in-Diff	.009** (.004)	.002 (.004)	.031*** (.008)
Long Diff	.017*** (.005)	.01* (.005)	.044*** (.01)
Controls	✓	✓	✓
Worker FE	✓	✓	✓
Firm FE	✓	✓	✓
Sector x Year FE	✓	✓	✓
# Clusters	10,458	10,309	8,938
N	112,621,077	84,007,708	25,118,914

Note: This table presents IV and reduced form (ITT) estimates for the worker-level sample. Difference-in-differences coefficient is estimated in equations 1.3 and 1.4, where there is only one post-period. The long difference comes from the period t=+3, in the event study design. Panel A reports the IV coefficients, which adjust for the imperfect compliance and are interpreted as the local average treatment effect on compliers. Panel B reports the reduced form coefficients, which are interpreted as the intention to treat (ITT). The dependent variable is log of workers' earnings. Column (1) presents the average effect in the all sample. Columns (2-6) present heterogeneity based on pre-reform occupation. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.4: Descriptive Statistics

	(1) Non-Eligible	(2) Eligible	(3) Pooled
Pre-Reform Characteristics			
Employment	53.17 (1,016.13)	55.47 (328.68)	53.31 (988.34)
Payroll Tax Rate	31.78 (13.14)	31.61 (14.50)	31.77 (13.23)
Share Male	0.55 (0.40)	0.77 (0.29)	0.56 (0.40)
Age	37.17 (8.91)	35.89 (7.69)	37.10 (8.85)
Share High School +	0.52 (0.41)	0.59 (0.37)	0.52 (0.41)
Share White	0.68 (0.38)	0.74 (0.33)	0.68 (0.37)
Share Blue Collar	0.89 (0.24)	0.85 (0.27)	0.89 (0.24)
N	1,759,400	112,780	1,872,180

Note: This table presents descriptive statistics of the baseline firm-level sample in the prereform period (2008 to 2011). Each observation is a firm x year. The descriptive statistics are presented for different groups of interest. Column (1) and (2) reports the pre-reform values for non-eligible and eligible firms, respectively. Column (3) pools these two groups together. Column (4) reports the value for eligible firms that eventually take-up the treatment. The variable "Payroll Tax Rate" informs the average payroll tax rates in (%). The variable "High School +" reports the share of workers that achieved high school education or higher. The variable "Share Male" reports the share of male workers. The variable "Share White" informs the average share of white workers per firm. Standard deviations are presented in parentheses.

Table 1.5: Informality Analysis

	(1) Log(1+τ)	(2) Log(#Employees)	(3) Log(Earnings)
Panel A: Low Informality Areas			
Diff-in-Diff	-0.133***	0.135***	0.025*
	(0.004)	(0.039)	(0.014)
Long Diff	-0.121***	0.204***	0.018
	(0.005)	(0.035)	(0.015)
Panel B: High Informality Areas			
Diff-in-Diff	-0.131***	0.031	-0.003
	(0.004)	(0.031)	(0.011)
Long Diff	-0.116***	0.011	0.022*
	(0.006)	(0.043)	(0.012)
Panel C: High Education Firms			
Diff-in-Diff	-.135***	.201***	.03**
	(.004)	(.034)	(.014)
Long Diff	-.119***	.216***	.038***
	(.005)	(.038)	(.014)
Panel D: Low Education Firms			
Diff-in-Diff	-0.129***	0.008	0.003
	(0.004)	(0.033)	(0.013)
Long Diff	-0.121***	0.004	0.010
	(0.006)	(0.039)	(0.012)
Controls	✓	✓	✓
Firm FE	✓	✓	✓
Sector x Year FE	✓	✓	✓
# Clusters	9,548	9,953	9,953
N	3,908,467	4,225,726	4,225,726

Note: This table reports results from the informality analysis, showing that effects are concentrated in low informality regions, and firms employing relatively more educated workforce, which are settings less prone to informality. Panel A presents results for low informality municipalities, which are defined as the bottom 50% of the informality distribution. Panel B presents results for high informality areas. Panel C presents results for firms that employ relatively more educated workers, which are defined as above the median, while Panel D presents results for below median on average education. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.6: Heterogeneity Across Liquidity Constraints

	(1)	(2)
	Employment Low Liquidity	Employment High Liquidity
Currently Treated	0.107***	0.109***
	(0.0283)	(0.0289)
Observations	228,087	233,691
Firm FE	Yes	Yes
Sector (1 digit) x Year FE	Yes	Yes
Worker FE	No	No

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: This table reports IV difference-in-differences coefficients for firms below/above the median on liquidity constraint, during the pre-reform period. Liquidity constraint is defined as the ratio of current assets over current liabilities. An example of current assets is cash, whereas an example of current liabilities is short term bills, such as the wage bill. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.7: Within-Firm Earnings Inequality

	Log(Earnings)				Occup Pctile
	firm (99p) (1)	firm (90p) (2)	firm (40p) (3)	firm (20p) (4)	firm level (5)
Panel A: IV					
Diff-in-Diff	.041***	.022	.01	.003	.001
	(.016)	(.013)	(.011)	(.01)	(.002)
Long Diff	.068***	.038***	.012	-.003	.005
	(.016)	(.013)	(.011)	(.011)	(.003)
Controls	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓
Sector x Year FE	✓	✓	✓	✓	✓
# Clusters	10,679	10,679	10,679	10,679	10,674
N	4,234,882	4,234,882	4,234,882	4,234,882	4,232,627

Note: This table presents IV estimates for the firm-level sample. Difference-in-differences coefficient is estimated in equations 1.3 and 1.4, where there is only one post period. The long difference comes from the period t=+3, in the event study design. Column (1)-(4) reports the earnings effect at different percentiles of the within-firm distribution, indicating that the pass-through predominantly affects employees at the higher end of the spectrum. Column (5) reports zero effect on the average occupation percentile. Occupations are ranked based on average earnings during the years prior to the reform. After each occupation has been allocated to a specific percentile, we calculate, for each t, the mean occupation percentile that firms are employing from. The zero occupation response reinforces that the within-firm inequality response is not driven by an upscale in employed occupations. It also reinforces that the tax cut did not induce a structural change in the production process at the firm-level. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.8: Heterogeneity on Structural Parameters

	(1)	(2)	(3)
Structural Estimates	**Baseline**	**Small Firms**	**Large Firms**
Labor Supply Elasticity, ϵ	4.15	5.75	4.25
	(0.20)	(0.33)	(0.28)
Labor-Capital Elasticity, σ_{KL}	1.72	5.01	1.25
	(0.08)	(0.34)	(0.08)
Output Demand Elasticity, η	1.43	5.21	1.10
	(0.07)	(4.21)	(0.06)
Empirical Estimates			
Employment effect, β_L	0.12	0.35	0.09
Capital effect, β_K	-0.03	0.01	-0.03
Earnings effect, β_W	0.03	0.06	0.02
Revenue effect, β_R	0.05	0.18	0.02
Cost Shares			
Labor	0.80	0.80	0.80
Capital	0.20	0.20	0.20
J-test			
Overid test (pvalue)	0.74	0.17	0.79
Observations			
N	450,387	184,924	265,452

Notes: This table presents the parameters estimated, according to the method presented in Section 1.5. Column (1) reports results for the baseline case, which includes all firms. Columns (2) and (3) restrict the analysis to small and large firms, respectively. Firm size is measured in the pre-reform years, and small/ large are defined based on below/ above the median, respectively. The empirical estimates are presented in Section 1.3. At the bottom, the table displays the cost shares, number of observations, and p-values associated with the J-test for overidentification.

1.9 Details on the Empirical Model

Derivation of the Reduced Form Equations

Given the set of k first stage equations, the reader might not be able to see immediately the reduced form equation. Starting with the firm-level design, we obtain the reduced form by substituting all first stage equations into the second stage,

$$Y_{jt} = \sum_{k=-4,\neq-1}^{3} \beta_k \left[\sum_{l=-4,\neq-1}^{3} \pi_{kl} \times \mathbb{I}(t = e_{s(j)} + l) \times L_{s(j)} + \alpha_j + \xi_{I(j),t} + X'_{jt}\delta_k + \eta_{jt} \right] + \\ X'_{jt}\gamma + \alpha_j + \xi_{I(j),t} + \epsilon_{jt}$$

where, $D^k_{jt} = 1$, if $t = e_j + k$; e_j is the year when firm j enters treatment; $L_{s(j)}$ indicates if firm j's sector is eventually eligible; $e_{s(j)}$ is the date when firm j's sector becomes eligible; X_{jt} set of controls such as education, race, age and its square; $\xi_{I(j),t}$ is industry (broader than sector) x year fixed effect; α_j is the firm fixed effect; η_{jt} and ϵ_{jt} are residuals. Standard errors are conservatively clustered at the 5-digit industry-by-state level. Reorganizing terms,

$$Y_{jt} = \sum_{l=-4,\neq-1}^{3} \left[\sum_{k=-4,\neq-1}^{3} \beta_k \pi_{kl} \times \mathbb{I}(t = e_{s(j)} + l) \times L_{s(j)} \right] + X'_{jt}\left[\gamma + \sum_{k=-4,\neq-1}^{3} \beta_k \delta_k \right] + \\ + (\alpha_j + \xi_{I(j),t}) \left[1 + \sum_{k=-4,\neq-1}^{3} \beta_k \delta_k \right] + \left[\epsilon_{jt} + \sum_{k=-4,\neq-1}^{3} \beta_k \eta_{jt} \right]$$

Thus, the reduced form coefficient is,

$$\rho_l = \sum_{k=-4,\neq-1}^{3} \beta_k \pi_{kl}$$

Note that if K=L and diagonal is such that $\pi_{kl} = 0$ (when k $\neq$ l), then $\rho_l = \beta_l \pi_{ll}$, and $\beta_l = \frac{\rho_l}{\pi_{ll}}$. However, if K$<$ L then the system $\rho_l = \sum_{k=-4,\neq-1}^{3} \beta_k$ for l=1,...., L is a system of L equations in K<L unknowns and generally cannot be solved. The off diagonal coefficients estimated in equation (1.2) are small and not statistically different than zero, which makes the interpretation of the reduced form coefficients equal to the one dimensional case, i.e., $\rho_l = \beta_l \pi_{ll}$. At the worker-level, the algebra to obtain the reduced form coefficient is analogous to the firm-level computations presented in this appendix.

Characterizing Compliers

Section 1.3 stresses that the causal interpretation for the LATE is restricted to the set of compliers. Oftentimes, compliers are not representative of the population, therefore it is useful to have a deeper understanding of who the compliers are. The challenge is that different from *always-takers* and *never-takers* compliers' characteristics are not observationally identified. Even though it is observable if an eligible firm took up treatment, it is not observable if the take-up decision is because the firm is an *always-taker* or *complier*. This comes from the fact that the counterfactual decision (what an eligible firm would do if it were not to be eligible) is not observable in the data.

Abadie, 2002 proposes a 2SLS approach to detect compliers. This method relies on the fact that *never-takers* (eligible firms that do not take-up) and *always-takers* (ineligible firms that take-up) are observable. Concretely, it estimates the pair of regressions:

$$X_{jt} \times \mathbb{I}_{D_j=d} = \alpha_d + \gamma_d \mathbb{I}_{D_j=d} + \nu_{jtd} \tag{1.1}$$

$$\mathbb{I}_{D_j=d} = \zeta_d + \pi_d L_{s(j)} + \eta_{jtd} \tag{1.2}$$

where X_{jt} is a vector of firm's characteristics at the baseline; d ={0,1} indicates if $L_{s(j)}$ is instrumenting eventual treatment or never treatment; and α_d, ζ_d are constants. The IV coefficients for d ={0,1} recover average characteristics for never and eventually treated compliers, respectively. To obtain baseline characteristics for *never-takers* I regress $X_{jt}(1-D_j)L_{s(j)}$ on $(1-D_j)L_{s(j)}$. Finally, the characterization of *always-takers* comes from regressing $X_{jt}D_j(1-L_{s(j)})$ on $D_j(1-L_{s(j)})$. Table 1.1 reports results for the same regressions when we incorporate the 1-digit sector x year dummies and set of controls that are included in the main specification.[19] The table shows that covariates' means for treated and untreated compliers are not statistically distinguishable between each other. As Angrist, Hull, and Walters, 2022 point out, the balance check across compliers is equivalent to the hidden complier RCT embedded in the treatment assignment with imperfect compliance. Comparisons to the remaining columns showcase that *always-takers* are larger firms, and *never-takers* are smaller firms compared to compliers.

[19]The interpretation of coefficients is compliers' weighted average characteristics within sector x year cells.

Table 1.1: Compliers' Characteristics

	Compliers			
	Untreated	Treated	Always-Takers	Never-Takers
	(1)	(2)	(3)	(4)
Employment	108.54	103.69	188.94	27.1
	(29.592)	(16.237)	(16.86)	(2.692)
Payroll Tax Rate	.33	.35	.35	.29
	(.011)	(.002)	(.002)	(.004)
Share Male	.73	.73	.71	.74
	(.02)	(.02)	(.017)	(.022)
Age	35.17	35.17	33.36	36.68
	(.326)	(.326)	(.087)	(.354)
High School +	.58	.6	.57	.58
	(.026)	(.022)	(.007)	(.022)
White	.75	.75	.76	.71
	(.025)	(.025)	(.007)	(.027)
Blue Collar	.8	.8	.92	.86
	(.018)	(.018)	(.003)	(.016)
Sector x Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Note: This table reports baseline estimates characteristics of compliers, always-takers and never-takers in the context of the Brazilian tax reform. Values for each covariate are computed in the pre-reform period at the firm x year level, and the regressions include 1-digit sector x year fixed effects and set of controls considered in the main specification (Section 1.3). Standard errors are reported in parentheses and conservatively clustered at the 5-digit industry-by-state level.

1.10 Model

In this appendix, I present the model derivation. For didactic purposes, I start by analyzing the revenue and payroll taxes, separately. In the end, I put both taxes together to map the structural equations to the reduced form estimates.

Microfounding the Labor Supply

As in Card et al., 2018, workers exhibit idiosyncratic preferences for employers. These preferences can be understood through non-pecuniary match factors such as corporate culture and commuting distance. Unlike traditional search models, this approach posits that wage-posting behavior induces firms to pay identical wages to all workers of the same quality. Upon meeting the requisite quality standards, a firm hires any worker willing to accept the posted wage. In this scenario, worker i is fully knowledgeable of available job opportunities, and derives the following utility from working at firm j:

$$u_{ij} = \epsilon \ln(w_j - b) + a_j + \nu_{ij}$$

where, w_j is the wage level paid by firm j, b is the competitive wage level defined by the workers' outside option, a_j is a firm-specific amenity, and ν_{ij} is the idiosyncratic preference for worker i to be at firm j. Assuming that ν_{ij} comes from an extreme type I distribution, I follow McFadden et al., 1973 to compute the logit probabilities to work at firm j:

$$p_j = \frac{\exp(\epsilon \ln(w_j - b) + a_j)}{\sum_{k=1}^{J} \exp(\epsilon \ln(w_k - b) + a_k)}$$

If the total number of firms J is large enough, the logit probabilities can be approximated by exponential probabilities of the form,

$$p_j = \lambda \exp(\epsilon \ln(w_j - b) + a_j)$$

where λ is a constant common to all firms in the market. Therefore, for large J, we can write the firm-specific supply function as:

$$\ln L_j(w_j) = \ln \mathbb{L}\lambda + \epsilon \ln(w_j - b) + a_j$$

where $\mathbb{L}$ represents the total number of workers in the market. Taking exponential transformations on both sides, we can compute the labor supply function:

$$L_j = \exp(\epsilon \ln(w_j - b)) \exp(a_j) \exp(\lambda \mathbb{L}) \iff \underbrace{L_j^{\frac{1}{\epsilon}} \exp\left(\frac{-\mathbb{L}\lambda - a_j}{\epsilon}\right)}_{\equiv A_j} = (w_j - b)$$

As b $\to 0$, then

$$w_j = A_j L_j^{\frac{1}{\epsilon}} \tag{1.1}$$

In this case, ϵ is the constant labor supply elasticity faced by the firm.

Effects of Payroll Taxation

The labor supply function gives rise to the cost function faced by firms,

$$C = A((1+\tau))L^{\frac{1}{\epsilon}+1} + rK$$

Production function exhibits constant returns to scale, and the firm faces demand at the product market given by, $P = Q^{\frac{-1}{\eta}}$. The firm solves two related problems. First, it chooses plant size to maximize profit. Second, for a given plant size (Q), it chooses inputs of production (L and K) to minimize costs, according to the following program:

$$\begin{aligned} \min_{K,L} \quad & A(1+\tau)L^{\frac{1}{\epsilon}+1} + rK \\ \text{s.t.} \quad & f(K,L) \geq Q \end{aligned} \tag{1.2}$$

Summing and rearranging the optimality conditions, I obtain the cost function:

$$C = \underbrace{\lambda(w,r,Q)}_{=\frac{\partial C}{\partial Q}} \underbrace{(Lf_L + Kf_K)}_{=Q} - \underbrace{A(1+\tau)\frac{1}{\epsilon}L^{1+\frac{1}{\epsilon}}}_{new} \tag{1.3}$$

Differently from the perfectly competitive labor market, under monopsony average and marginal cost no longer align. Lemma 1 proves this point.

Lemma 1. *In a perfectly competitive labor market, the marginal cost of production is constant in the quantity Q.*

Proof. From FOC,

$$C(w,r,Q) = \lambda(w,r,Q)Q \iff C(w,r,\alpha Q) = \lambda(w,r,\alpha Q)\alpha Q$$

From constant returns,

$$C(w,r,\alpha Q) = \alpha C(w,r,Q) = \alpha\lambda(w,r,Q)Q$$
$$\lambda(w,r,\alpha Q)\alpha Q = \lambda(w,r,Q)\alpha Q \Rightarrow \lambda(w,r,Q) = \lambda(w,r)$$

$\square$

The profit maximizing firm chooses output Q,

$$\max_Q P(Q)Q - c(Q,\tau)Q + \frac{1}{\epsilon}A(1+\tau)L^{1+\frac{1}{\epsilon}}$$

At the optimal, marginal cost and marginal revenue are equated:

$$\left(\frac{\eta-1}{\eta}\right)Q^{\frac{-1}{\eta}} = \lambda(Q,\tau) \tag{1.4}$$

To evaluate the policy induced scale effect, I take logs and differentiate with respect to the labor cost $(1+\tau)$,

$$\epsilon^{Q}_{1+\tau} = \frac{-\epsilon^{\lambda}_{1+\tau}}{\left(\frac{1}{\eta}+\epsilon^{\lambda}_{Q}\right)} \tag{1.5}$$

Also note that from 1.4,

$$P\left(\frac{\eta-1}{\eta}\right) = \lambda \iff \frac{\partial \log P}{\partial \log(1+\tau)} = \frac{\partial \log \lambda}{\partial \log(1+\tau)} = \epsilon^{\lambda}_{1+\tau} + \epsilon^{\lambda}_{Q}\epsilon^{Q}_{1+\tau}$$

$$\frac{\partial \log Q}{\partial \log(1+\tau)} = \underbrace{\frac{\partial \log Q}{\partial \log P}}_{-\eta}\overbrace{\frac{\partial \log P}{\partial \log(1+\tau)}}^{\epsilon^{\lambda}_{1+\tau}+\epsilon^{\lambda}_{Q}\epsilon^{Q}_{1+\tau}}$$

$$\frac{\partial \log Rev}{\partial \log(1+\tau)} = \frac{\partial \log PQ}{\partial \log(1+\tau)} = (1-\eta)(\epsilon^{\lambda}_{1+\tau} + \epsilon^{\lambda}_{Q}\epsilon^{Q}_{1+\tau}) \tag{1.6}$$

Applying the envelope theorem to derive equation (1.3) with respect to $(1+\tau)$,

$$AL^{\frac{1}{\epsilon}+1} = \lambda_{1+\tau}Q - \frac{AL^{1+\frac{1}{\epsilon}}}{\epsilon} - \frac{A(1+\tau)}{\epsilon}\left(\frac{1+\epsilon}{\epsilon}\right)L^{\frac{1}{\epsilon}}\frac{\partial L}{\partial(1+\tau)}$$

$$\frac{\partial \log \lambda}{\partial \log(1+\tau)} = \frac{(1+\tau)AL^{1+\frac{1}{\epsilon}}}{\lambda Q}\left(\frac{\epsilon+1}{\epsilon}\right)\left(1+\frac{(1+\tau)}{\epsilon L}\frac{\partial L}{\partial(1+\tau)}\right) \tag{1.7}$$

Equation (1.7) refers to the elasticity of the marginal cost with respect to the labor cost, which is a key aspect of the incidence analysis. Taking this expression to the data is challenging because we do not observe either λ, or Q. However, by manipulating equation (1.3) and dividing both sides by the total wage bill we obtain,

$$\frac{\lambda Q}{(1+\tau)AL^{1+\frac{1}{\epsilon}}} = \frac{C+(1+\tau)AL^{1+\frac{1}{\epsilon}}(\frac{1}{\epsilon})}{\underbrace{(1+\tau)AL^{1+\frac{1}{\epsilon}}}_{wL}} = \frac{1}{s_L}+\frac{1}{\epsilon} \tag{1.8}$$

The right hand side of equation (1.8) depends on s_L and ϵ. It turns out that we do observe labor share (s_L), and we can estimate ϵ. Plugging 1.8 in 1.7,

$$\epsilon_{1+\tau}^{\lambda} = \left(\frac{1}{\frac{1}{s_L} + \frac{1}{\epsilon}}\right)\left(\frac{\epsilon+1}{\epsilon}\right)\left(1 + \frac{\epsilon_{1+\tau}^{L}}{\epsilon}\right) \tag{1.9}$$

Equation (1.9) shows that the effect of the labor cost on the marginal cost depends on three components. First, is the monopsony-adjusted labor share. The more relevant is the labor share, which means that reducing labor costs will have a greater impact on the marginal cost. Second, is the inverse markdown. The intuition for this term is that as labor market power increases, there is more rents to be shared with incumbent workers when the firm expands plant size. Finally, the last term says that the pass-through to marginal cost is directly affected by the pass-through to the marginal cost of labor. Differentiating both sides of equation (1.3) by Q, after some manipulation I obtain,

$$\epsilon_{Q}^{\lambda} = \left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{\epsilon_{Q}^{L}}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L} + \frac{1}{\epsilon}}\right) \tag{1.10}$$

Note that,

$$\epsilon_{1+\tau}^{L} = \frac{\partial \log L}{\partial \log(1+\tau)} = \frac{\partial \log L}{\partial \log Q}\frac{\partial \log Q}{\partial \log(1+\tau)}$$

$$\epsilon_{Q}^{L} = \frac{\epsilon_{1+\tau}^{L}}{\epsilon_{1+\tau}^{Q}} \tag{1.11}$$

Using 1.5 in 1.11,

$$\epsilon_{Q}^{L} = \frac{-\epsilon_{1+\tau}^{L}(\frac{1}{\eta} + \epsilon_{Q}^{\lambda})}{\epsilon_{1+\tau}^{\lambda}} \tag{1.12}$$

Now, 1.12 and 1.9 in 1.10,

$$\epsilon_{Q}^{\lambda} = \frac{-\epsilon_{1+\tau}^{L}}{\eta(2\epsilon_{1+\tau}^{L} + \epsilon)} \tag{1.13}$$

To compute $\epsilon_{1+\tau}^{Q}$ substitute 1.9 and 1.13 in 1.5,

$$\epsilon_{1+\tau}^{Q} = -\left(\frac{\epsilon+1}{\epsilon}\right)\left(1 + \frac{\epsilon_{1+\tau}^{L}}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L} + \frac{1}{\epsilon}}\right)\left(\frac{\eta(\epsilon + 2\epsilon_{1+\tau}^{L})}{\epsilon + \epsilon_{1+\tau}^{L}}\right) \tag{1.14}$$

To compute the tax reduction pass-through to employment and capital, I can differentiate optimal choices in 1.2 with respect to the labor cost ($(1+\tau)$):

$$\epsilon_{1+\tau}^{L} = \frac{\epsilon}{1 - \epsilon\rho + \epsilon}(\epsilon_{1+\tau}^{\lambda} + \epsilon_{Q}^{\lambda}\epsilon_{1+\tau}^{Q} - 1) + \left(\frac{(1-\rho)\epsilon}{1 - \epsilon\rho + \epsilon}\right)\epsilon_{1+\tau}^{Q}$$

Plugging 1.9, 1.13 and 1.14, I obtain the model's prediction for the pass-though to employment, in terms of observables and parameters to be estimated:

$$\epsilon_{1+\tau}^{L}=\left(\frac{\epsilon}{1+\epsilon(1-\rho)}\right)\left[\left(\frac{(\epsilon+2\epsilon_{1+\tau}^{L})(\sigma-\eta)}{\sigma\epsilon}\right)\left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)-1\right] \tag{1.15}$$

Recall, that the elasticity of employment with respect to labor cost $\epsilon_{1+\tau}^{L}$ I empirically estimate in the reduced form analysis. The remaining structural parameters are jointly estimated in Section 1.5. Similarly, I can find equations for the pass-through to capital, and revenue.

$$\epsilon_{1+\tau}^{K}=\left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\left(\frac{\epsilon+2\epsilon_{1+\tau}^{L}}{\epsilon}\right)\left(\sigma-\eta\right) \tag{1.16}$$

$$\epsilon_{1+\tau}^{R}=(1-\eta)\left[\left(\frac{\epsilon+1}{\epsilon}\right)\left(\frac{\epsilon+2\epsilon_{1+\tau}^{L}}{\epsilon}\right)\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\right] \tag{1.17}$$

Taking logs and differentiating the labor supply function,

$$\beta_W=\frac{\epsilon_{1+\tau}^{L}}{\epsilon}\phi_1 \tag{1.18}$$

Effects of Revenue Taxation

Under revenue taxation (τ_r), the firm solves the following program in the product market:

$$\max_{Q} P(Q)Q-\frac{C(Q)}{1-\tau_r}$$

The firm equates marginal revenue to marginal cost,

$$\left(\frac{\eta-1}{\eta}\right)Q^{\frac{-1}{\eta}}=\frac{\lambda(Q)}{1-\tau_r}$$

where the right-hand side is a direct application of the envelope theorem on the cost minimization problem. The plant size has a direct implication on prices through demand, so if we take logs and differentiate with respect to $\log\tau_r$,

$$\frac{\partial\log P}{\partial\log\tau_r}=\frac{\tau_r}{1-\tau_r}$$

I know the relationship between the elasticity of prices and quantity with respect to revenue taxes,

$$\frac{\partial\log P}{\partial\log Q}\frac{\partial\log Q}{\partial\log\tau_r}=\frac{\partial\log P}{\partial\log\tau_r}\iff\frac{\partial\log Q}{\partial\log\tau_r}=-\frac{\tau_r}{1-\tau_r}\eta \tag{1.19}$$

where the $\frac{\partial \log P}{\partial \log Q} = \frac{-1}{\eta}$ is known based on the iso-elastic demand function. The price and quantity responses allow me to compute the effect of revenue taxes on revenue,

$$\epsilon^R_{1+\tau_r} = \frac{\tau_r}{1-\tau_r}(1-\eta)$$

Once firms, choose the plant size, they will choose the inputs mix to minimize cost,

$$\begin{aligned} C(Q) = \min_{K,L} \quad & AL^{\frac{1}{\epsilon}+1} + rK \\ \text{s.t.} \quad & (s_L L^\rho + s_K K^\rho)^{\frac{1}{\rho}} \geq Q \end{aligned}$$

The optimal choices of capital and labor are:

$$L = \left[\left(\frac{\epsilon}{\epsilon+1}\right)\frac{s_L}{A}\lambda(Q)\right]^{\frac{\epsilon}{1-\epsilon\rho+\epsilon}} Q^{\frac{(1-\rho)\epsilon}{1-\epsilon\rho+\epsilon}} \qquad K = \left(\frac{r}{\lambda(Q)s_K}\right)^{\frac{1}{\rho-1}} Q$$

Taking logs and differentiating with respect to $\log \tau_r$, we obtain the revenue tax pass-through to employment and wages,

$$\frac{\partial \log L}{\partial \log \tau_r} = \frac{-\epsilon}{1-\epsilon\rho+\epsilon} + \left(\frac{(1-\rho)\epsilon}{1-\epsilon\rho+\epsilon}\right)\frac{\partial \log Q}{\partial \log \tau_r} + \frac{\epsilon}{1-\epsilon\rho+\epsilon}\left(\frac{\partial \log \lambda(Q)}{\partial \log Q}\frac{\partial \log Q}{\partial \log \tau_r}\right) \tag{1.20}$$

$$\frac{\partial \log K}{\partial \log \tau_r} = \frac{\partial \log Q}{\partial \log \tau_r} - \left(\frac{1}{\rho-1}\right)\left[\frac{\partial \log \lambda(Q)}{\partial \log Q}\frac{\partial \log Q}{\partial \log \tau_r}\right] \tag{1.21}$$

To obtain closed form solution for the pass-through expressions we need to compute the elasticity of marginal cost with respect to quantity ϵ^λ_Q, which we can pin down by differentiating the cost function with respect to Q,

$$\epsilon^\lambda_Q = \left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\left(\frac{\epsilon+1}{\epsilon}\right)\frac{\epsilon^L_Q}{\epsilon} \tag{1.22}$$

Note that,

$$\epsilon^L_Q = \frac{\epsilon_{L\tau_r}}{\epsilon^Q_{\tau_r}} \iff \epsilon^L_Q = \frac{-\epsilon^L_{\tau_r}(1-\tau_r)}{\tau_r\eta} \tag{1.23}$$

Plugging 1.23 in 1.22,

$$\epsilon^L_{\tau_r} = -\left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\left(\frac{\epsilon+1}{\epsilon}\right)\frac{\epsilon^L_{\tau_r}(1-\tau_r)}{\tau_r\eta}$$

Plugging ϵ^λ_Q and $\epsilon^Q_{\tau_r}$ in 1.20 and 1.21, we obtain the closed form pass-through expressions for the revenue taxation,

$$\epsilon^{L}_{\tau_r} = \frac{-(1-\rho)\epsilon}{1+\epsilon(1-\rho-\chi(\epsilon,s_L))}\frac{\tau_r}{1-\tau_r}\eta \quad (1.24)$$

$$, \quad \epsilon^{K}_{\tau_r} = \frac{\tau_r\eta}{1-\tau_r}\left(\frac{-\chi(\epsilon,s_L)\epsilon}{1+\epsilon(1-\rho-\chi(\epsilon,s_L))}-1\right) \quad (1.25)$$

where, I denote $\chi(\epsilon, s_L) = \left(\frac{1}{\frac{1}{s_L}+\frac{1}{\epsilon}}\right)\left(\frac{\epsilon+1}{\epsilon}\right)$ to simplify notation. The elasticity η makes the model versatile to accommodate different degrees of competition in the product market. As η increases, we move to a more competitive product market. At first, we will be agnostic about its value, and let η be determined by the data. For the specific case of the Brazilian tax reform, the revenue tax rate is small (around 1.5%). For this reason, the effects depicted on equations 1.24 and 1.25 are negligible compared to the responses coming from the payroll tax side.

1.11 Deadweight Loss

Payroll taxes depresses wages, profits, and consumption, while increases Government revenue. To compute the efficiency effect of taxation, Equation (1.1) relies on a money metric approach that aggregates the net benefit and costs of payroll taxes.

$$W = \underbrace{wL - \int_0^L Ak^{\frac{1}{\epsilon}} dk}_{\text{worker surplus}} + \underbrace{PQ - wL(1+\tau) - rK}_{\text{firm owner surplus}} + \underbrace{\int_0^Q z^{\frac{-1}{\eta}} dz - PQ}_{\text{consumer surplus}} + \underbrace{wL\tau}_{\text{Gov revenue}}$$

$$W = -\int_0^L Ak^{\frac{1}{\epsilon}} dk + \int_0^Q z^{\frac{-1}{\eta}} dz - rK \tag{1.1}$$

Therefore, the efficiency gain induced by a discrete payroll tax cut can be computed according to the following formula:

$$\Delta W = B\left[\underbrace{\beta_w}_{\text{worker, dw}} + \underbrace{\frac{\beta_\pi s_\pi}{s_L}}_{\text{firm owner, } d\pi} + \underbrace{\frac{\beta_R}{s_L(\eta-1)}}_{\text{consumer, dp}} + \underbrace{\overbrace{(\tau-\tau_0)}^{<0 \text{ (tax cut)}} + \tau_0 \frac{\beta_L(\epsilon+1)}{\epsilon}}_{\text{Governement, dT}}\right] \tag{1.2}$$

Taking Equation 1.2 to the data, we obtain a precise measure of the deadweight loss associated with payroll taxation. To obtain further theoretical intuition about the forces at play on the determinants of the deadweight loss of payroll taxes, I totally differentiate Equation 1.1:

$$dW = -\frac{\partial L}{\partial \tau} AL^{\frac{1}{\epsilon}} + \frac{\partial Q}{\partial \tau} Q^{\frac{-1}{\eta}} - r\frac{\partial K}{\partial \tau}$$

$$dW = \underbrace{\frac{\partial L}{\partial \tau}\left[\overbrace{\frac{w}{\epsilon}}^{\text{Monopsony}} + \overbrace{w\left(\frac{\epsilon+1}{\epsilon}\right)\tau}^{\text{Tax wedge}}\right]}_{\text{labor wedge}} + \underbrace{\left[\frac{\partial Q}{\partial \tau} Q^{\frac{-1}{\eta}} - \frac{\partial L}{\partial \tau}\overbrace{\mu_L}^{\text{MCL}} - \overbrace{\mu_K}^{\text{MCK}=r}\frac{\partial K}{\partial \tau}\right]}_{\text{product wedge}}$$

The product market wedge can be expressed as a function of $\frac{\partial Q}{\partial \tau}$:

$$dW = \frac{\partial L}{\partial \tau}\left[\frac{w}{\epsilon} + w\left(\frac{\epsilon+1}{\epsilon}\right)\tau\right] + \frac{\partial Q}{\partial \tau}\left[Q^{\frac{-1}{\eta}} - \frac{\frac{\partial L}{\partial \tau}}{\frac{\partial Q}{\partial \tau}} a - r\frac{\frac{\partial K}{\partial \tau}}{\frac{\partial Q}{\partial \tau}}\right] \tag{1.3}$$

To compute the ratio of derivatives in equation 1.3, I recall the optimal input choices from the cost minimization problem:

$$\mathbb{L} = A(1+\tau)L^{\frac{1}{\epsilon}+1} + rK + \lambda[Q - (s_L L^\rho + s_K K^\rho)^{\frac{1}{\rho}}]$$

The lagragean multiplier λ is the shadow price of output, and it is equal to the marginal cost of production. The first order conditions are:

$$[L]: \underbrace{\frac{\epsilon+1}{\epsilon}A(1+\tau)L^{\frac{1}{\epsilon}}}_{\text{MCL}} = \underbrace{\lambda}_{\text{Mg Cost}} \underbrace{s_L(s_L L^\rho + s_K K^\rho)^{\frac{1}{\rho}-1}L^{\rho-1}}_{\text{MPL}=\frac{\partial Q}{\partial L}}$$

Therefore,

$$\frac{1}{\frac{\partial Q}{\partial L}} = \frac{\lambda}{\mu_L} \tag{1.4}$$

$$[K]: \underbrace{r}_{\text{MCK}} = \underbrace{\lambda}_{\text{Mg Cost}} \underbrace{s_L(s_L L^\rho + s_K K^\rho)^{\frac{1}{\rho}-1}K^{\rho-1}}_{\text{MPK}=\frac{\partial Q}{\partial K}}$$

$$\frac{1}{\frac{\partial Q}{\partial K}} = \frac{\lambda}{MCK} = \frac{\lambda}{r} \tag{1.5}$$

Given that Q depends on K and L, I can write the derivative of Q as:

$$\frac{\partial Q}{\partial \tau} = \frac{\partial Q}{\partial L}\frac{\partial L}{\partial \tau} + \frac{\partial Q}{\partial K}\frac{\partial K}{\partial \tau}$$

$$\frac{\partial L}{\partial \tau}/\frac{\partial Q}{\partial \tau} = \frac{\partial L}{\partial \tau}/\left(\frac{\partial Q}{\partial L}\frac{\partial L}{\partial \tau} + \frac{\partial Q}{\partial K}\frac{\partial K}{\partial \tau}\right) = \frac{\partial L}{\partial \tau}/\left(\frac{\mu_L}{\lambda}\frac{\partial L}{\partial \tau} + \frac{r}{\lambda}\frac{\partial K}{\partial \tau}\right) \tag{1.6}$$

$$\frac{\partial K}{\partial \tau}/\frac{\partial Q}{\partial \tau} = \frac{\partial K}{\partial \tau}/\left(\frac{\partial Q}{\partial L}\frac{\partial L}{\partial \tau} + \frac{\partial Q}{\partial K}\frac{\partial K}{\partial \tau}\right) = \frac{\partial K}{\partial \tau}/\left(\frac{\mu_L}{\lambda}\frac{\partial L}{\partial \tau} + \frac{r}{\lambda}\frac{\partial K}{\partial \tau}\right) \tag{1.7}$$

where the last equalities in 1.6 and 1.7 come from the optimal input choices, as depicted in equations 1.4 and 1.5. Plugging 1.6 and 1.7 back into 1.3:

$$dW = \frac{\partial L}{\partial \tau}\left[\frac{w}{\epsilon} + w\left(\frac{\epsilon+1}{\epsilon}\right)\tau\right] + \frac{\partial Q}{\partial \tau}\left[\frac{Q^{\frac{-1}{\eta}}}{\eta}\right] \tag{1.8}$$

The first term in Equation 1.8 depicts the deadweight loss originated from the labor market, while the second term depicts the deadweight loss originated from the product market. The terms outside the brackets are the behavioral responses, which capture the deadweight loss from quantity distortions. The terms inside the first bracket are the monopsony wedge, and the payroll tax wedge, respectively. The term inside the second bracket captures the monopoly wedge due to price markup. If we take the limit of ϵ and η to infinity, Equation 1.8 reduces to the standard textbook deadweight loss formula.

1.12 Revenue Maximizing Payroll Tax Rate

The payroll tax rate variation induces a mechanical change on Governments' revenue, measured by $dM = B(\tau_1 - \tau_0)$. It also induces a behavioral response to tax revenue given the labor supply responses induced by the policy: $H = \tau B(\frac{\epsilon+1}{\epsilon})\frac{\partial L}{\partial 1+\tau}\frac{1}{L}$. To compute the behavioral effect we rely on the empirically estimated employment response ($\frac{\partial L}{\partial 1+\tau}$). Note that this response is locally estimated. To extrapolate the counterfactual employment response at hypothetical tax rates far from the observed level, I undertake a Taylor expansion, with rates varying from τ_0 to τ_1:

$$\frac{\partial L}{\partial 1+\tau}(1+\tau_1) = \frac{\partial L}{\partial 1+\tau}(1+\tau)\Big|_{\tau=\tau_0} + \frac{\partial L}{\partial^2 1+\tau}(1+\tau)\Big|_{\tau=\tau_0}(\tau_1-\tau_0) + \frac{1}{2}\frac{\partial L}{\partial^3 1+\tau}(1+\tau)\Big|_{\tau=\tau_0}(\tau_1-\tau_0)^2 + ... \quad (1.1)$$

$$\frac{\partial L}{\partial 1+\tau}(1+\tau_1) = \frac{L\epsilon_{L,1+\tau}}{1+\tau}\Big|_{\tau=\tau_0}\Bigg[1 + \frac{(\epsilon_{L,1+\tau}-1)}{1+\tau}\Big|_{\tau=\tau_0}(\tau_1-\tau_0) + \frac{1}{2(1+\tau)^2}\Big|_{\tau=\tau_0}(\epsilon_{L,1+\tau}(\epsilon_{L,1+\tau}-1)+2)(\tau_1-\tau_0)^2\Bigg] \quad (1.2)$$

In counterfactual scenarios, where the payroll tax rate moves to τ_1, I compute the behavioral response by evaluating $dH = \tau B(\frac{\epsilon+1}{\epsilon})\frac{\partial L}{\partial 1+\tau}\frac{1}{L}$ at the counterfactual employment response delineated in Equation 1.2. With this framework, I simulate the revenue impact of perturbing the labor tax rate. Figure 1.1 presents a shape similar to the so-called, *Laffer curve*, and shows that the Brazilian tax revenue would be maximized if the labor tax rate were 56%.[20]

Figure 1.1 plots the net effect of two opposing forces at play when the payroll tax rate is increased. On one hand, the mechanical effect increases tax revenue. On the other hand, the behavioral response decreases tax revenue, as the tax rate is increased. This curve illustrates the zone where the mechanical effect outweighs the behavioral response, thereby enabling us to visually observe the revenue maximizing rate.

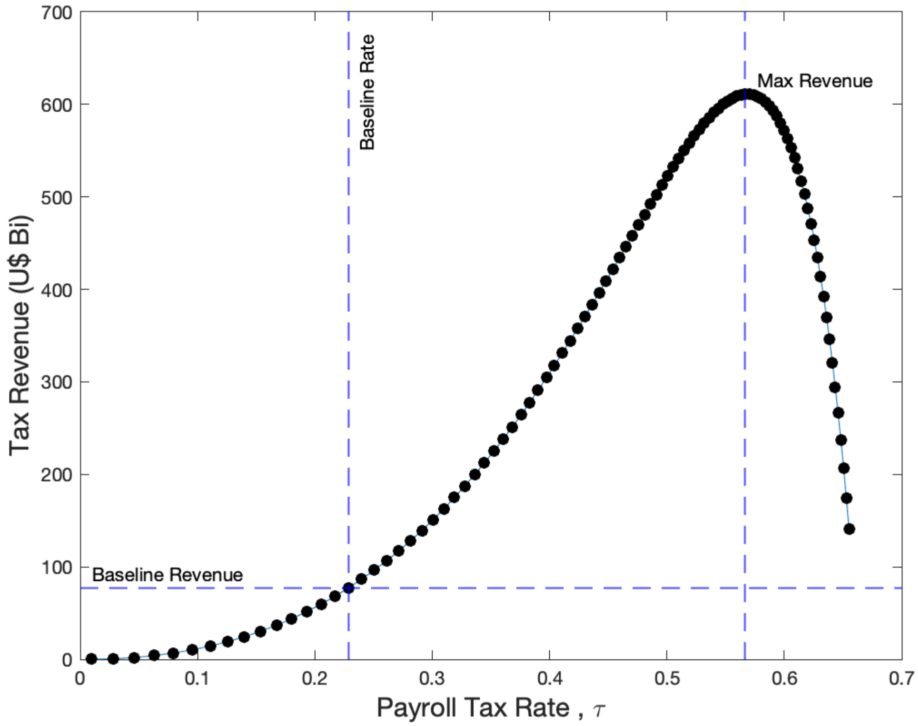
One immediate takeaway from this exercise is that payroll tax rates in Brazil are fairly far from the revenue-maximizing rate, which is indicative that existing average tax rates

[20] Alternatively, this could be expressed as a firm's payroll tax rate of 130%. There is a one-to-one relationship between the firms' and workers' take-home tax rates:

$$\underbrace{\frac{wL}{1-t}}_{\text{Received by worker}} = \underbrace{wL(1+\tau)}_{\text{Paid by firm}}$$

are on the "right side of the Laffer curve". The direct consequence is that Brazilian policymakers can increase the payroll tax without fearing a decline in tax revenue. This conclusion is further supported by the positive MVPF reported in Table 1.5.

Figure 1.1: Laffer Curve for Payroll Taxation



Note: This figure plots the "Laffer curve" for the Brazilian payroll tax system. As we simulate increases in the payroll tax rate, there are two opposing forces: mechanical and behavioral effects. When payroll tax rates are increased, the behavioral response prompts a drop in revenue as a result of adjustments in labor supply. This curve illustrates the zone where the mechanical effect outweighs the behavioral response, thereby enabling us to visually observe the revenue maximizing rate.

1.13 Capital-Skill Complementarity

Inequality in modern society is not only persistent, but it has also risen over time, a concern emphasized by Saez and Zucman, 2019. An array of recent research, including studies by Katz and Murphy, 1992 and Autor, Goldin, and Katz, 2020, explores this escalating phenomenon through the perspective of capital-skill complementarity. This theory suggests that capital and skilled labor are complementary inputs, with technological advancements increasingly benefiting skilled workers. To examine the plausibility of this theory, I leverage the quasi-experimental payroll tax variation in an extension of the model that includes two types of labor.

Extended Model. Consider two types of workers, say high (L_h) and low skill (L_l). Consequently, a firm's production decisions are now based on three inputs: high-skilled labor (L_h), low-skilled labor (L_l), and capital (K). We maintain the constant elasticity of substitution (CES) technology with constant returns but introduce an additional nesting layer to the model.

$$f = (s_{hl}(s_h L_h^{\rho} + s_l L_l^{\rho})^{\frac{\gamma}{\rho}} + s_k K^{\gamma})^{\frac{1}{\gamma}}$$

where, s_{hl} is the labor (high plus low skill) share; ρ is the parameter driving the substitution across the two types of workers. Consider the high and low skill labor supply elasticity given respectively by,

$$w_h = A_h L^{\frac{1}{\epsilon_h}}$$
$$w_l = A_l L^{\frac{1}{\epsilon_l}}$$

where, ϵ_i represents the labor supply of worker type i $\in (l, h)$. Note from the minimization program that marginal productivity of high-skill labor is,

$$f_{lh} = f^{1-\rho} L_h^{1-\gamma}(s_h L_h^{\rho} + s_l L_l^{\rho})^{\frac{\gamma}{\rho}-1} s_{hl} s_h L_h^{\rho-1}$$

By examining the optimal decisions of firms, I can calculate the demand for high and low-skilled labor. More importantly, I derive the labor cost pass-through for each type of labor, as a function of the auxiliary elasticities ($\epsilon_{1+\tau}^{\lambda}$, ϵ_{Q}^{λ}, $\epsilon_{1+\tau}^{Q}$):

$$\epsilon_{1+\tau}^{L_h} = \frac{1 + \frac{s_l}{s_h}(\frac{L_l}{L_h})^{\rho}}{1 + \frac{s_l}{s_h}(\frac{L_l}{L_h})^{\rho} - (\gamma - \rho)} \left[\frac{\epsilon_h}{1 + \epsilon_h(1-\rho)}(\epsilon_{1+\tau}^{\lambda} + \epsilon_{Q}^{\lambda}\epsilon_{1+\tau}^{Q}) - 1 + (1-\rho)\epsilon_{1+\tau}^{Q} + \frac{(\gamma-\rho)\epsilon_{1+\tau}^{L_l}}{1 + \frac{s_h}{s_l}(\frac{L_h}{L_l})^{\rho}} \right] \tag{1.1}$$

$$\epsilon_{1+\tau}^{L_l} = \frac{1 + \frac{s_h}{s_l}(\frac{L_h}{L_l})^{\rho}}{1 + \frac{s_h}{s_l}(\frac{L_h}{L_l})^{\rho} - (\gamma - \rho)} \left[\frac{\epsilon_l}{1 + \epsilon_l(1-\rho)}(\epsilon_{1+\tau}^{\lambda} + \epsilon_{Q}^{\lambda}\epsilon_{1+\tau}^{Q}) - 1 + (1-\rho)\epsilon_{1+\tau}^{Q} + \frac{(\gamma-\rho)\epsilon_{1+\tau}^{L_h}}{1 + \frac{s_l}{s_h}(\frac{L_l}{L_h})^{\rho}} \right] \tag{1.2}$$

To compute the auxiliary elasticities and obtain a closed form solution for the labor elasticities with respect to the labor cost, I start by re-writing the cost function in terms of the marginal productivity of each input, and the marginal cost. Standard envelope arguments enable me to compute ($\epsilon^{\lambda}_{1+\tau}$, ϵ^{λ}_{Q}, $\epsilon^{Q}_{1+\tau}$), and obtain an expression for the labor cost pass-through as a function of observables and structural parameters.

$$\epsilon^{L_h}_{1+\tau} = C_h\left[\left(K_h\left(\frac{\epsilon_h + 2\epsilon^{L_h}_{1+\tau}}{\epsilon_h}\right)+K_l\left(\frac{\epsilon_l + 2\epsilon^{L_l}_{1+\tau}}{\epsilon_l}\right)\right)\left(\frac{\epsilon_h}{1+\epsilon_h(1-\rho)}-(1-\rho)\eta\right)-1+\frac{(\gamma-\rho)\epsilon^{L_l}_{1+\tau}}{1+\frac{s_h}{s_l}(\frac{L_h}{L_l})^{\rho}}\right] \tag{1.3}$$

$$\epsilon^{L_l}_{1+\tau} = C_l\left[\left(K_l\left(\frac{\epsilon_l + 2\epsilon^{L_l}_{1+\tau}}{\epsilon_l}\right)+K_h\left(\frac{\epsilon_h + 2\epsilon^{L_h}_{1+\tau}}{\epsilon_h}\right)\right)\left(\frac{\epsilon_l}{1+\epsilon_l(1-\rho)}-(1-\rho)\eta\right)-1+\frac{(\gamma-\rho)\epsilon^{L_h}_{1+\tau}}{1+\frac{s_l}{s_h}(\frac{L_l}{L_h})^{\rho}}\right] \tag{1.4}$$

where,

$$K_h = \frac{-1}{\frac{1}{s_{L_h}} + \frac{W_l}{W_h}\frac{1}{\epsilon_l} + \frac{1}{\epsilon_h}}\left(\frac{\epsilon_h+1}{\epsilon_h}\right) \qquad K_l = \frac{1}{\frac{1}{s_{L_l}} + \frac{W_h}{W_l}\frac{1}{\epsilon_h} + \frac{1}{\epsilon_l}}\left(\frac{\epsilon_l+1}{\epsilon_l}\right)$$

$$C_h = \frac{1+\frac{s_l}{s_h}(\frac{L_l}{L_h})^{\rho}}{1+\frac{s_l}{s_h}(\frac{L_l}{L_h})^{\rho}-(\gamma-\rho)} \qquad C_l = \frac{1+\frac{s_h}{s_l}(\frac{L_h}{L_l})^{\rho}}{1+\frac{s_h}{s_l}(\frac{L_h}{L_l})^{\rho}-(\gamma-\rho)}$$

For the effect on capital and revenue, not very different from the main model specification with one type of labor, I find:

$$\epsilon^{K}_{1+\tau} = \left(K_h\left(\frac{\epsilon_h + 2\epsilon^{L_h}_{1+\tau}}{\epsilon_h}\right)+K_l\left(\frac{\epsilon_l + 2\epsilon^{L_l}_{1+\tau}}{\epsilon_l}\right)\right)\left(\underbrace{\frac{1}{1-\gamma}}_{subst}\underbrace{-\eta}_{scale}\right) \tag{1.5}$$

$$\epsilon^{R}_{1+\tau} = (1-\eta)\left[K_h\left(\frac{\epsilon_h + 2\epsilon^{L_h}_{1+\tau}}{\epsilon_h}\right)+K_l\left(\frac{\epsilon_l + 2\epsilon^{L_l}_{1+\tau}}{\epsilon_l}\right)\right] \tag{1.6}$$

The associated elasticity of substitution between low and high skill labor is:

$$\sigma_{LH} = \frac{1}{1-\rho}$$

Identification. In this augmented model, it is not feasible to obtain closed-form analytical solutions for all the structural parameters as functions of the reduced-form estimates. The notable exceptions are the labor supply elasticities (ϵ_h and ϵ_l), which can be directly computed from the employment and wage responses for each type of worker. To structurally estimate the parameters ρ, γ, and η, I employ the Classical Minimum Distance

(CMD) approach. The CMD methodology is a non-parametric technique that draws on the moment conditions outlined in equations 1.3, 1.4, 1.5, and 1.6. Formally, the program solves, $\min_\beta [\hat{\beta} - \xi(\beta)]' \hat{W}^{-1} [\hat{\beta} - \xi(\beta)]$, where $\xi(\beta)$ is the vector of model predictions, and $\hat{\beta}$ is the vector of reduced-form estimates. Given the availability of four moments to estimate three parameters, it's possible to assess the validity of the model by conducting a J-test for overidentification. The null hypothesis posits that the model is correctly specified. Notably, a J-test yielding a p-value of 0.86 provides support for the null. Table 1.1 reports the structural estimates.

Structural Estimation. The elasticity of substitution between high and low-skill workers is tighly estimated at 1.27, corroborating the extensive literature that endorses the concept of capital-skill complementarity. This estimate sits comfortably within the range surveyed by Hamermesh, 1996, and micro studies that found 1.5 (Johnson, 1997), and 1.67 (Krusell et al., 2000). The smaller earnings pass-through to low skill workers identify greater elasticities, implying that firms exert greater labor market power over high-skilled workers. While initially, this finding might seem counterintuitive, it aligns with the fact that there are relatively fewer firms hiring in the high-skill market. I find that labor market concentration, proxied by HHI, is 32% greater in the high-skill labor market, reinforcing that unskilled labor operates more as in a commodity market. Such logic rationalizes extensive empirical evidence on the unequal pass-through presented on this paper. Table 1.1 summarizes the results.

Policy Implication. Indeed, understanding the dynamics between skilled and unskilled labor is important for policy implications, as highlighted by Krusell et al. (2000). For example, increasing trade barriers to protect domestic unskilled labor may not be effective if foreign low-wage labor is not the only competitor. Other factors such as automation and technological advancements also play a significant role in the substitution dynamics of labor. Domestic unskilled labor also faces competition from increasingly affordable and advanced capital equipment. Therefore, a more impactful policy for combating inequality might be an investment in basic education, as posited by numerous studies and corroborated in the Brazilian context. By enhancing workers' skills, they can utilize new equipment and increase their productivity, reducing the risk of being replaced by machinery.

Table 1.1: Structural Estimation *(Extended Model)*

Structural Elasticities	(1) **Baseline**
Low-High Skill Elasticity, σ_{LH}	1.27
	(0.04)
High Skill Labor Supply, ϵ_H	3.58
	(1.32)
Low Skill Labor Supply, ϵ_L	6.01
	(2.54)
Output Demand Elasticity, η	1.20
	(0.07)
Empirical Estimates	
High Skill Employment, β_{L_H}	0.14
Low Skill Employment, β_{L_L}	0.12
High Skill Earnings, β_{W_H}	0.04
Low Skill Earnings, β_{W_L}	0.02
Capital, β_K	-0.04
Revenue effect, β_R	0.05
Cost Shares	
High Skill Labor	0.12
Low Skill Labor	0.68
Capital	0.20
J-test	
Overid test (pvalue)	0.86

Notes: This table presents estimates based on the extended model with two types of labor. In the empirical section, the table displays coefficients empirically estimated, and used for the structural estimation. At the bottom, the table displays the p-values associated with the J-test for overidentification. The standard errors for the labor supply elasticities are directly computed from the reduced form estimates, which rely on the Delta Method. In contrast, the standard errors for the remaining structural elasticities are computed using the bootstrap method.

1.14 Robustness Checks

This section presents additional robustness tests to further validate the findings from the main empirical analysis. These exercises help address potential concerns related to sample selection and empirical assumptions. Regarding sample restrictions, there may be concerns that our primary results are influenced by changes in firm composition, namely their initiation and dissolution. To mitigate this, I reapply the empirical analysis on a balanced sample. In terms of identification assumptions, we broaden our approach beyond the assumed exogenous legal variations and re-conduct the empirical study using a matched difference-in-differences methodology, which relies on the conditional independence assumption (CIA). It is noteworthy that across these alternative tests, all findings remain qualitatively the same.

Balanced Sample

The balanced sample is comprised of firms that consistently appear in the data across all sample years from 2008 to 2017. Tables 1.2 and 1.1 below showcase the estimates derived from the firm-level analysis, fitted to this balanced sample. If anything, these point estimates are slightly above compared to the main estimates. However, balanced and unbalanced estimates are statistically indistinguishable from each other.

Table 1.1: Firm Level Estimates

	(1) Log(1+τ)	(2) Log(#Employees)	(3) Log(Earnings)
Panel A: IV			
Diff-in-Diff	-.136***	.114***	.019*
	(.003)	(.03)	(.011)
Long Diff	-.121***	.155***	.024**
	(.004)	(.03)	(.01)
Panel B: ITT			
Diff-in-Diff	-.075***	.063***	.01*
	(.003)	(.017)	(.006)
Long Diff	-.075***	.083***	.014**
	(.003)	(.019)	(.006)
Controls	✓	✓	✓
Firm FE	✓	✓	✓
Sector x Year FE	✓	✓	✓
# Clusters	7,824	7,924	7,824
N	2,422,141	2,491,523	2,422,141

Note: This table presents IV and reduced form (ITT) estimates for the firm-level balanced sample. Difference-in-differences coefficient is estimated in equations 1.3 and 1.4, where there is only one post period. The long difference comes from the period t=+3, in the event study design. Panel A reports the IV coefficients, which adjust for imperfect compliance. Panel B reports the reduced form coefficients, which are interpreted as the intention to treat (ITT) coefficients. Column (1) reports the policy induced labor cost variation, which provides evidence on the first stage. The remaining columns have log of employment, as the dependent variable. Column (2) presents the average effect in the whole sample. Columns (3-5) present heterogeneity based on pre-reform firm size. Firms are categorized as small if they have less than 9 workers in the pre-period. Medium if they have 10-49, and large if they have more than 50 workers. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.2: Within-Firm Earnings Inequality

	Log(Earnings)				Occup Pctile
	firm (99p) (1)	firm (90p) (2)	firm (40p) (3)	firm (20p) (4)	firm level (5)
Panel A: IV					
Diff-in-Diff	.054*** (.015)	.025** (.013)	.015 (.011)	.007 (.01)	.001 (.003)
Long Diff	.082*** (.017)	.038*** (.013)	.016 (.011)	-.002 (.01)	.003 (.003)
Controls	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓
Sector x Year FE	✓	✓	✓	✓	✓
# Clusters	7,924	7,924	7,924	7,924	7,921
N	2,491,523	2,491,523	2,491,523	2,491,523	2,491,146

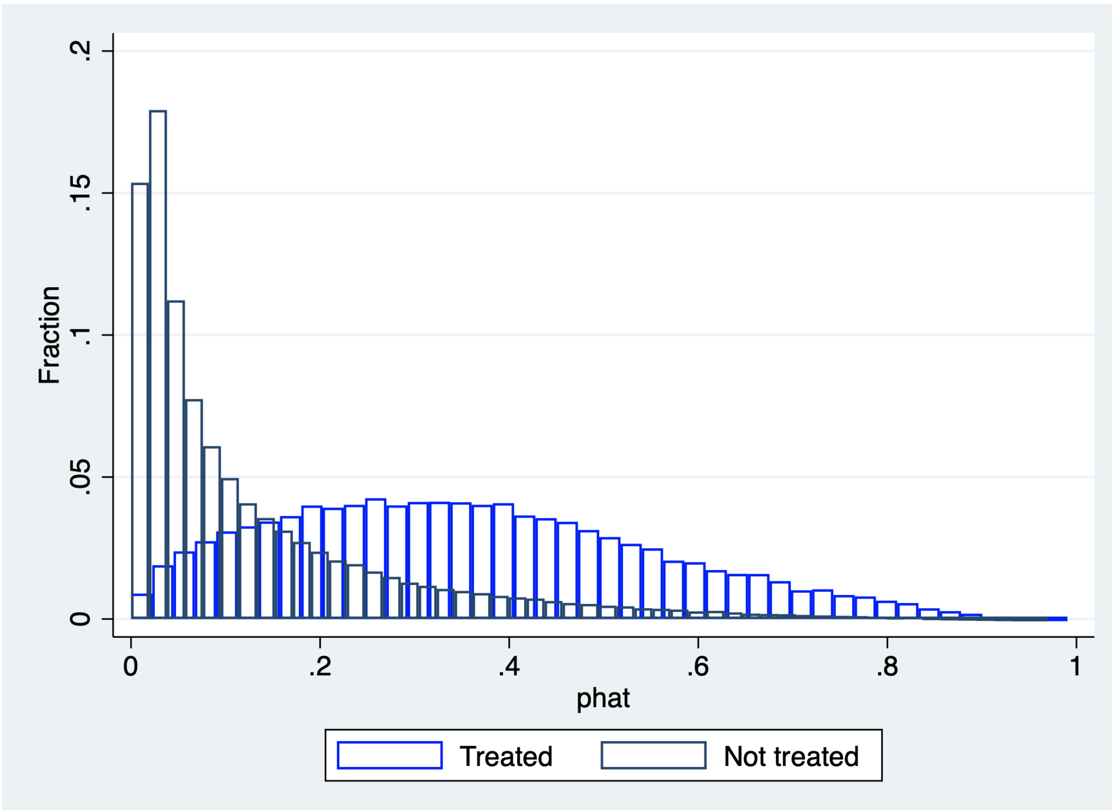
Note: This table presents IV estimates for the causal impacts of the reform on outcomes labeled on each column for the balanced sample. The instrument is the sector eligibility. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Matched Sample

I follow extensive theoretical (Cochran and Rubin, 1973; Rosenbaum and Rubin, 1984; Ho et al., 2007) and applied (Campos and Kearns, 2022) literature that propose matching methods to deal with potential imbalances at baseline.

Procedure. To ensure that pre-trends are not mechanically satisfied, the matching occurs only in two out of the four pre-reform years (2010 and 2011). The procedure goes as follows: each eventually treated firm matches a never treated one that belongs to a non-eligible sector and shares the same pre-reform deciles on average employment, workers' earnings, firm age, net revenue, and profits. In the case of multiple control firms matching the same treated one, I use propensity score to break ties. To compute the propensity score, I fit a logit in the pre-reform period to predict treatment status based on a vector of observables such as $\log$ of employment, wage bill, gross revenue, payroll taxes, profit, and some labor force average characteristics such as age, race, gender, and education. A coefficient ($\hat{\beta}$) is then estimated for each firm, enabling the calculation of the propensity score: $\hat{p} = \frac{\exp\{\hat{\beta}\}}{1+\exp\{\hat{\beta}\}}$. The distribution of propensity scores across the sample is illustrated in Figure 1.1. The noticeable overlap between groups provide evidence of support across the estimated propensity score distribution, validating the matching procedure.

Figure 1.1: Histogram of Propensity Scores



Note: This histogram plots the propensity-score overlap between eventually and never treated firms. The propensity scores are computed in the pre-reform years, and it is based on a logit regression of treatment status on firm-level characteristics.

Balance. The matched sample consists of 30,761 firms in each group. These are firms that appear at least once in the pre-reform years and have a matched counterpart that satisfies the matching conditions. Table 1.3 presents descriptive statistics for both treated and control firms within the matched sample during the pre-reform years. The top five rows report variables used in the matching procedure. Noticeably, the balance holds even across dimensions that were not directly targeted. For instance, for both groups, payroll tax rates are about 34%, the total wage bill BRL 0.85 million, the average worker's age is 33.8 years, 70% are male, 73% are white, 60% have completed high school, and 12% have a college education. The minor discrepancies between the groups do not reach statistical significance at conventional confidence levels, for any characteristic.

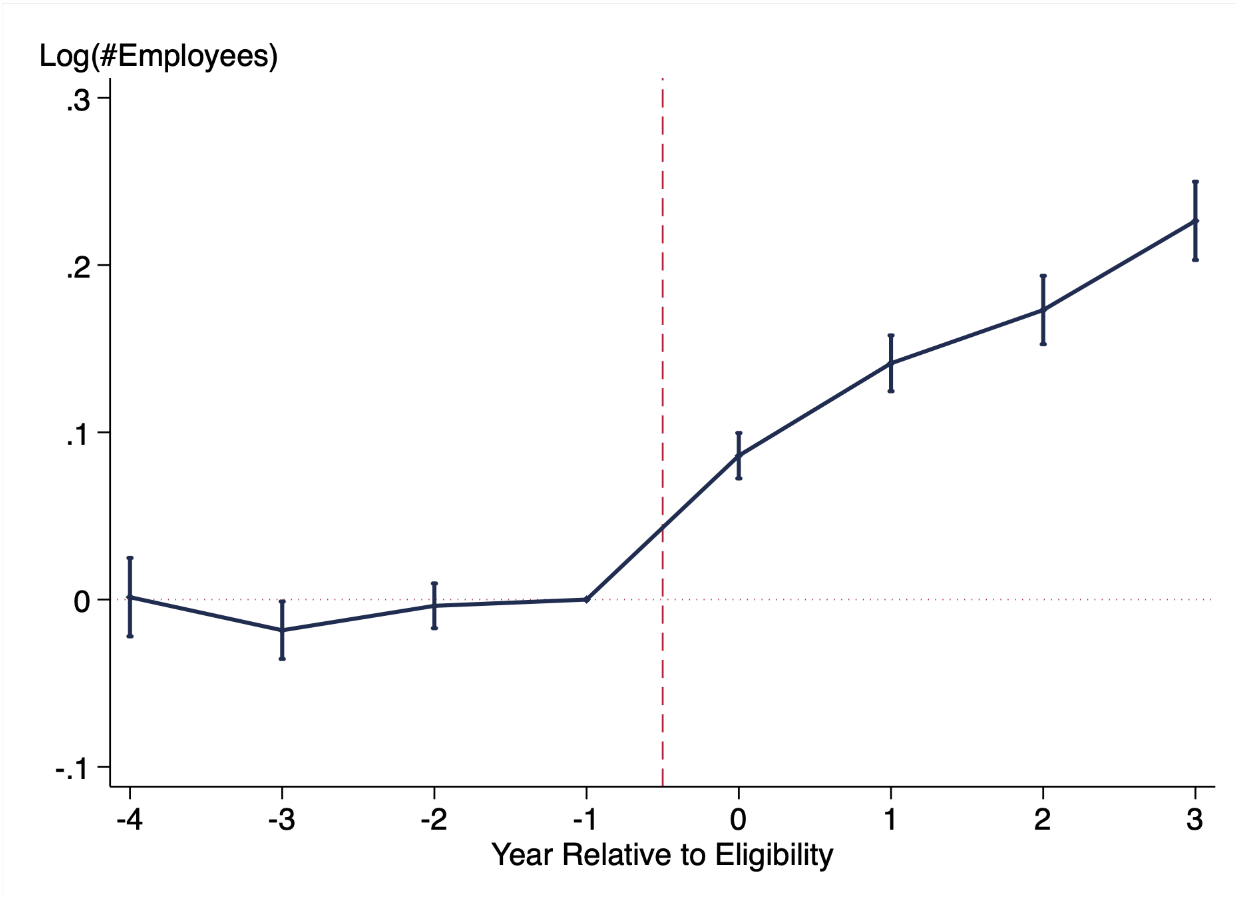
Table 1.3: Balance on Matched Sample

	Treatment	Control
Employment	45.70 (48.21)	45.69 (48.38)
Avg Monthly Earnings	1,461.83 (1,086.82)	1,449.26 (1,348.23)
Firm Age	11.51 (11.80)	11.52 (11.67)
Capital (Mil)	11.28 (17.50)	11.64 (18.16)
Gross Revenue (Mil)	37.63 (48.28)	38.32 (49.29)
Ebit (Mil)	1.40 (3.98)	1.42 (4.08)
Payroll Tax Rate	0.34 (0.08)	0.33 (0.09)
Total Payroll Tax (Mil)	1.29 (8.94)	1.39 (19.59)
Total Wage Bill (Mil)	0.86 (0.90)	0.84 (0.90)
Age	33.80 (5.65)	33.86 (5.64)
Gender	0.73 (0.26)	0.70 (0.29)
Share White	0.76 (0.27)	0.73 (0.28)
Share High School +	0.61 (0.31)	0.60 (0.33)
Share College +	0.11 (0.21)	0.12 (0.21)
Observations	30761	30761

Note: This table provides mean characteristics for eventually treated versus never treated firms in the pre-period. Each observation depicts is a unique firm, which will be followed over time.

Results. I follow treated and control firms over time and estimate the difference-in-differences outlined in equations 1.3 and 1.4. The results are qualitatively similar to the main specification, which validates the empirical findings. Notably, the pre-trends in the matched sample are not statistically significant, as shown in Figure 1.2. It is worth noting that this is not entirely attributable to a mechanical consequence of the matching procedure itself, as only two out of four pre-reform years are used in the matching.

Figure 1.2: Event Study on Matched Sample



Note: This figure presents the event study estimates for the log of employment estimated at the matched sample. In this sample, firms are matched based on pre-reform characteristics in the years of 2010 and 2011. Standard errors are clustered at the firm-level.

Placebo. To further validate the matching design, I conducted a placebo test, randomly assigning firms to treatment, and applied the same matching procedure based on this fake treatment assignment. Given the absence of real tax variation in the fake treatment bucket, we should expect to see zero effects in this analysis. This is precisely what Table 1.5 reports. To showcase that the matching algorithm still works in the placebo sample, Table 1.4 shows that fake treatment and control are balanced in pre-reform characteristics. This finding provides compelling evidence that the main results in the matched sample are actual tax responses and are not mistakenly generated by the matching procedure.

Table 1.4: Balance on Placebo Matched Sample

	Treatment	Control
Employment	15.46 (33.02)	15.29 (32.50)
Avg Monthly Earnings	1,061.97 (1,048.59)	1,057.45 (980.10)
Firm Age	13.82 (10.93)	13.82 (11.00)
Capital (Mil)	8.71 (16.13)	8.58 (15.86)
Gross Revenue (Mil)	26.82 (42.55)	26.93 (42.45)
Ebit (Mil)	0.79 (3.29)	0.80 (3.30)
Payroll Tax Rate	0.31 (0.10)	0.31 (0.09)
Total Payroll Tax (Mil)	0.28 (3.70)	0.35 (13.88)
Total Wage Bill (Mil)	0.27 (0.59)	0.27 (0.58)
Age	37.12 (8.97)	36.41 (8.83)
Gender	0.55 (0.40)	0.51 (0.40)
Share White	0.67 (0.37)	0.69 (0.37)
Share High School +	0.55 (0.41)	0.59 (0.40)
Share College +	0.10 (0.23)	0.11 (0.23)
Observations	35188	35188

Note: This table provides mean characteristics for eventually treated versus never treated firms in the pre-period. Each observation depicts is a unique firm, which will be followed over time.

Table 1.5: Reduced Form on Placebo Matched Sample

	(1) Log Labor Cost $(1+\tau)$	(2) Log Employment	(3) Log Earnings
Panel A: Diff-in-Diff			
Baseline	.0011	.0009	.0002
	(.0023)	(.0179)	(.0071)
Controls	✓	✓	✓
Firm FE	✓	✓	✓
Sector x Year FE	✓	✓	✓
N	450, 666	464, 031	464, 031

Note: This table reports difference-in-differences coefficients instrumented by sector eligibility, estimated at the placebo matched sample. In this sample, randomly selected firms were assigned to a placebo treatment group, and then the same matching procedure is implemented. Given the absence of real tax variation in this fake treatment bucket, we should expect to see zero effects. Each column reports different outcomes, such as labor cost, employment, and earnings. Standard errors are clustered at the firm level and reported in parentheses.

1.15 Additional Figures and Tables

Figure 1.1: Tax Forms Information

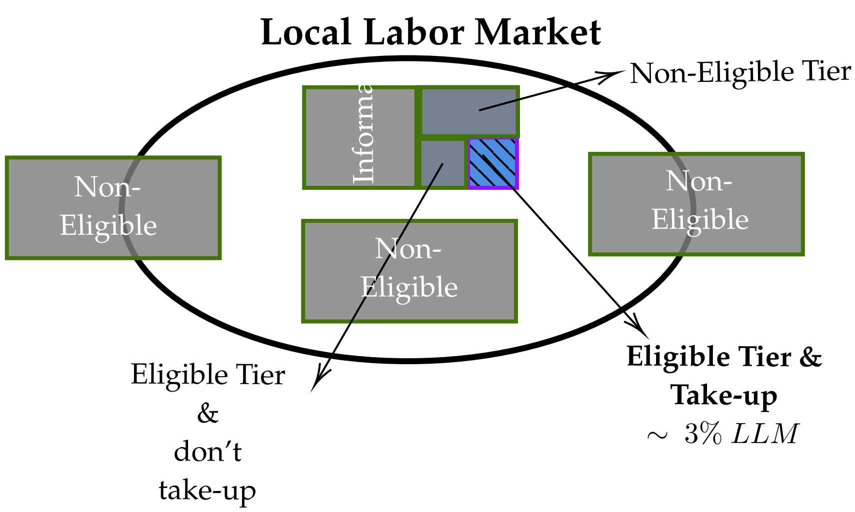
Nº	Campo	Descrição	Tipo	Tam	Dec	Obrig
01	REG	Texto fixo contendo “0145”.	C	004*	-	S
02	COD_INC_TRIB	Código indicador da incidência tributária no período: 1 – Contribuição Previdenciária apurada no período, exclusivamente com base na Receita Bruta; 2 – Contribuição Previdenciária apurada no período, com base na Receita Bruta e com base nas Remunerações pagas, na forma dos nos incisos I e III do art. 22 da Lei nº 8.212, de 1991.	N	001*	-	S
03	VL_REC_TOT	Valor da Receita Bruta Total da Pessoa Jurídica no Período	N	-	02	S
04	VL_REC_ATIV	Valor da Receita Bruta da(s) Atividade(s) Sujeita(s) à Contribuição Previdenciária sobre a Receita Bruta	N	-	02	S
05	VL_REC_DEMAIS_ATIV	Valor da Receita Bruta da(s) Atividade(s) não Sujeita(s) à Contribuição Previdenciária sobre a Receita Bruta	N	-	02	N
06	INFO_COMPL	Informação complementar	C	-	-	N

Nº	Campo	Descrição	Tipo	Tam	Dec	Obrig
01	REG	Texto fixo contendo "P100"	C	004*	-	S
02	DT_INI	Data inicial a que a apuração se refere	C	008*	-	S
03	DT_FIN	Data final a que a apuração se refere	C	008*	-	S
04	VL_REC_TOT_EST	Valor da Receita Bruta Total do Estabelecimento no Período	N	-	02	S
05	COD_ATIV_ECON	Código indicador correspondente à atividade sujeita a incidência da Contribuição Previdenciária sobre a Receita Bruta, conforme Tabela 5.1.1.	C	008*	-	S
06	VL_REC_ATIV_ESTAB	Valor da Receita Bruta do Estabelecimento, correspondente às atividades/produtos referidos no Campo 05 (COD_ATIV_ECON)	N	-	02	S

Nº	Campo	Descrição	Tipo	Tam	Dec	Obrig
01	REG	Texto fixo contendo "P100"	C	004*	-	S
07	VL_EXC	Valor das Exclusões da Receita Bruta informada no Campo 06	N	-	02	N
08	VL_BC_CONT	Valor da Base de Cálculo da Contribuição Previdenciária sobre a Receita Bruta (Campo 08 = Campo 06 – Campo 07)	N	-	02	S
09	ALIQ_CONT	Alíquota da Contribuição Previdenciária sobre a Receita B ruta	N	008	04	S
10	VL_CONT_APU	Valor da Contribuição Previdenciária Apurada sobre a Receita Bruta	N	-	02	S
11	COD_CTA	Código da conta analítica contábil referente à Contribuição Previdenciária sobre a Receita Bruta	C	255	-	N
12	INFO_COMPL	Informação complementar do registro	C	-	-	N

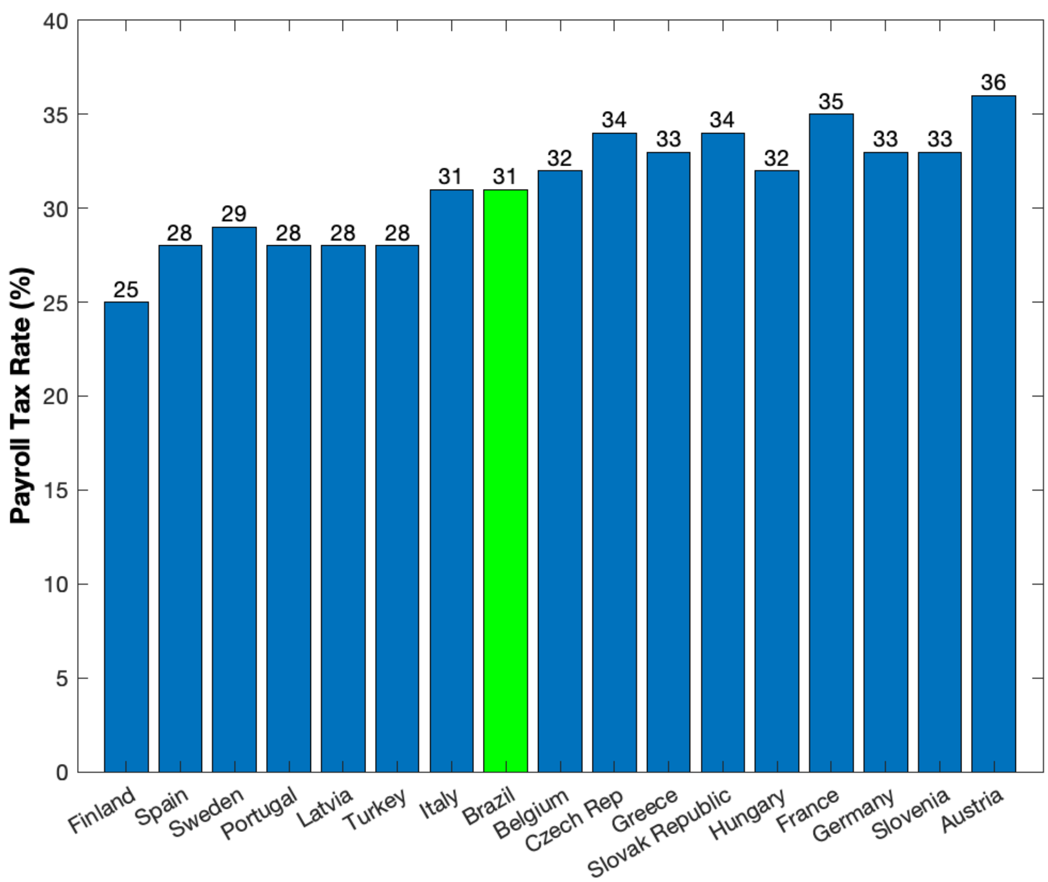
Note: This figure shows instructions for eligible firms to request the payroll tax benefit. It describes detailed information to be provided in Tax Administration software, in order to substitute part of the payroll tax by revenue taxes.

Figure 1.2: Illustration of Treatment Coverage in LLM



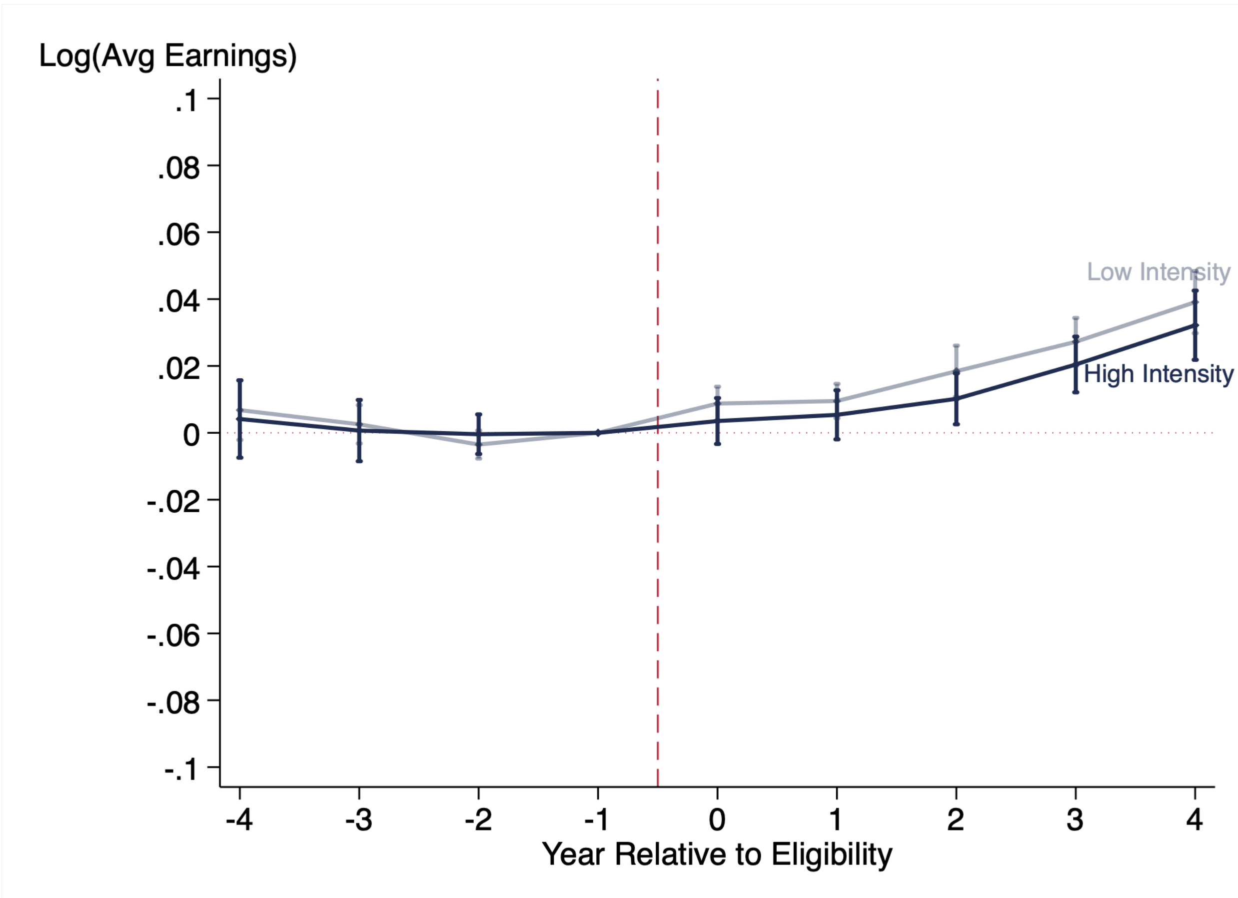
Note: This figure illustrates the policy coverage within local labor markets (LLM), which are defined as occupation x region cells according to job switching patterns. The figure shows that within LLM there are eligible and non-eligible sectors. In the eligible sectors, there is ineligibility due to informality, non-eligible tax tier ("Simples"), and imperfect take-up. All together, the figure illustrates that conditional on existence of an eligible sector in a local labor market, the share of treated firms in the LLM is approximately 3%.

Figure 1.3: Payroll Tax Rates Around the World



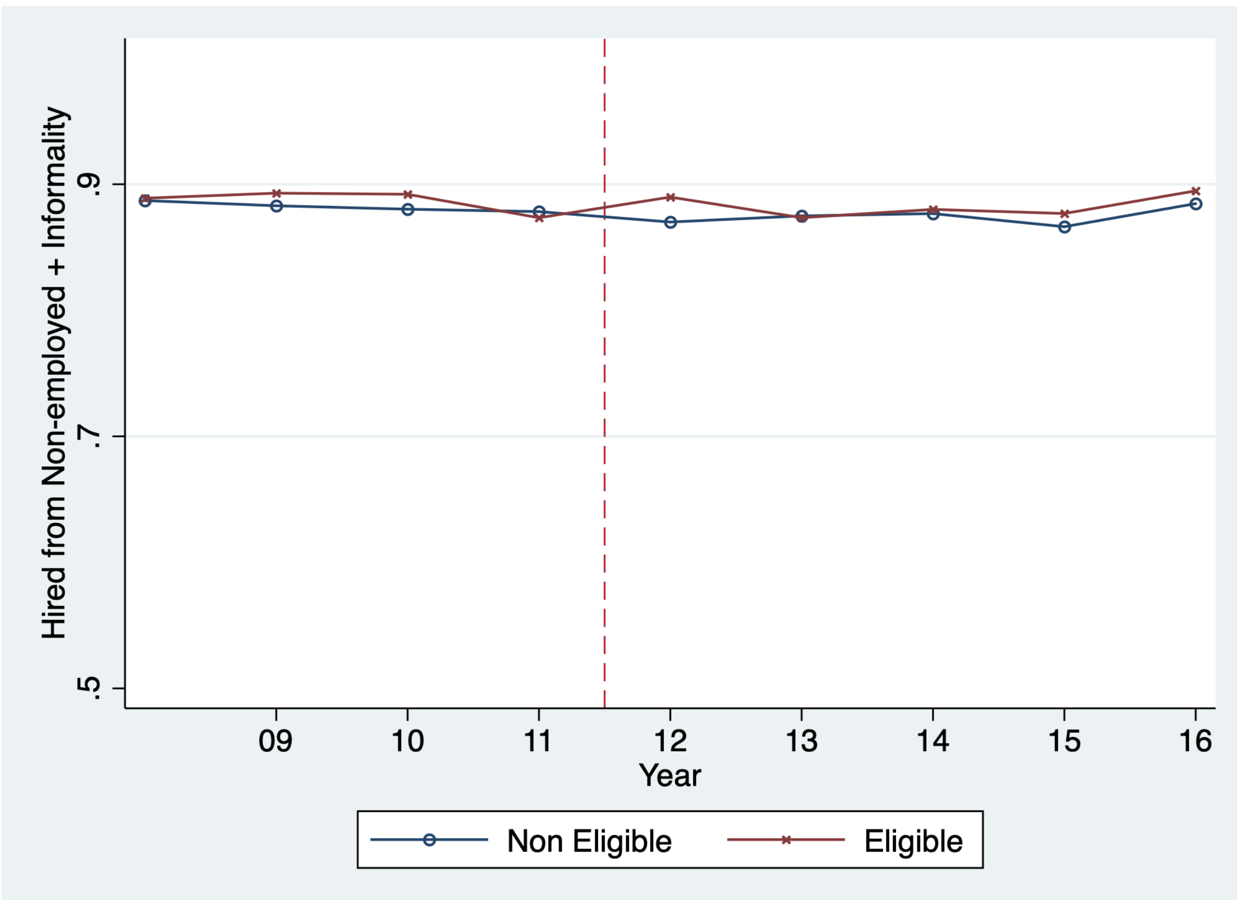
Note: This figure reports payroll tax rates around the world. The payroll tax rate is composed by the sum of employer and employee's contributions.
Source: Elaborated by author, based on information from OECD OECD, 2019.

Figure 1.4: Firm vs Market Level Shock



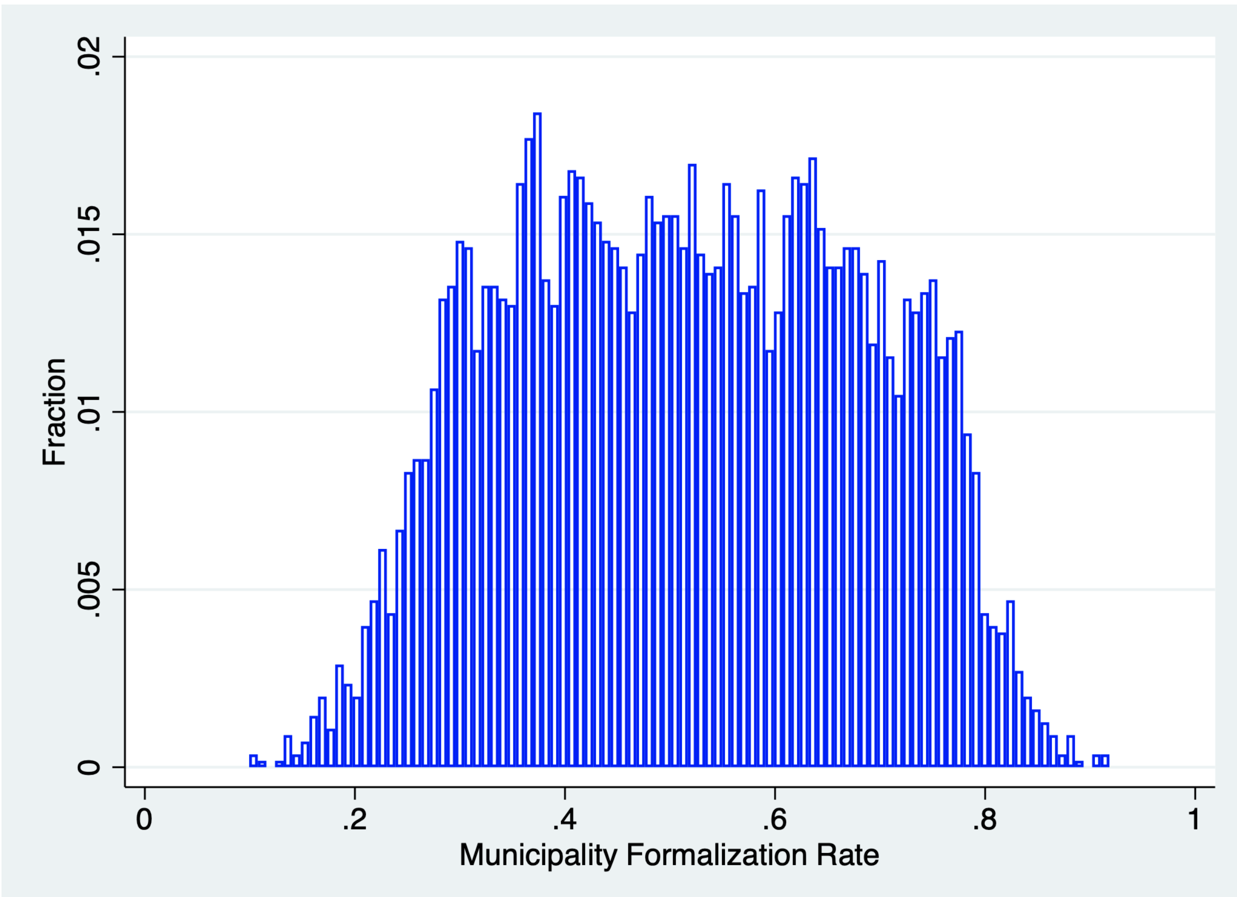
Note: This figure provides an additional test on the spillover effect. It compares the worker's earnings effect for high and low intensively treated markets. To measure market treatment intensity I compute the share of treated workers in each labor market, which are defined by the occupation x region cells. Then it separately estimates the earnings pass-through, for workers in markets below and above the median in market intensity. Standard errors are conservatively clustered at the 5-digit industry-by-state level. If the driving force for the earnings increase was a bump on workers' outside options through market spillover, we would expect to see more pass-through on high intensity markets. The figure shows no significant difference across market intensity.

Figure 1.5: New Hires Origin by Eligibility Status



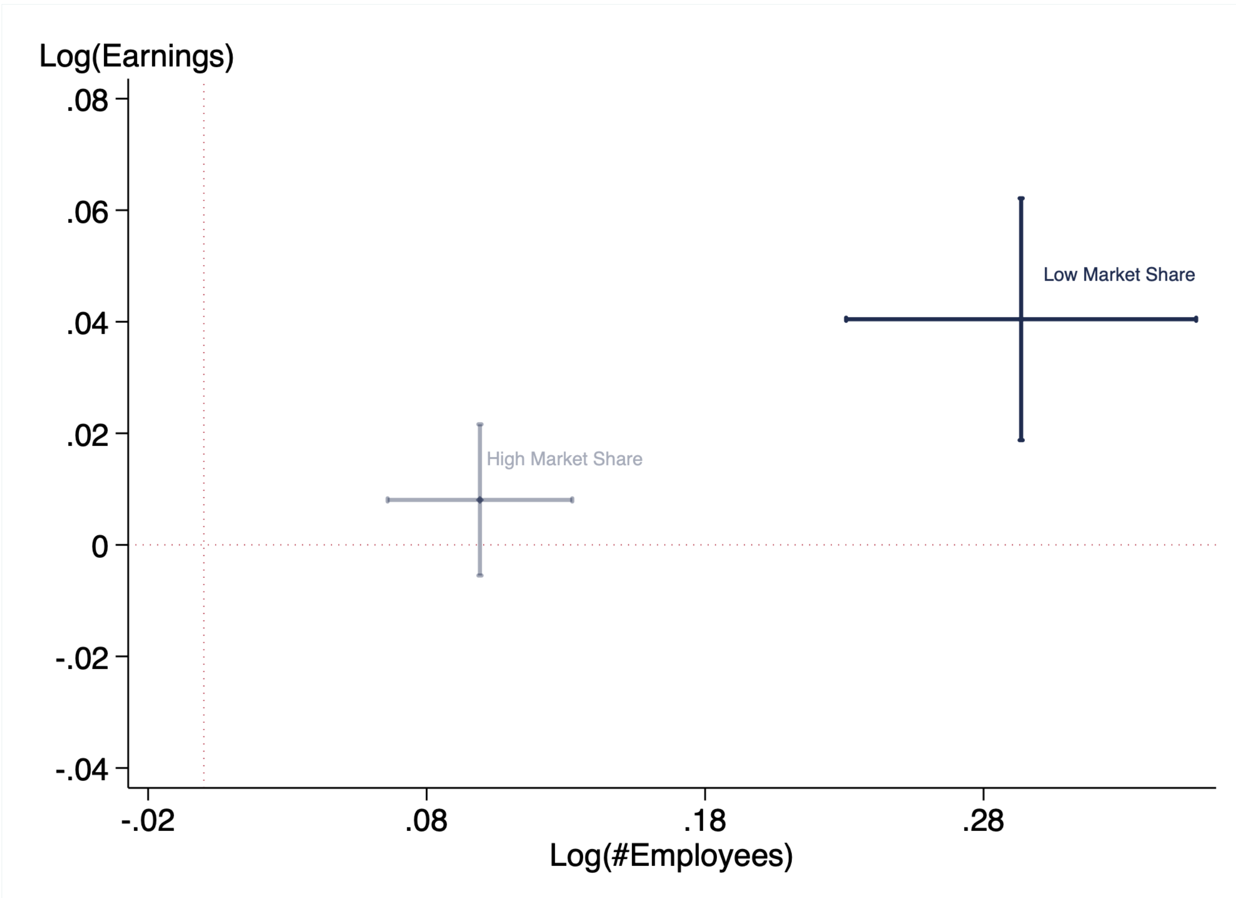
Note: This figure plots the share of new hires coming from non-employment or informality. A new hire is classified as previously informal or non-employed if she was not holding a formal job in the three months prior to being hired. Eligibility is defined based on the sector of employment.

Figure 1.6: Formalization Rates per Municipality



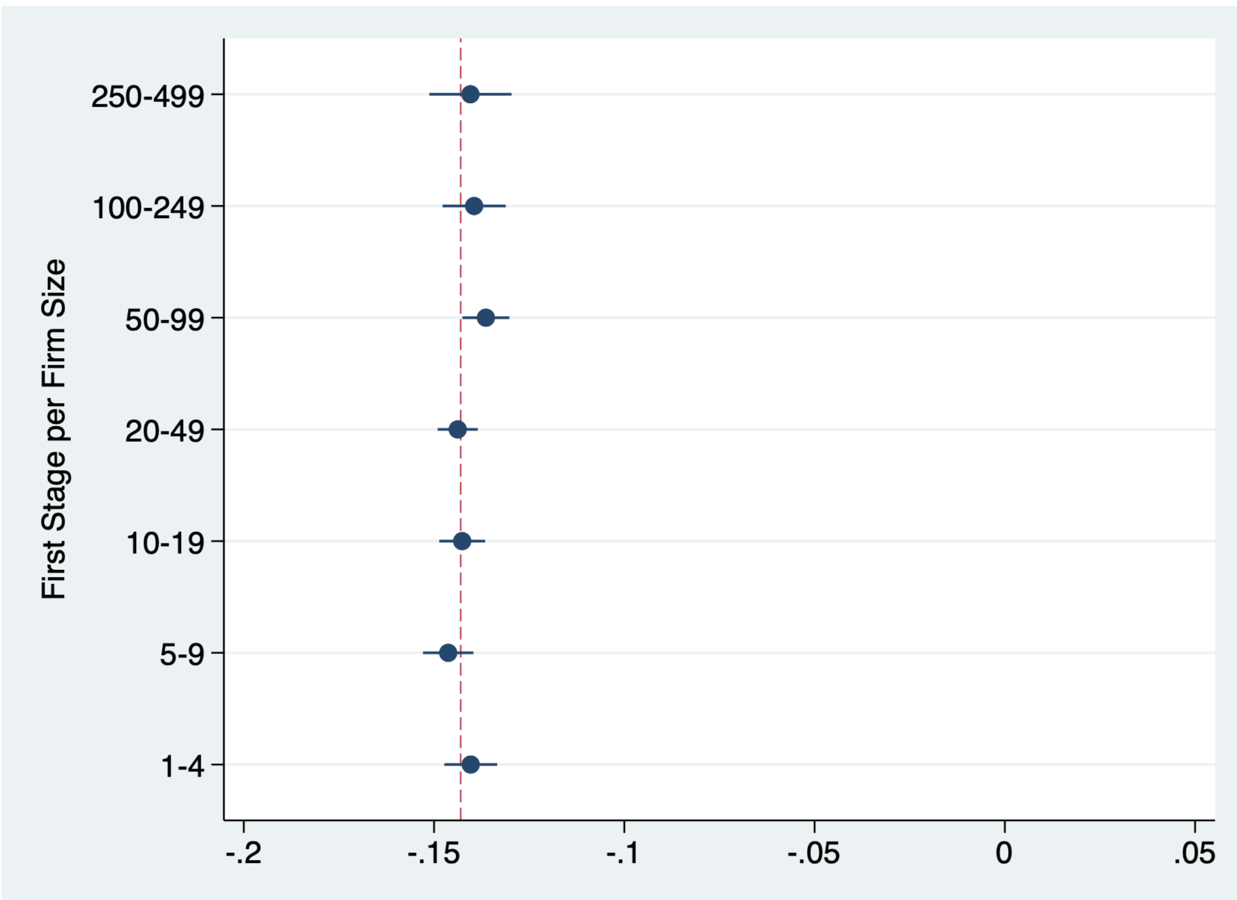
Note: This figure presents the distribution of formalization rates per municipalities in Brazil, according to the 2010 Census. There are 5,300 municipalities with heterogeneous informality rates.

Figure 1.7: Earnings and Employment per Market Concentration



Note: This figure presents firm-level IV difference-in-differences coefficients for above and below the median on pre-reform employment market share within each local labor market. The outcomes are employment and earnings. The blue marker plots the effect for firms below the median (low market power), whereas the gray marker plots the effect for firms with high market power. Horizontal and vertical lines plot the confidence intervals for the employment and earnings estimates, respectively. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Figure 1.8: First Stage per Firm Size



Note: This figure presents the difference-in-differences IV estimate for the effect of the tax treatment on the log of labor cost. Firms are categorized on size bins based on pre-reform levels. It shows that the effect of the reform on labor cost is similar for all size bins. Standard errors are conservatively clustered at the 5-digit industry-by-state level.

Table 1.1: Descriptives on Market Level Treatment

		Average
	Share of Treated Firms in a Treated Market (1)*(2)*(3)*(4)	**0.03**
(1)	Share of Treated Sectors per Market — At Least One	0.211
(2)	Formality Rate	0.550
(3)	Share of Eligible Tax Tier	0.520
(4)	Take Up Within Eligible	0.517

Note: This table breaks down the calculation of treatment share per local labor market. Row (1) reports the local labor market (LLM) share of eligible sectors, conditional on the existence at least one eligible sector on the given LLM. Row (2) reports workers' average formality rate in Brazil; row (3) reports the share of firms in the eligible tax tier; row (4) reports the take-up rate within eligible firms. The product of these 4 rows gives the share of treated firms in a treated market.

Chapter 2

Corporate Taxation and Evasion Responses: Evidence from a Minimum Tax in Honduras

2.1 Introduction

The landscape of corporate taxation has changed significantly in the last few decades. Average statutory corporate tax rates have fallen from over 40% in the 1990s to 30% in low-income countries, and by even more in middle- and high-income countries International Monetary Fund, 2019b. At the same time, technological changes such as the rise of digital companies and the emergence of tax havens mean that governments face increasing challenges to assure compliance in corporate tax payments Zucman, 2014. These trends pose particularly stark threats to the tax base in lower-income countries, which often do not have the institutional capacity to fight tax evasion.

One tool deployed by several governments to assure tax payments by corporations is minimum taxes. While corporate taxes are usually assessed on declared profits, minimum taxes are assessed on a broader base (possibly gross revenue) when reported profits are very low. The International Monetary Fund (IMF) recommends the use of minimum taxes as part of "simple measures protecting against base erosion" International Monetary Fund, 2019a. Some form of minimum taxation on corporations is also at the core of recent international tax cooperation initiatives, such as the G20/OECD Inclusive Framework on Base Erosion and Profit Shifting (BEPS).

In this paper we study corporate responses to the introduction of a minimum tax in Honduras between 2014-2017. Despite the prominence of minimum taxes in economic debates, evidence is scarce on their impact on the behavior of firms Best, Brockmeyer, et al., 2015; Mosberger, 2016; Alejos, 2018.

Before the introduction of the minimum tax, corporations in Honduras faced a flat 25% tax on reported profits, defined as gross revenues minus total claimed deductions.

Starting in FY2014, the country introduced a minimum tax provision mandating that taxpayers declaring yearly gross revenue above L10 million (approximately USD 400,000) pay the maximum between their liability under profit taxation and 1.5% of declared gross revenue. The policy effectively introduced a floor on the effective tax rate paid by large corporations, even when reported profits were low. It was also important from a tax collection perspective, representing approximately 20% of total corporate income tax in the period.

Using the universe of corporate tax declarations between 2011 and 2018, we start by documenting that taxpayers responded strongly to the incentives created by the minimum tax. Since firms reporting gross revenue below L10 million are exempt from the minimum tax, its introduction created a *notch*: a threshold where tax liability might change discontinuously in response to small changes in declared revenue. As an illustration, a firm declaring L9.99 million in gross revenue and close to zero profits will pay virtually no taxes, but declaring L10 million would create a tax liability of L150,000 (1.5%*L10 million) under the minimum tax. This notch generates strong incentives for firms to strategically locate below the exemption threshold. We show that the distribution of firms declaring gross revenue in the vicinity of the exemption threshold was smooth between 2011 and 2013, but presents a clear and increasing excess mass immediately below the threshold when the minimum tax went into effect in 2014. When the exemption threshold was increased to L300 million in 2018, the excess mass around the previous notch immediately disappeared.

We use tools from the bunching literature Henrik J. Kleven and Waseem, 2013b; Henrik Jacobsen Kleven, 2016, adapted to our context, to recover bounds on the elasticity of reported revenue with respect to one-minus the tax rate - a key behavioral parameter to assess the response of firms to a policy taxing revenues. Our estimates suggest that the marginal buncher reduces their reported revenue by 15-30% to avoid being subject to the minimum tax and facing higher tax liability. We estimate revenue elasticities in the range of [0.35,1], considerably higher than previous estimates for similar contexts – Bachas and Soto, 2021, for example, estimate elasticities of reported revenue in the range [0.08 - 0.33] for corporations in Costa Rica.

The large estimated elasticity emphasizes the limits faced by the tax authority in broadening the tax base: increasing tax rates will lead to a substantially smaller tax base. While the revenue response could be entirely driven by real production decisions (firms decreasing sales in order to be exempt) we offer evidence that misreporting revenue is part of the explanation. We construct firm-level measures of revenue observability, which we define as the share of self-declared revenue that is independently observed by the tax authority through third-party reporting. We show that taxpayers are more likely to locate immediately below the exemption threshold when the tax authority has limited ability to independently assess declared revenue: the excess mass below the exemption threshold is 65% larger for firms with below median revenue observability. We also explore different levels of revenue observability across industries and document the same pattern of behavior: firms in high-observability industries are much less likely to bunch below the threshold, implying a lower elasticity of reported revenues. Taken together, we interpret these as evidence that at least part of the observed response of declaring revenue below the exemption threshold is explained by misreporting and thus potentially responsive to the enforcement environment.

While firms that would have declared gross revenue slightly above the exemption threshold might report lower revenue to escape the minimum tax, larger firms will not be exempt. We document that taxpayers with revenues significantly above the threshold reduce their reported costs and increase their reported profit margins, consistent with the fact that under revenue taxation firms cannot decrease their tax liability by inflating costs. We interpret this as clear evidence of evasion under the profit taxation regime. In order to quantify these evasion responses, we explore the fact that a minimum tax creates a *kink* in the tax schedule faced by taxpayers Best, Brockmeyer, et al., 2015: both the tax rate and the tax base change discontinuously at the profit margin level that separates the two regimes, while the tax liability changes continuously.

We show that corporations increased their reported profit margin by 0.9 - 1.1 percentage points when incentives to over report deductions disappeared. Decomposing the profit margin change between real production and misreporting components, we estimate that under profit taxation corporations increase their reported costs by 13 - 17% of their profits in order to reduce their tax liability.

We also explore the rich administrative data to show that not all deduction categories respond in the same manner. We document that firms systematically over-report hard-to-trace deductions, like costs linked to the purchase of goods and materials, when taxed on profits. No over-reporting is observed in categories that generate a paper trail that is easier to verify, like labor or financial costs. This is similar to findings from Mosberger, 2016 in Hungary and strongly suggests a focus for tax authorities' efforts in assessing the veracity of claimed deductions under profit taxation.

To quantify the impacts of the minimum tax on government revenue collection and profit of firms, as well as compare these with alternative tax policies, we impose more structure on the profit maximization problem of firms and calibrate a model using behavioral parameters estimated above. We present two exercises. First, under our parametric assumptions, we quantify the impact of the specific minimum tax policy introduced in Honduras, considering that previously firms were taxed on profits. We estimate that the reform increased tax revenues by up to 30%, but at the cost of reducing aggregate corporate profits by 10% due to larger tax liability and production distortions. Second, we consider a potentially simpler policy change to increase tax revenue from large taxpayers: an increase in the *average* profit tax rate faced by corporations declaring gross revenue above L10 million. We show that to collect the same amount of revenue as in the minimum tax regime would require an average tax rate of 40%, 15 p.p. higher than the tax rate below the threshold. While production is not distorted under the increased profit tax rate, aggregate profits fall by 20% in this scenario driven by increased evasion related losses.

Two caveats about our results should be taken into account. First, we do not attempt to estimate who bears the incidence of corporate taxes Auerbach, 2005; Bastani and Waldenström, 2020. While the classic result of Harberger, 1962 is that capital owners economy-wide bear the full incidence of corporate taxation in a closed economy, recent empirical evidence suggests that a substantial share of the tax burden is also borne by workers Suárez Serrato and Zidar, 2016a; Fuest, Peichl, and Siegloch, 2018a. For those reasons we also do not discuss any possible redistribution motives from the minimum tax reform, since such exercises would require attributing incidence. Second, our model of firm optimization and our simulations do not consider general equilibrium effects of a broader tax base. Limiting cost deduction not only distorts firm size directly, but also cascades down production networks, distorts input prices and the size of downstream firms, and can lead to firm exit. Best, Brockmeyer, et al., 2015 develop a general equilibrium model and show that introducing some degree of production inefficiency is still optimal when enforcement is imperfect.

The first contribution of this paper is to provide new evidence on corporate minimum taxes, a widely-used policy tool to ensure corporations contribute to tax revenue mobilization. We complement previous findings about minimum taxes in Pakistan Best, Brockmeyer, et al., 2015, Hungary Mosberger, 2016 and Guatemala Alejos, 2018. One crucial advantage of our setting in Honduras is that the specific design of the minimum tax allows us to estimate the elasticity of reported revenue for corporations, a key behavioral parameter to understand the impact of revenue-based taxes.

Second, our work provides a detailed anatomy of the nature of tax evasion among medium and large corporations in low- and middle-income countries. We contribute to a small but growing literature using tax administrative records to gain insights on mechanisms behind evasion by corporations Bustos et al., 2022; Carrillo, Donaldson, et al., 2022; Waseem, 2020; Mittal, Reich, and Mahajan, 2018; Almunia and Lopez-Rodriguez, 2018. Our paper shows that the availability of third-party information reduces the elasticity of reported revenue and that cost misreporting is concentrated in categories that are hard to verify such as the costs of goods and materials[1]. These findings provide a more nuanced understanding of the nature of tax non-compliance and reinforce the idea that evasion responses are not fundamental primitives that govern the behavior of firms, but are to some degree sensitive to the enforcement context Fack and Landais, 2016; Joel Slemrod and Kopczuk, 2002; Basri et al., 2019.

Finally, we contribute to the growing literature on bunching methodologies that use discontinuities in the tax design to identify structural parameters (see Henrik Jacobsen Kleven, 2016 for a recent review). While there exists extensive research on how individuals react to discontinuities in the tax schedule Saez, 2010; Bastani and Selin, 2014; Henrik J. Kleven and Waseem, 2013b, we contribute to the more limited literature on how corporations respond to these incentives, similarly to the work of Bachas and Soto, 2021 in Costa Rica and Devereux, Li Liu, and Loretz, 2014 in the United Kingdom.

[1] In the context of personal income taxes, Londoño-Vélez and Ávila-Mahecha, 2019 document substantial evasion of a wealth tax in Colombia, highlighting the use of offshore accounts and of harder-to-observe wealth components as a relevant mechanism.

2.2 Institutional Context and Data

We study a reform that introduced a minimum tax on corporations in Honduras, a lower middle-income country in Central America with a population of 9 million and per capita GDP of $5,800 PPP in 2018. The level and composition of government tax revenues in Honduras is comparable to other countries with similar per capita income. First, total tax revenues represent around 18% of GDP, significantly below the average of 25% observed in high-income OECD countries. Second, the country is much more reliant on goods and services taxes, representing over 50% of total tax revenue, than on income taxes, which amount to one-third of total tax revenue. Finally, corporate income taxes are equivalent to 4% of GDP, almost twice as much as personal income taxes International Monetary Fund, 2018. These last two facts are broadly consistent with the perception that lower-income countries face significant informational constraints in assessing more complex tax liabilities and therefore rely more on broader sales taxes and/or taxing large corporations Gordon and W. Li, 2009[2].

Corporations face a 25% flat tax rate on taxable income, defined as gross revenues minus standard deductions such as wages, raw materials, depreciation of capital, interests paid and carryover losses. Fiscal years in Honduras run according to the calendar year and taxpayers must file the income tax declaration by April 30th.

The minimum tax studied in this paper was introduced in 2014 as part of a broader tax reform that also increased VAT rates from 12% to 15%. The two main features of the minimum tax are as follows. First, it exempts taxpayers reporting gross revenue below L10 million[3], which are still liable for a 25% rate on declared taxable income. Second, taxpayers reporting gross revenue above L10 million are liable for a minimum of 1.5% of their reported revenue. When filing the yearly income tax declaration, corporations must compute their tax liability under the usual profit regime and the 1.5% regime, and are liable for the larger of the two. Since profits are taxed at 25%, a taxpayer declaring 6% profit margin (reported profits divided by gross revenue) will face a tax liability equivalent to 25%*6% = 1.5% of gross revenues and will be located exactly at the edge between the two regimes.

The immediate objective of the minimum tax was to create a floor to the effective tax rate (tax liability divided by gross revenue) faced by large taxpayers: regardless of declared profits, corporations with revenue above L10 million should pay no less than 1.5% of their declared gross revenues in taxes. In 2.1, panel A, we present evidence that the policy substantially raised the effective rate faced by large corporations.

In the period 2011-2013, before the minimum tax was in place, the median effective rate faced by firms with gross revenue around L10 million was approximately 0.5%. Between 2014 and 2017, when the minimum tax is in place for firms declaring revenue above L10 million, the median effective rate substantially changes around the threshold. Firms

[2]Figure A1 illustrates how Honduras compares to other countries in terms of overall and corporate income tax collection.

[3]Approximately USD 400,000 using the average market exchange rate in 2018 (USD 1 = L24.5).

declaring gross revenues below that level still face an effective rate close to 0.5%. Corporations with revenue above L10 million, however, are now subject to the minimum tax and the median firm faces an effective rate of exactly 1.5%[4]. The figure also illustrates the *notch* generated by the minimum tax: by declaring gross revenue marginally above L10 million firms face a discontinuous increase in their tax liability. While in panel A we focus on corporations around the exemption threshold, in panel B we document that the policy was effective in increasing the median effective rate for all firms declaring gross revenue well above the threshold.

The increase in effective tax rate for firms above the exemption threshold is driven by firms that declare low profit margins but no longer pay very small tax liabilities. We illustrate that fact in 2.2, where we plot effective tax rates for firms declaring different profit margins. In the period 2011-2013, before the introduction of the minimum tax, the relationship between declared profit margin and tax liability is approximately linear for all profit margin levels. With the introduction of the minimum tax, the relationship between profitability and tax liability changes for firms with profit margins below 6%. They now face a minimum tax liability equivalent to 1.5% of their gross revenue and the incentive to declare lower profits in order to reduce their tax liability disappears. The figure also illustrates that the policy introduces a *kink* in the budget set of taxpayers exactly at the 6% threshold, with a change in the slope of the tax schedule.

Three special provisions of the minimum tax law are worth discussing in more detail. First, taxpayers in certain sectors (cement, state enterprises, pharmaceuticals and bakery) face a 0.75% rate instead of 1.5%. Firms in those sectors are less than 2% of taxpayers, so we exclude them from our main analysis and present separate results showing that their behavior is also consistent with predictions from theory. Second, we also exclude from our main analyses firms operating in petroleum-related sectors and those in their first two years of operations, which are exempt from the minimum tax[5]. Finally, firms declaring losses are also exempt from the minimum tax. This feature is potentially relevant to our empirical exercises, since that might create strong incentives for low profit firms to report negative taxable income. In practice this behavior seems to be limited in the data. We discuss the likely reasons for that in 2.5.

Despite being part of a larger tax reform, the minimum tax provision was highly salient and widely debated in the public sphere. A previous attempt to institute a 1% minimum tax in 2011 was ruled unconstitutional by the Supreme Court and never went into effect. The 2014 reform was again challenged in the courts but eventually upheld as constitutional in 2015, and stayed in place until FY2017. In the aftermath of highly contested elections in that year, the government approved a series of policy reforms "conced[ing] to long-standing demands from interest groups" International Monetary Fund, 2018, including the gradual phasing out of the minimum tax provision. For FY2018, the exemp-

[4]A2 shows a similar pattern when plotting the average instead of median effective rate.

[5]Both exemptions in the first years of operation and lower rate for sectors such as pharmaceuticals are common features of minimum tax regimes across the world. We provide a summary of minimum tax provisions in several countries in Online Appendix G.

tion threshold was raised from L10 million to L300 million. While approximately 20% of corporations declared gross revenue above L10 million before the introduction of the minimum tax, only 1.3% declared revenues above L300 million in 2017. The law additionally established further increases in the exemption threshold to L600 million in FY2019 and L1 billion in FY2020, meaning that very few corporations would be affected by the minimum tax at the end of the period.

Data and descriptive statistics

The main analyses in this paper are based on administrative data comprising the universe of income tax declarations from corporations in the 2011-2018 period. We supplement this data, in additional exercises, with monthly VAT declarations and third-party information on taxpayers' transactions. Throughout the paper, we exclude taxpayers in special regimes that exonerate them from paying any income taxes – they represent less than 5% of all corporations. The resulting dataset is an unbalanced panel of over 180,000 firm-year observations and approximately 41,000 unique firms.

We present basic descriptive statistics of our sample for years 2013-2018, highlighting the following facts. First, the number of corporations filing income tax has steadily increased throughout the period, from less than 20,000 in 2013 to approximately 30,000 in 2018. While in our main estimates we use an unbalanced panel of taxpayers, we show that firms' responses to the minimum tax are qualitatively similar in a balanced panel of corporations that file every year. Second, average reported gross revenue was around L30 million (USD 1.2 million) but with wide dispersion: the median corporation in the sample had yearly gross revenues of L1.2 million (USD 48,000) and over 80% reported revenues below L10 million. Third, average pre-tax profit margins steadily increase throughout the period, from less than 2% in 2013 to almost 5% in 2018. As discussed below, part of this increase is likely explained by the introduction of the minimum tax, which induced a decrease in claimed deductions and consequent increase in reported profits for large corporations. Despite that, average profit margins are always well below 6%, meaning that the average tax liability under profit taxation is less than 1.5% of gross revenues. Fourth, even though the minimum tax is not directly aimed at multinational corporations (MNC) operating in the country, these are disproportionately large and thus potentially affected by the policy: even though MNCs represent only 2-4% of corporate filers, they pay approximately 60% of taxes[6]. Finally, even though only a small fraction of firms end

[6]Multinational corporations are defined as firms filing transfer price declarations at some point in the period 2014-2017. The potential for the minimum tax to increase tax collection from MNCs depend not only on their gross revenues but also on their profit margin in the absence of minimum taxation. In A3 we show that large MNCs declare higher profit margins than domestic firms in 2013, but still only 30% declare margins above 6%, implying an effective tax rate above 1.5%. In Online H, we investigate whether MNEs have reacted to the minimum tax policy by changing their use of transfer pricing activities. Given data limitations and the small number of MNEs, our estimates are very noisy and we cannot precisely assess those impacts.

up liable for minimum taxes (between 6-8% in 2014-2017), they contribute 20-30% of total corporate tax revenues. Indeed, despite the number of firms liable for minimum taxes falling by an order of magnitude in 2018, when the exemption threshold increased, their contribution to total corporate tax revenues was still close to 15%[7].

In order to illustrate the relevance of the largest corporations to tax collection, we present in 2.1 the share of total revenue and taxes declared by the largest taxpayers. In 2013, before the introduction of the minimum tax provision, the largest twenty corporations in terms of gross declared revenue (top 0.1%) declared almost 30% of total revenues and accounted for 32% of total corporate taxes. Almost 70% of taxes were generated by the top 1% corporations and the top 10% (approximately 2,000 firms) paid more than 90% of taxes[8]. This skewness in the distribution of firm size highlights the potential of the minimum tax to significantly increase revenue collection despite exempting approximately 80% of firms.

2.3 Conceptual framework

Model of firm optimization

In this section we present a stylized model of profit maximization by firms in line with the classical approach of Allingham and Sandmo, 1972 and adapted by Best, Brockmeyer, et al., 2015 to illustrate the incentives introduced by a minimum tax and motivate the empirical exercises that follow. Firms choose a production level y and the level of costs $\hat{c}$ reported to the tax authority, which might be higher than true costs of production given by an increasing and convex function $c(y)$. We assume output prices are fixed and equal to p = 1, so we can express revenue equal to production. Firms face an increasing and convex loss in the amount of cost misreported given by $g\big(\hat{c} - c(y)\big)$, with $g(0) = 0$[9]. Since a regime with a minimum tax allows for both profit and revenue taxation, we model the possibility that only a share $\mu \in [0, 1]$ of costs can be deducted to obtain the taxable income, taxed at rate τ. Firms then choose the vector $(\hat{c}, y)$ to maximize after-tax profits:

$$\underset{(\hat{c},y)}{Max} \quad \Pi(\hat{c}, y) = y - c(y) - \tau\big(y - \mu\hat{c}\big) - g\big(\hat{c} - c(y)\big) \tag{2.1}$$

[7] In A4 we show that corporate tax liabilities substantially increase from approximately L10 million in 2013 to almost L14 million in the year after the introduction of the minimum tax.

[8] This is similar to what Devereux, Li Liu, and Loretz, 2014 report for corporations in the United Kingdom (top 1% account for 80% of corporate income taxes) and Almunia and Lopez-Rodriguez, 2018 report for Spain (top 2% report 80% taxable profits.). In the United States, Auerbach, 2005 mentions that the largest 0.04% corporations in terms of assets account for 62% of all corporate income tax in 2001. In a more similar context, Bachas and Soto, 2021 document that the largest 20% corporations account for 87% of corporate taxes, which is a substantially smaller share than in Honduras.

[9] In our stylized model we consider that firms can only misreport costs and not revenues. This is a simplifying assumption we make to illustrate the idea that it is easier to misreport costs than revenue.

Under a linear tax schedule, first-order conditions are:

$$g'\big(\hat{c} - c(y)\big) = \tau\mu \tag{2.2}$$

$$c'(y) = \frac{1-\tau}{1-\tau\mu} = 1 - \tau\frac{1-\mu}{1-\tau\mu} = 1 - \tau_E \tag{2.3}$$

When choosing how much costs/deductions to report, firms equalize the marginal cost of misreporting deductions to the marginal benefit $\tau\mu$ (not paying tax rate τ on share μ of marginal reported cost). Similarly, the level of production is obtained by equalizing the marginal benefit of producing one extra unit of output $1-\tau$ to the marginal cost $c'(y)(1-\tau\mu)$, which depends on how much of costs can be deducted to obtain taxable income. We re-write 2.3 so that firms equalize the marginal cost of production to $1-\tau_E$, the net-of-tax benefit of marginally increasing production.

Under a pure profit taxation regime, when all production costs can be deducted ($\mu = 1$), we have that $\tau_E = 0$ and $c'(y^*) = 1$: taxes on pure profits are non-distortionary and firms choose the efficient level of production. In the other extreme, when $\mu = 0$ firms pay taxes on their gross revenue and $\tau_E = \tau$ and $c'(y_r) = 1-\tau \implies y_r \leq y^*$. That is, firms are sub-optimally small since the marginal benefit of an extra unit of revenue is $1-\tau$. For any interior value of $\mu \in (0,1)$, production levels will be below optimal.

While taxing a broader base than profits induces distortions in production levels, the opposite is true for evasion levels: under revenue taxation 2.2 becomes $g'\big(\hat{c} - c(y)\big) = 0$ and then $\hat{c} = c(y)$. When costs are not deductible, firms have no incentive to misreport and so report truthfully. Increases in costs deductibility μ induce firms to increase their reported costs in order to reduce tax liability, but also produce misreporting losses[10].

Incentives under the minimum tax

Informed by the model, we now discuss the change in incentives faced by firms that were initially subject to a 25% flat rate on profit and face the introduction of a minimum tax. We can write the tax liability faced by firms as

$$T(y,\hat{c}) = \begin{cases} 0.25 * \big(y - \hat{c}\big), & \text{if } y < 10,000,000 \\ Max\{0.25 * \big(y - \hat{c}\big), 0.015 * y\}, & \text{if } y \geq 10,000,000 \end{cases} \tag{2.4}$$

Consider first firms with gross revenue significantly above L10 million and therefore not exempt from the minimum tax. From the expression above, the tax liability under profit and revenue taxation will be the same whenever the declared profit margin $(y-\hat{c})/y$

[10]Importantly for welfare evaluation, we interpret these evasion losses as social losses, such as the costs of keeping parallel accounting systems or avoiding entering certain economic transactions that might reveal true costs. As discussed by Chetty, 2009, implications for welfare analysis differ if evasion costs are actually seen as transfer between agents (fines paid to the government, for example) or if perceived costs are different from actual costs.

is equal to $0.015/0.25 = 6\%$. Corporations which in the absence of the minimum tax would have reported profit margins above 6% have no incentive to change their behavior: their liability under profit taxation is still larger than 1.5% of their revenues, so they effectively do not face a different regime. Firms which declare positive profit margins below 6%, on the other hand, now face a tax of 1.5% on their gross revenues instead of 25% on declared taxable income. According to the model discussed, this induces changes in two dimensions. First, production decisions are now distorted (since $\tau_E = 0.015$) and firms will reduce production/revenues. Under the assumption of decreasing returns to scale, that effect will lead to an increase in firms' profit margins Best, Brockmeyer, et al., 2015. Second, under revenue taxation firms will not over-report costs, since misreporting entails losses but no longer provides the benefit of reducing tax liability. Both effects will cause the pre-tax profit margin distribution to shift right. Since taxpayers reporting profit margins above 6% are not affected, only the distribution below 6% is shifted and we should observe an excess mass around that threshold.

Consider now the incentives faced by firms that, absent the minimum tax, would have declared gross revenue slightly above the L10 million exemption threshold. Just as discussed above, firms that would have declared profit margins above 6% face no change in incentives and will still choose the same revenue and declared cost levels as they would under pure profit taxation. Low-profit firms, however, now face a different decision. They might declare gross revenue above L10 million and adjust their production and evasion decisions in response to the 1.5% minimum tax liability. But they might also decide to decrease revenue to slightly below L10 million so that they are exempt from the minimum tax and pay the profit tax. Unlike notches generated by wealth Londoño-Vélez and Ávila-Mahecha, 2019 or gross income taxes Henrik J. Kleven and Waseem, 2013b, where all taxpayers above the notch see their liability discontinuously increase, in our setting only a subset of taxpayers are affected by the notch Bachas and Soto, 2021. The benefit of declaring revenue below the exemption threshold, i.e., of bunching, is inversely proportional to the profit margin that would be declared in the absence of the minimum tax.

To see this, consider the profits of a hypothetical taxpayer that must decide between choosing a production level marginally below the exemption threshold (bunching) y^T and reporting cost $\hat{c}$, or producing y_0 above the threshold, reporting true costs $\hat{c}_0 = c(y_0)$ and paying the minimum tax:

$$\Pi(y^T, \hat{c}|Bunch) = y^T - \tau_\pi\big(y^T - \hat{c}\big) - c(y^T) - g\big(\hat{c} - c(y^T)\big) \tag{2.5}$$

$$\Pi(y_0, \hat{c}_0|NotBunch) = y_0 - \tau_y y_0 - c(y_0) - \underbrace{g\big(\hat{c}_0 - c(y_0)\big)}_{=0} \tag{2.6}$$

in which the term of cost misreporting will be zero since staying above the threshold means being taxed on revenue, so there is no incentive to overreport costs.

The gains from deciding to bunch can therefore be written as

$$Bunching\ Gains \approx \underbrace{(y^T - y_0)}_{\leq 0} - \underbrace{\big(c(y^T) - c(y_0)\big)}_{\leq 0} - \underbrace{(\tau_\pi y^T - \tau_y y_0)}_{\geq 0} + \tau_\pi \hat{c} - g\big(\hat{c} - c(y)\big) \tag{2.7}$$

The expression above breaks down the change in profits when deciding to bunch. The first two terms capture the fact that, when bunching, firms will reduce real output, therefore losing revenue, but also reducing costs. The third term captures the fact that bunching means paying a much larger tax rate on gross reported revenues (25% vs. 1.5%), while the fourth term captures the main benefit of bunching: the opportunity to deduct 25% of all reported costs when being taxed on profits instead of revenue. This highlights the fact that the incentive to bunch is directly proportional to costs: for any given level of revenues, firms with higher costs have a stronger incentive to bunch since they will be able to deduct those costs from their tax base when bunching. The fifth term captures the negative effects for the firm in misreporting costs, which is increasing in the distance between true and reported costs.

2.4 Empirical results

We start this section providing non-parametric evidence that taxpayers responded to the introduction of the minimum tax in a manner consistent with the model described above. We then proceed to explore these behavioral responses in order to recover structural parameters of interest.

Evidence of behavioral responses

We start presenting evidence that, consistent with the simple model outlined previously, taxpayers responded to the existence of the exemption threshold by reporting gross revenue immediately below L10 million. In 2.3, we present the empirical densities of reported gross revenues separately for three periods: 2011-2013, before the introduction of the minimum tax; 2014-2017, when the policy was in place with a L10 million exemption threshold; and 2018, when the exemption threshold was increased to L300 million. In the absence of the notch created by the minimum tax, the distribution of reported revenue is smooth throughout the interval. In the period when the minimum schedule creates a notch at L10 million, corporations respond by adjusting their reported revenue to slightly below the threshold: there is a clear excess mass of firms in that region, and a more diffuse absence of mass slightly above. Consistent with the theory presented previously, there is no "hole" in the distribution immediately above the L10 million notch, since the minimum 1.5% effective rate is not binding for firms with high enough profit margin[11].

While firms immediately to the right of the notch have a strong incentive to bunch at the L10 million threshold, firms that would have reported much larger revenue are inframarginal to this bunching behavior. The introduction of the minimum tax leads affected

[11] As discussed by Henrik J. Kleven and Waseem, 2013b and Gelber, Jones, and Sacks, 2020, among others, some firms might not respond to the incentives to bunch due to inattention, high adjustment costs or some combination of other frictions. We discuss below how we interpret the existence of such taxpayers in our elasticity estimates.

firms to decrease evasion through misreporting and decrease scale, increasing reported profit margins. Since only firms otherwise declaring profit margins below 6% are affected we should observe an excess mass of firms exactly at the kink. In practice we often observe a diffuse mass in the vicinity of the kink Saez, 2010. In 2.4, Panel A, we present the empirical density of reported profit margin for firms declaring revenue above L13 million, and therefore infra-marginal to the bunching behavior at the notch, separately for 2011-2013 and 2014-2017. In the period before the introduction of the minimum tax, we observe a steep negative slope in the density of profits, smoothly distributed around the 6% kink. With the introduction of the minimum taxation in 2014, the distribution becomes starkly different as predicted by theory: there is much less mass around positive but close to zero profit margins and firms bunch around the 6% kink.

While in Panel A of 2.4 we illustrate the change in profit margin density before and after the introduction of the minimum tax, in panel B we present empirical densities for the period 2014-2017, while the minimum tax was in place, separately for firms with reported revenue significantly below and above the L10 million exemption threshold. The pattern is remarkably similar to Panel A: firms declaring revenue below the exemption threshold, and therefore unaffected by the minimum tax, are much more likely to declare low profit margins, while those under the minimum tax regime declare higher profit margins and bunch at the 6% kink. We interpret these differences in reported profit margin as evidence that corporations over-report costs under profit taxation to evade taxes, and adjust their behavior when taxed on revenues.

The previous set of figures are strong evidence that the minimum tax was a highly salient policy change that induced behavioral responses from the taxpayers[12]. In the remainder of this section we explore how these responses can be used to identify parameters of interest.

Revenue elasticity at the L10 million notch

We use tools from the bunching literature to translate the observed behavioral responses presented above into estimates of parameters underlying firms' behavior. The core insight developed by Saez, 2010 is that non-linearities in the tax schedule faced by taxpayers will generate bunching, the amount of which is proportional to the elasticities governing the behavior of taxpayers. Our first step is then estimating the counterfactual distribution that would have prevailed in the absence of these discontinuities, so that we can obtain an estimate of the excess bunching and relate that to underlying behavior.

We first discuss how the bunching in response to the L10 million threshold can be used to estimate the elasticity of reported revenue. As previously shown, the exemption threshold generates a notch, where tax liability discontinuously changes for some taxpay-

[12] In A5 we present jointly the change in reported revenue and profit margins using heatmaps. With the introduction of the minimum tax (Panel B), an excess mass of firms declare revenue immediately below the L10 million exemption threshold and, for larger firms, increase their reported profit margins up to 6%.

ers. According to our model, firms deciding to locate exactly at the notch come from a continuous region $[y^T, y^T + \Delta Y]$, where $y^T = L10$ million.

To recover the counterfactual gross revenue density, we fit a polynomial regression to the empirical density of revenue, including dummies for the "excluded region" - the area around the notch affected by the policy 2010; 2011. We then predict the counterfactual density for the entire distribution ignoring the dummies, extrapolating the polynomial prediction to the bunching area and assuring a smooth counterfactual distribution around the notch[13].

We first collapse the data in bins of L100,000 (USD 4,080) of revenue and estimate:

$$n_j = \sum_{k=0}^{5} \beta_k y_j^k + \sum_{b=y_L}^{y_H} \gamma_b \mathbb{1}\{y_j = b\} + \epsilon_j \tag{2.8}$$

where n_j is the number of observations in bin *j*, y_j are the revenue midpoint of bin *j*, $[y_L, y_H]$ is the excluded region affected by the notch and $\mathbb{1}\{y_j = b\}$ are dummies indicating that bin *j* belongs to the excluded region.

The predicted counterfactual density is defined as $\hat{n}_j = \sum_{k=0}^{5} \hat{\beta}_k y_j^k$. We can then obtain the excess mass of taxpayers below the threshold as the difference between the empirical and predicted densities $\hat{B} = \sum_{b=y_L}^{y_N} (n_j - \hat{n}_j)$, where y_N is the bin with upper bound equal to the notch.

The credible estimation of the counterfactual density requires the excluded region to be correctly determined - all those bins affected by the existence of the notch/kink in the tax schedule should not be used to estimate the counterfactual density. We follow the method pioneered by Henrik J. Kleven and Waseem, 2013b when taxpayers face notches: while the lower bound of bunching is visually determined, we use the convergence method to obtain an upper bound for the affected region. We exploit the fact that, according to our model, the excess mass observed immediately below the notch ($\hat{B}$) must be equal to the missing mass above $\left(\hat{M} = \sum_{b=y_N}^{y_u} (n_j - \hat{n}_j)\right)$, so we recursively estimate 2.8 increasing the upper bound y_H until $\hat{B} \approx \hat{M}$[14], at which point we determine that to be the upper bound.

In 2.5 we pool all corporate filings in the 2014-2017 period and present the empirical revenue density as well as the estimated counterfactual density. We provide estimates of the total excess number of firms (B), the excess mass of firms as a share of average density in the bunching region (b), the upper bound of the bunching region estimated using the

[13]The assumption of a smooth distribution is not a trivial one, as pointed out by Blomquist and Newey, 2017 and Bertanha, McCallum, and Seegert, 2018. In particular, they show that kinks cannot identify the elasticity of taxable income if we allow for unrestricted heterogeneity of preferences. In our setting, we can partially alleviate concerns about the counterfactual density by showing, as we do in 2.3, that the density was indeed smooth around the threshold before and after the existence of the notch.

[14]Since we estimate the regression using discrete bins, we determine $\hat{B} \approx \hat{M}$ to mean that $|(\hat{B} - \hat{M})/\hat{B}| \leq 0.03$.

convergence method (y_u) and the number of underlying observations used in each graph (N)[15]. Our estimates indicate that the excess mass below the notch is equivalent to 5.5 times the predicted counterfactual density and that the marginal buncher would have reported gross revenue of L11.8 million in the absence of the notch, effectively reducing their declared revenue by over 15% in order to avoid the minimum tax. The results for each year and for the pooled sample are presented in columns (1) - (4) of 2.2[16].

In order to recover the elasticity of reported revenue from the behavioral responses estimated above, we adapt the reduced-form approximation developed by Henrik J. Kleven and Waseem, 2013b (we present the derivation of the formula in Online Appendix B). We can show that, for a given revenue response ΔY by the marginal buncher, the elasticity of reported revenue is given by:

$$\epsilon_{y,(1-t)} = \left(\frac{1}{\tau_y \left(2 + \frac{\Delta Y}{Y^T}\right) - 2\tau_\pi \frac{(Y^T - \hat{c})}{Y^T}} \right) \left(\frac{\Delta Y}{Y^T} \right)^2 \tag{2.9}$$

Importantly, the estimated elasticity depends not only on the change in reported revenue, but also on the cost that would have been reported when bunching, since the tax base changes from gross revenue above the notch to reported profits below it.

We will compute lower and upper bounds on the true elasticity. The convergence method used to obtain the upper bound of the bunching region provides an estimate of the counterfactual revenue of the marginal buncher. Under the assumption of homogeneous elasticity across all taxpayers, the response of the marginal buncher allows us to recover the structural revenue elasticity. However, if elasticities are heterogeneous the convergence method recovers the response of the taxpayer with higher elasticity Henrik J. Kleven and Waseem, 2013b; Londoño-Vélez and Ávila-Mahecha, 2019. For that reason, we consider our estimate using that method as an upper bound on the true structural elasticity.

While the convergence method provides the revenue response of the marginal buncher, we still need the counterfactual cost to estimate the elasticity. Our model indicates the answer: since the marginal buncher is the taxpayer with the strongest incentive to bunch and incentives are inversely proportional to the profit margin, the marginal buncher has close to zero profits[17]. That allows us to set $Y^T - \hat{c} = 0$ in 2.9 and write the reported revenue elasticity as a function of known policy parameters and the estimated revenue

[15]Standard errors are obtained by bootstrapping the entire estimating procedure, resampling errors from 2.8 500 times.

[16]We also present graphical representation of the estimates for each year in A6.

[17]If firms with real low profits instead decide to declare negative profits to benefit from the exemption to loss-making taxpayers, that would lead to an even higher implied elasticity. In 2.5, we show that this behavior is very muted in the data and discuss the possible reasons for that.

response of the marginal buncher:

$$\epsilon_{y,(1-\tau)} \approx \left(\frac{1}{\tau_y}\right)\left(\frac{1}{2+\frac{\Delta Y}{Y^T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2 \tag{2.10}$$

We present results of the estimated upper bound of the elasticities in column (5) of 2.2. The key quantity needed to obtain the upper bound estimate is the revenue response of the marginal buncher, estimated using the convergence method and presented in column (4). These estimates yield upper-bound revenue elasticities in the interval of $[0.6, 2.6]$. Estimates are particularly large in 2014 (1.3) and 2015 (2.6), when the upper bound of the bunching region is estimated to be above L12 million. Estimates for 2016 and 2017 are very similar (0.61) and smaller than our preferred estimate using the pooled sample ($\epsilon_y = 0.99$). We also note estimates are noisy, with very large standard errors[18].

We now turn to the estimation of the lower bound of the revenue elasticity. Our approach is similar to the "bunching-hole" method proposed by Henrik J. Kleven and Waseem, 2013b, but adapted to take into account the fact that bunching incentives depend on firms' profit margins Bachas and Soto, 2021. We provide a brief description here and save details for Online Appendix C. Since the decision to bunch depends both on counterfactual revenue and costs, we can rewrite 2.9 to find the counterfactual cost that would make a taxpayer indifferent between bunching or not, given a distance ΔY from the threshold and elasticity ϵ_y:

$$\hat{c}^* = Y^T\Big(1-\frac{\tau_y}{\tau_\pi}\Big) - \frac{\tau_y}{\tau_\pi}\frac{\Delta Y}{2} + \frac{(\Delta Y)^2}{2\epsilon_y\tau_\pi Y^T} \tag{2.11}$$

Since the incentives to bunch are inversely related to profit margins, we know that if a taxpayer with revenue $Y^T + \Delta Y$ and cost $\hat{c}^*$ is indifferent to bunching, all taxpayers with lower profit margins should also bunch since they face even stronger incentives. If we knew the counterfactual profit margin distribution, we could compute the share of taxpayers bunching for each revenue bin, for a given elasticity, and compare the total amount of predicted bunching to our estimated excess mass below the notch. In order to implement that strategy, we need to make an assumption about the unobserved counterfactual profit margin distribution above the threshold. We assume the profit margin distribution for firms reporting revenue in the interval L6 - 8 million, significantly below the notch, is a good approximation for the unobserved distribution Bachas and Soto, 2021[19]. We then compute the estimated elasticity as the one generating a predicted amount of bunching equal to the excess mass observed below the notch, among a range of elasticity values

[18]Standard errors are estimated by bootstrap and the empirical distribution of estimated elasticities is highly non-symmetrical: for the pooled sample where the point estimate is 0.99 the empirical 95% confidence interval is $[0.7, 5.7]$, meaning there is significant uncertainty on the upper bound of the estimate, but little on the lower bound. We present the histogram of our bootstrap estimates for the pooled sample in A7.

[19]We show in A8 that the profit margin distribution is similar for the L6 - 8 million and L10-12 million range in the period before the introduction of the minimum tax.

One important caveat of the lower bound methodology is that we consider that all taxpayers that have an incentive to bunch will do so. There is ample evidence, nonetheless, that even when facing strictly dominated regions some taxpayers do not bunch Henrik J. Kleven and Waseem, 2013b; Gelber, Jones, and Sacks, 2020. While notches often give rise to strictly dominated regions for all taxpayers and allow researchers to estimate optimization frictions, we show in Online Appendix D that is not the case with the exemption notch in Honduras. Since the size of the discontinuous change in tax liability depends on counterfactual profit margins, the existence and extent of a dominated region also depends on the counterfactual profitability. While it is possible to make stronger assumptions, ruling out extreme preferences in order to estimate optimization frictions 2020, we abstain from doing so and consider our estimates to be lower bounds for the true reported revenue elasticity: the existence of optimization frictions require, all else equal, a larger elasticity to obtain the same amount of predicted bunching mass.

We present lower bound estimates for ϵ_y in column (6) of 2.2. Here estimated elasticities are both much lower and more stable across years, and likewise much more precise and statistically different from zero in every period. While the elasticity is lower (0.2) in 2014, when we observe significantly less bunching, for the period 2015-2017 and the pooled sample estimates lie tightly between 0.35 - 0.4.

We take results for the pooled sample as our preferred estimates, where we obtain a range for the reported revenue elasticity of $[0.35, 0.99]$[20]. These are substantially larger than the estimates obtained by Bachas and Soto, 2021 for corporations in Costa Rica, for example, where the similar range using lower and upper bound estimates is $[0.08, 0.33]$. They are also much larger than estimates of individual earnings elasticities in Pakistan obtained by Henrik J Kleven, 2018, which mostly fall in the range $[0.05, 0.3]$. Our results suggest that, under the existing enforcement environment while the minimum tax was in place, the reported gross revenue of corporations was highly elastic, limiting to some extent the ability of the tax authority to increase revenues through higher tax rates.

Real or misreporting response at L10 million notch?

The observed response in declared gross revenues under the minimum tax could be due to real production decisions, to under reporting of realized revenues or to a mix of both. In this section we explore the evidence related to these possibilities.

We investigate whether the amount of bunching is related to the availability of third-party information (TPI) about the sales of taxpayers. Previous studies have documented much less bunching in response to change in marginal tax rates among wage-earners than among the self-employed Saez, 2010 and also less evasion (measured by audits)

[20] We perform robustness exercises for the estimated elasticity of reported revenue in A1, using different polynomial orders. For the lower bound elasticity, the estimate is unchanged using a higher order polynomial but somewhat larger (0.5 - 0.6) when using a lower order polynomial. Consistent with noisy estimates in our preferred specification, however, estimates for the upper bound vary significantly when using different polynomials, ranging from 1.6 to 6.

for income with third-party information 2011; 2019. We hypothesize that observing less bunching among taxpayers with high "revenue observability" is evidence that misreporting is at least partially responsible for the behavior observed.

Several transactions in which firms engage, such as selling to the government or exporting, generate third-party information: these sales are directly reported to the tax authority, allowing them to independently assess part of the revenue declared by taxpayers[21]. Nonetheless, the availability of this information is limited: less than 60% of corporations have any third-party information available, and even among larger firms declaring revenue above L5 million more than 15% are not covered at all[22]. We use these reports to create a firm-level measure of revenue observability, defined as the share of self-declared revenue that is independently observed by the tax authority[23]. Conditional on having any third-party information available, the median ratio between third-party informed and self-declared revenue is 25%.

In 2.6, panel A, we plot the empirical density of revenue for the period 2015-2017 around the L10 million threshold separately for two groups: corporations for which some third-party information is available and those for which it is not. We observe bunching in both distributions, although there is slightly more mass below the threshold among those firms with no third-party information available. Since for a significant number of taxpayers the amount reported by third-parties is very small, we repeat the exercise in panel B, now separating the sample in those above and below the (unconditional) median of revenue observability (15%). Here we observe a much sharper bunching behavior for firms with lower revenue observability, although excess mass is still clearly present for firms with a higher degree of third-party coverage. We quantify these differences in panel A of 2.3. Whereas we estimate the excess mass at the notch for firms with above median revenue observability as four times the counterfactual density, for firms with below median observability we estimate seven times as much mass, and this difference is precisely estimated.

We provide additional evidence that bunching below the exemption threshold is par-

[21]The tax authority has access to five sources of information on taxpayers' revenues. The most important are sales to some large companies, which are mandated to report individual purchases as part of the credit system used for VAT. The other sources are withholding of sales using credit and debit cards; sales to the government, exports, and services provided to a subset of very large companies. Data on third-party information is only consistently available since 2015 so we restrict our analysis to the period 2015-2017.

[22]In A9 we provide the distribution of third-party information coverage for all firms and for those located around the L10 million threshold. Even for firms with above-median TPI coverage the tax authority only has limited information on their revenues: for only 1 out of every 10 firms the tax authority observes more than 90% of their revenues independently recorded by third-parties

[23]While available information is an important condition for tax authorities to enforce tax compliance, it is not sufficient. Audits in Honduras are rare but strongly size-dependent: there were less than 160 full- or partial-audits in 2014, the year the minimum tax was introduced, but almost all of them were focused on the top 20% of taxpayers in terms of revenue. We provide some information on the enforcement environment in A10 and A2. We also note that penalties for non-compliance can be high, including fines and the loss of tax exemptions, and that approximately 7% of all corporations received some fine in 2018 for not presenting a declaration, presenting it late or including incorrect information.

tially driven by revenue misreporting by evaluating heterogeneity across industries. The availability of TPI varies systematically across industries given the nature of their economic activities. Since the main source of third-party information is withholding through the VAT credit system, revenues from firms in upstream sectors are more likely to be reported to the tax authority. On one extreme, the median corporation operating in construction or retail sees less than 15% of their total self-declared revenue being reported directly to the tax authority by third-parties. On the other, for the median firm in the manufacturing sector the revenue reported by third-parties amounts to approximately 40% of their self-reported revenue. We then evaluate whether bunching at the sectoral level is systematically correlated with the degree of revenue observability in each industry, in the spirit of the analysis in Almunia and Lopez-Rodriguez, 2018 but using firm-level data on revenue observability, allowing for a direct measure of the information set available for the tax authority on the revenue of taxpayers[24].

In panel B of 2.3 we present estimates of excess bunching at the notch, normalized by the predicted density at the threshold (column 2). First, we estimate large and precisely estimated excess bunching for firms in all industries. The amount of bunching, however, varies significantly across sectors: the excess mass ranges from 3.5 times the counterfactual density in manufacturing to approximately 8 times in agriculture and construction. To assess whether the amount of bunching is correlated with the availability of TPI, in 2.7 we plot the estimated excess mass below the notch and the median revenue observability in each industry. We observe a strong negative correlation between the two measures: in industries where third-party reporting covers a larger share of a firm's revenue much less bunching is observed immediately below the L10 million notch. Consider retail, where the majority of sales are to final customers and a low penetration of debit and credit cards means that only a small fraction of corporations' revenues are reported to the tax authority. The excess mass observed below the notch is seven times the predicted density, indicating a large amount of response to the incentives provided by the minimum tax. Manufacturing firms, on the other extreme, mostly supply to other firms and see a much larger share of their total sales directly informed to the tax authority. Here the excess mass at the notch is only half that observed among retail firms. While other factors might contribute to the observed negative correlation, we interpret this as further evidence that misreporting revenues plays a role in explaining the observed bunching below the exemption threshold.

Estimating evasion under profit taxation

We now turn to firms with gross revenue significantly above L10 million and therefore inframarginal to the bunching behavior below the notch. As documented above, the intro-

[24]Almunia and Lopez-Rodriguez, 2018 rely on input-output tables to compute the share of sales from each sector to final consumers. Our sectoral definition is somewhat different from theirs, but we show in A11 that our results are qualitatively similar when we use a similar industry grouping.

duction of the minimum tax led to an increase in the reported profit margins and bunching around the 6% threshold, which separates the profit and revenue taxation regimes.

Let B be the excess mass of taxpayers locating around the threshold. These bunchers are coming from a continuous segment $[\Pi^T - \Delta\Pi, \Pi^T]$ below the kink: these are taxpayers that otherwise would have reported lower profit margins, but under revenue taxation increase their reported profit. The area where these bunchers come from is not empty, however, since the entire distribution shifts to the right as taxpayers declare higher profit margins.

Following a very similar approach to the one used above, our goal is again to estimate the counterfactual distribution and use it to obtain an estimate of excess bunching at the kink. We estimate a counterfactual distribution of profits using a polynomial regression akin to 2.8 and obtain estimates of the excess mass of taxpayers located around the kink[25].

In A12 we present the empirical and estimated counterfactual profit margin densities for each year in the period 2014-2017. Between 90 and 210 firms are estimated to bunch around the 6% profit kink each year, an excess mass equivalent to 3-6 times the average density in the interval. In 2.8 we present results for the pooled sample, where we estimate a similar excess mass equivalent to 5.4 times the average counterfactual density around the kink. We present the same results in the first two columns of 2.4.

Starting from the estimated excess mass around the kink, we can recover the change in reported profit margin by the marginal buncher noting that the bunching mass B around the threshold can be expressed as:

$$B = \int_{\Pi^T - \Delta\Pi}^{\Pi^T} f_0(\Pi) d\Pi \approx \Delta\Pi f_0\Big(\Pi^T\Big) \implies \Delta\Pi \approx \frac{B}{f_0\Big(\Pi^T\Big)} \tag{2.12}$$

where $f_0(.)$ is the counterfactual profit margin density and the approximation assumes the density is constant on the bunching segment. Empirically, we estimate $f_0(\Pi^T)$ as the average predicted density in the bunching region, and use the estimated excess mass at the kink to obtain $\Delta\hat{\Pi} \approx \frac{\hat{B}}{\hat{f}_0(\Pi^T)}$.

We present results for the estimated change in profit margins in column (3) of 2.4. With the exception of 2014, when we observe less bunching, estimates for 2015-2017 and for the pooled sample are very similar: the marginal buncher increased declared profit margin between 0.9 - 1.1 percentage points, a narrow range of precisely estimated responses. To put it differently, the marginal buncher would have declared a profit margin of approximately 5% under profit taxation, when incentives to misreport are strong and production decisions are undistorted.

In order to interpret the magnitude of these changes in reported profit and separate the total effect between cost evasion and production decisions, we use the decomposition of reported profit margin response developed by Best, Brockmeyer, et al., 2015. Totally

[25] We compute the number of taxpayers in bins 0.2 p.p. wide. Following the literature, we determine visually the lower and upper bounds of the bunching region.

differentiating the reported profit margin and considering the incentives of a taxpayer around the kink yields:

$$\Delta\hat{\Pi} = \frac{\tau_y^2}{\tau_\pi}\epsilon_{y,(1-\tau)} - \frac{d(\hat{c} - c(y))}{y} \tag{2.13}$$

The main insight provided by the decomposition is that, since the tax rate on revenues is often very small (0.015 in the case of Honduras), even large revenue elasticities will only generate second-order effects on the change in reported profit margins. If we observe large increases in reported profit margins from the marginal buncher, then changes due to evasion incentives must be playing a large role. We illustrate that point in column (4) of 2.4, where we consider the implied revenue elasticity in a model where there is no cost evasion. For all years and for the pooled sample, the implied elasticities under no cost evasion are implausibly high: with the exception of 2014 when the estimate is 6.7, the remaining elasticities of 10-12 are four times larger than our largest estimate in 2.2 and an order of magnitude higher than our preferred estimates, suggesting that cost evasion must be playing a significant role in explaining the observed response.

We present our estimates of misreporting in column (6). We use the upper bound elasticity $\epsilon_y = 0.99$ obtained for the pooled sample, so evasion estimates are a lower bound of the true evasion, and express evasion as a share of reported profits. With the exception of 2014, where bunching is smaller, in the period 2015-2017 and using the pooled data we estimate that cost misreporting is in the range of 13-17% of reported profits. Results are practically unchanged when we consider alternative polynomial orders in our estimates, as reported in A3. They are also very similar if we consider that evasion is purely driven by misreporting in revenue instead of costs (A4).

These estimates are very similar to evasion documented by Best, Brockmeyer, et al., 2015 for most corporations in Pakistan, which also fall in the range of $[0.13, 0.17]$, and by Alejos, 2018 for corporations in Guatemala that fail to claim an exemption to the local minimum tax. Our results reinforce these previous findings that evasion through cost misreporting in lower income countries is significant even for large corporations, making the use of taxation of broader bases a potential tool to increase tax revenues.

The composition of cost adjustments

In the previous section we document that corporations evade a substantial amount of taxes by over reporting costs under a profit regime, and immediately change their reporting behavior when evasion incentives disappear under the minimum tax. One relevant policy question arising from these evasion responses is whether firms adjust all cost categories similarly between these regimes, or if some cost items seem particularly prone to evasion.

We first present evidence, in 2.9, that deduction levels change discontinuously at the L10 million revenue threshold, consistent with the fact that, under the minimum tax, firms above the threshold increase their reported profits. Reassuringly, we observe no discontinuity in claimed costs in the period 2011-2013, before the minimum tax was in place. In order to assess whether specific cost categories are more responsive to the change in incentives, we use detailed cost items claimed in corporate income tax filings to construct five broad cost categories: Labor, Goods and Materials, Operations, Financial and Losses & others[26]. In 2.10, Panel A, we present costs as a share of gross revenue for each bin of declared revenue. The figure suggests that costs related to the purchase of goods and materials are the only ones that significantly change at the L10 million threshold. While for firms declaring revenue below L10 million the participation of goods and materials steadily increases, the average share of those costs falls discontinuously by over 5 p.p. at the threshold and remains at a lower level for firms declaring up to L15 million in revenue. We do not observe a similar discontinuous fall in claimed deductions for other categories that generate more paper trails, such as financial or labor costs. In Panel B of the same figure we focus on the goods and materials category, showing that the discontinuous change observed at the notch is not observed in 2018, when the exemption threshold increases to L300 million.

We present a more formal test of whether these discontinuities can be attributed to the minimum tax. Since we previously presented strong evidence that taxpayers strategically locate below the revenue threshold in order to avoid the minimum tax, we cannot simply estimate a regression discontinuity at the notch. Instead we estimate a linear "donut-hole" discontinuity regression, evaluating whether the level of costs change at the threshold but extrapolating from revenue levels not affected by bunching behavior.

In Column (1) we present results from a specification using median deductions by bin as the dependent variable. We estimate that the amount of claimed deductions falls by approximately L260,000 at the threshold, consistent with the non-parametric evidence presented. Since the median deduction at the threshold is L9.8 million, the estimated effect implies that the median firm above the threshold decreases deduction claims by 2.7% and doubles the reported profit margin. In Columns (2) through (5) we repeat the same exercise but use the ratio of deductions to revenue as the dependent variable. The only estimate statistically different from zero and meaningful in magnitude is goods and material costs: they fall by almost 5 p.p. from an average of 37% below the notch. Mosberger, 2016, using a different empirical strategy, also documents a significant change in goods and materials costs by firms facing a minimum tax in Hungary, suggesting this seems to be a deduction category particularly over reported by firms trying to reduce profit tax liabilities and therefore a potential focus for tax authorities[27].

[26]The detailed breakdown of cost categories only exists for firms declaring using the electronic form introduced in 2015. In all exercises using detailed cost data, we restrict our sample to the period 2015-2018 and to taxpayers filing electronically (70 - 80% of all corporations).

[27]The enforcement environment to assess the veracity of claimed deductions in Honduras, as in other low- and middle-income countries, is limited. Carrillo, Donaldson, et al., 2022, for example, show that in Chile corporations "are required to file purchase annexes, which includes supplying valid invoice numbers" to validate non-labor cost deductions. This relates directly to the existence of "ghost firms" or "invoice

2.5 Robustness and additional exercises

In this section we provide additional evidence that the empirical patterns discussed previously are indeed the result of corporate responses to the minimum tax.

Our main sample consists of an unbalanced panel of corporations. Since the number of firms filing income tax increases significantly during the period, one might worry that results are purely driven by sample composition. We show that this is not the case by restricting the sample to a subset of approximately 12,000 firms observed in every year between 2013 and 2018. In panel A of A13 we present empirical revenue densities and in panel B we present profit margin densities for each year. The same pattern observed in the full sample is present in the balanced panel: an excess of firms reporting revenue slightly below L10 million and larger firms bunching around a 6% profit margin in 2014-2017, but not before or after the exemption threshold was substantially increased.

We perform two additional exercises that strengthen our case that the shift observed in declared profit margins by firms above the revenue exemption threshold was a response to the specific features of the minimum tax. First, as mentioned in 2.2, a small number of industries were subject to a reduced minimum tax rate of 0.75% instead of 1.5%. Corporations in those industries therefore face a kink in the tax schedule not at 6% rate of profit margin but at $\frac{0.0075}{0.25} = 3\%$, and according to our model we should observe excess mass around that threshold. In A14 we show that is precisely what happens: between 2014-2017, the distribution of profit margins for firms in these industries is shifted to the left when compared to corporations facing the 1.5% minimum tax and the peak of the distribution is exactly around 3%.

Second, we also investigate whether the increase in declared profit margins is induced by "lazy cost reporting" Best, Brockmeyer, et al., 2015. If there are fixed-costs in filing different cost line items, taxpayers might respond to revenue taxation by reducing the number of items filed and therefore generating an increase in profit margins, even if they were reporting truthfully under a profit taxation regime. We investigate whether there are significant changes in the share of cost line items reported in A15. Panel A presents the share across the 6% profit margin kink, for firms reporting revenue above L13 million, while panel B reports shares across the L10 million notch. If the observed changes in deductions/profit were being driven by filing costs, we should expect an increase in the share of items reported when firms report profit margins above 6% (Panel A) and a decrease for firms reporting above the exemption threshold (Panel B). Instead, shares are mostly smooth across the thresholds, and no different from the behavior of firms in 2018, when the exemption threshold was much higher and fewer firms were subject to the minimum tax. These results suggest it is unlikely that costly filing drives our results, at least on the extensive margin, and point to the importance of evasion under profit taxation.

mill", that exist only to supply plausible VAT credits and purchasing costs to existing firms Waseem, 2020; Mittal, Reich, and Mahajan, 2018. In Honduras, corporations are not required to file this detailed information to claim credits: the only mechanism for the Tax Authority to verify these deductions is to perform costly audits which, as we discuss above, are limited in number.

The introduction of the minimum tax might also have affected firm survival: corporations that might be viable under profit taxes might cease to be if they must pay taxes based on gross revenue, making the enterprise unprofitable. While we do not have a clear design that allows us to estimate the causal impact of the policy on firm exit, in I we perform a series of empirical exercises to assess whether affected firms were more likely to stop filing taxes after the 2014 reform. Our exercises rely on the assumption that firms likely to be affected by the reform – those declaring gross revenue above L10 million and profit margins below 6% before the reform – would have exited at a similar rate as those less likely to be affected. Our estimates are quite sensitive to sample selection and not robust to placebo tests that consider effects in the years before the introduction of the minimum tax. For those reasons, we are unable to make statements about the effects of the policy on firm survival.

We also consider whether the provision exempting firms declaring losses from paying the minimum tax might be a serious concern for our estimates. First, as we previously noted, our elasticity estimates assume that the incentives to bunch and profit margins are inversely correlated. If some corporations with lower profits systematically decide to declare losses instead of bunching, that would require higher elasticities for a given estimate of bunching below the L10 million threshold. Empirically, nonetheless, we do not see a strong response from firms to that exemption. As we document in A25, we do not see a sharp increase in firms not paying taxes above the L10 million threshold: the decrease in firms paying profit taxes is accompanied by an increase in firms paying the minimum tax. We also do not see a strong reaction of firms declaring profits right below zero in order to escape the minimum tax after 2014 (see A16). The reasons for this are likely manyfold, but seem to include ex-ante uncertainty on the exact nature of that exemption and whether firms declaring losses would indeed be exempt and/or required to be audited in order to claim the benefit. The behavior was also likely curbed by the existence of a net asset tax, discussed in F, that applied to all firms with net assets above L3 million, including those incurring in losses.

2.6 Assessing the impact of counterfactual policies

In order to make progress in quantifying the impacts of the minimum tax and alternative tax policies, we make stronger parametric assumptions about the profit function of firms and calibrate a model. We consider firms with isoelastic production costs and cost misreporting loss functions so we can rewrite 2.1 as follows:

$$\hat{\Pi}(y, \hat{c}) = y - \alpha_i - \frac{\theta_i}{1 + 1/e}\Big(\frac{y}{\theta_i}\Big)^{(1+1/e)} - \tau(y - \mu\hat{c}) - \frac{B_i}{1 + 1/\gamma}\Big(\hat{c} - c(y)\Big)^{(1+1/\gamma)} \tag{2.14}$$

Taxpayers are heterogeneous in three dimensions, characterized by the vector $(\theta_i, \alpha_i, B_i)$ that define productivity, production fixed cost and evasion ability, respectively. Heterogeneity in productivity allows firms to have different optimal production levels, while varying fixed costs generates a distribution of profit margins. We consider the maximization problem of firms under a simple profit taxation regime and calibrate the model using the parameters previously estimated and data from 2013, before the introduction of the minimum tax. We set $e = 0.99$, the upper bound revenue elasticity from our pooled sample, and use the estimates from Best, Brockmeyer, et al., 2015 for evasion cost elasticity $\gamma = 1.5$. We then calibrate the remaining parameters to match the distributions of reported revenue and reported costs, considering that firms evade 17% of profits through cost over-reporting (additional details are presented in Online Appendix E).

We perform two main exercises. First, we simulate the actual minimum tax system implemented in Honduras in 2014, with an exemption threshold for firms reporting gross revenue below L10 million and minimum effective tax of 1.5% for larger firms. Second, we consider an alternative to the minimum tax regime where the tax authority increases the average tax rate that large firms pay on profits.

First, consider the actual minimum tax implemented, in which firms reporting gross revenue below L10 million are exempt and those above face a minimum tax liability of 1.5% of gross revenue. We estimate that over 60% of corporations declaring revenue above the exemption threshold are liable for the minimum tax and that total government revenues increase by over 30% when compared to a flat profit tax rate of 25%[28]. This is attained by a 120% increase in the aggregate tax liability of firms paying the minimum tax and a decrease of 10% in aggregate profit for all firms in the economy. The fall in aggregate profits shows that, under the parameters of the actual policy implemented, the potential gains for firms when moving from profit to revenue taxation (decrease in losses from misreporting costs) are much smaller than losses from higher tax liability and production distortions.

Our calibrated model also allows us to quantify the strong incentives introduced by the exemption notch: the total tax liability of bunching firms is less than 25% of what they would have paid had they stayed above the threshold and paid the minimum tax. Despite that strong reaction at the margin, the increase in taxes paid by infra marginal firms dwarfs this loss: reduction in taxes from bunching firms is only 1% of total revenue from the minimum tax. While in our model bunching below the exemption threshold is exclusively driven by real production decisions, we provided evidence that at least part of this behavior seems to be explained by revenue misreporting. That finding highlights that, despite generating relatively small aggregate losses, notches can generate large horizontal inequities: firms otherwise similar might be liable for vastly different tax burdens simply due to willingness to misreport revenue.

We also assess the impact of alternative minimum tax specifications, in which we vary both the exemption threshold and the minimum tax rate. We highlight two features of our simulations. First, holding constant the minimum tax rate on gross revenues, increasing the exemption threshold only slowly decreases total revenue gains due to the long

[28] In these simulations we exclude taxpayers that were liable for Net Asset tax in 2013, since we do not model firms' asset accumulation and reporting decisions.

right tail of firm size. A L50 million exemption threshold, for example, still increases tax revenue by 23%. Second, small changes in the minimum tax rate generate large impacts in aggregate tax revenue and firms' profit, given the very broad base. Using the same L10 million exemption threshold and considering a minimum tax rate of 0.5%, for example, generates a tax revenue increase of less than 4% and aggregate profit loss of 0.5%. When comparing these magnitudes with the actual policy implemented, the decrease in tax revenue gain is driven by two forces. First, the minimum "allowable" profit margin is now lower: corporations with a 5% profit margin, for example, are allowed to pay an effective tax rate of 25%*5% = 1.25% when the minimum tax is 0.5%, while they would be liable for the 1.5% minimum tax under the previous regime. Second, firms with very low profit margins now only pay 0.5% in effective tax rate instead of 1.5%.

Our second exercise considers a progressive tax schedule in which firms declaring gross revenue above L10 million face an increase in *average* tax rates, without a change in the tax base (reported profits). We consider that the average tax rate is still 25% for firms below the exemption threshold, so firms also face a discontinuous change in tax liability when reporting revenue above L10 million and will have a strong incentive to bunch below the threshold. Unlike in our setting where firms with low profits benefit the most from bunching, here firms with high profit margins face the strongest incentives to locate below the notch, since they have the most to lose from higher tax rates. We present results for scenarios that consider an average profit tax rate between 30% and 50% in 2.5. Increasing the average tax rate by 5 p.p. to 30%, for example, would increase tax revenues by 12% and reduce aggregate corporate profits by 7%. In order to generate the same amount of tax revenue gains as the minimum tax, average taxes have to increase by 15 p.p. to 40%. While production efficiency is preserved under high tax rate profit taxation, evasion costs are exacerbated in this scenario and lead to large losses in aggregate profits, which fall by 20%.

2.7 Conclusion

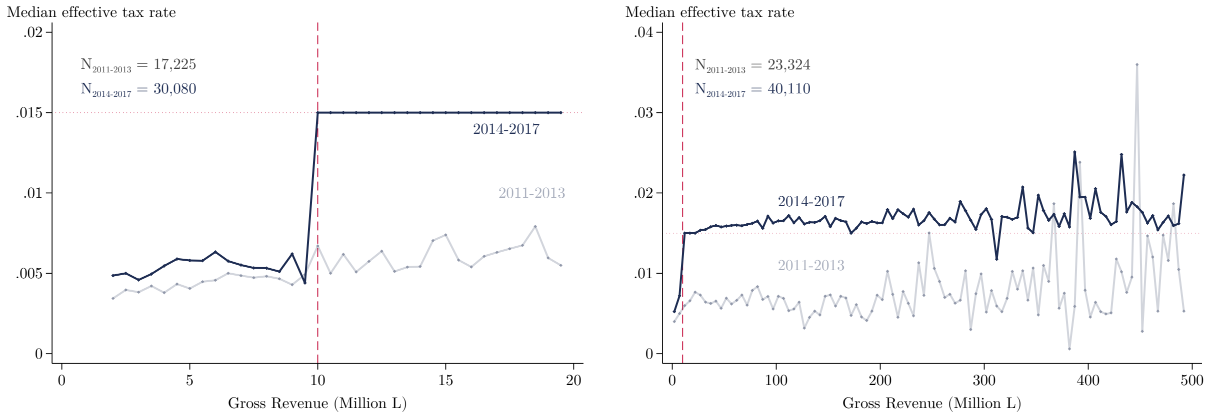
Minimum taxes are seen as effective tools for tax authorities to curb tax evasion in low- and middle-income countries and are at the heart of recent debates on global tax cooperation. In this paper we provide new evidence on corporate reaction to minimum taxes in Honduras.

We document meaningful evasion under profit taxation. Corporations liable for a minimum tax declare much larger profit margins when the incentives to over report costs disappears. We quantify that response and estimate that inflated costs allowed these firms to reduce tax liabilities by up to 17%. Curbing evasion through excessive reporting of deductions is costly to tax authorities 2017 since it requires time-intensive verification of receipts. We provide evidence that taxpayers exploit these limitations and use hard-to-verify cost categories to reduce tax liability. Improving oversight of these specific deductions seems to be a natural focal point for the efforts of tax authorities.

Using the response of taxpayers to the notch created by the exemption threshold, we bound the elasticity of reported revenue with respect to the net-of-tax rate at [0.35, 1]. These estimates are substantially higher than previous results for corporate taxpayers in similar settings and illustrate the limits faced by authorities in imposing high tax rates on broader bases. Whereas the elasticity of reported revenue summarizes responses both through real production and reporting decisions, we provide evidence that at least part of the observed response is due to revenue under-reporting. Firms with high revenue observability are less likely to strategically locate below the exemption threshold.

These results highlight the fact that behavioral responses of taxpayers are endogenous to the enforcement environment Fack and Landais, 2016; Joel Slemrod and Kopczuk, 2002. Building state capacity and properly designing rules to enforce tax compliance, therefore, might substantially change the trade-offs between available instruments. In the case of minimum taxes, improving the ability to assert the veracity of claimed deductions should decrease evasion through cost misreporting, making profit taxation more attractive. Improvements in independent verification of taxpayers' declared revenue, conversely, make broadening the tax base more attractive by reducing the elasticity of reported revenue.

Figure 2.1: Median effective tax rate across declared revenue distribution

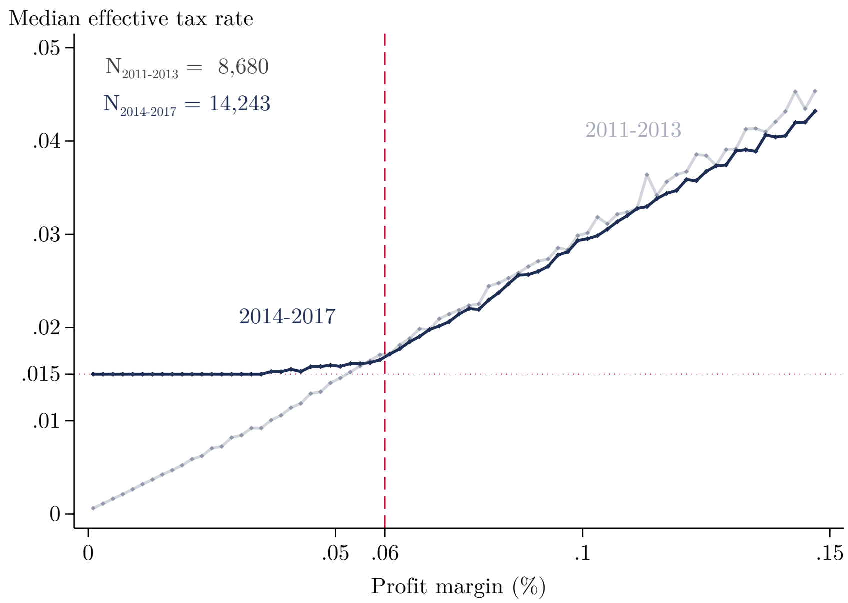


(a) Around L10 million exemption threshold

(b) Across gross revenue distribution

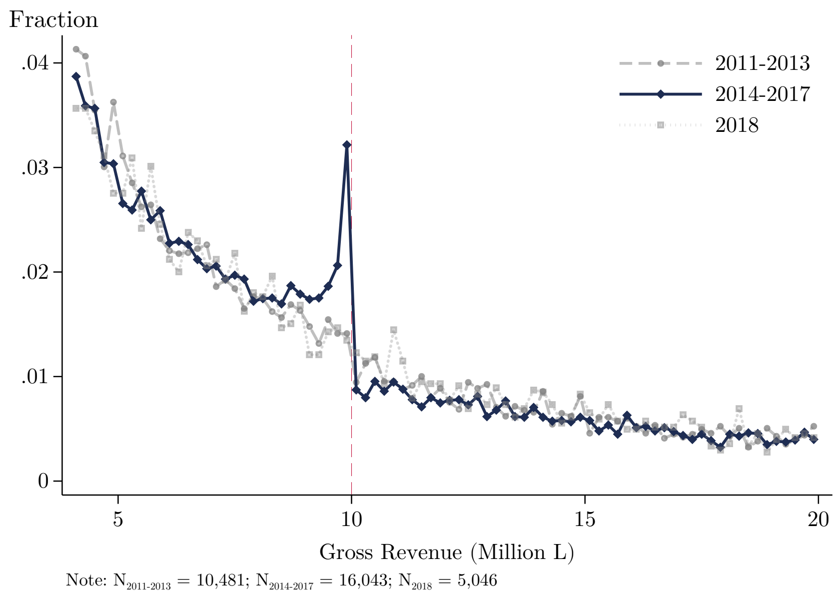
Note: This figure presents median effective tax rates, defined as the ratio between tax liability and gross revenue, for each bin of declared gross revenue. Panel A restricts the sample to taxpayers declaring gross revenue between L2-20 million, while panel B includes taxpayers with gross revenue between L2 - 500 million. Bins are L500,000 wide in Panel A and L5 million in Panel B.

Figure 2.2: Median effective tax rate across declared profit margin distribution



Note: This figure presents median effective tax rates, defined as the ratio between tax liability and gross revenue, for each bin of declared profit margin. The sample is restricted to firms declaring gross revenue above L13 million, and therefore inframarginal to bunching at the L10 million threshold. Bins are 0.2 p.p. wide.

Figure 2.3: Empirical Density of Gross Revenue around L10 Million threshold



Note: This figure presents the empirical density of gross revenues from firms pooled for three periods: 2011-2013 (before the minimum tax introduction); 2014-2017 (when the exemption threshold was L10 million); and 2018 (after the threshold for eligibility increased to L300 million). Bins are L200,000 wide. The sample is restricted to taxpayers declaring gross revenue between L4-20 million and excludes taxpayers exempt from the minimum tax.

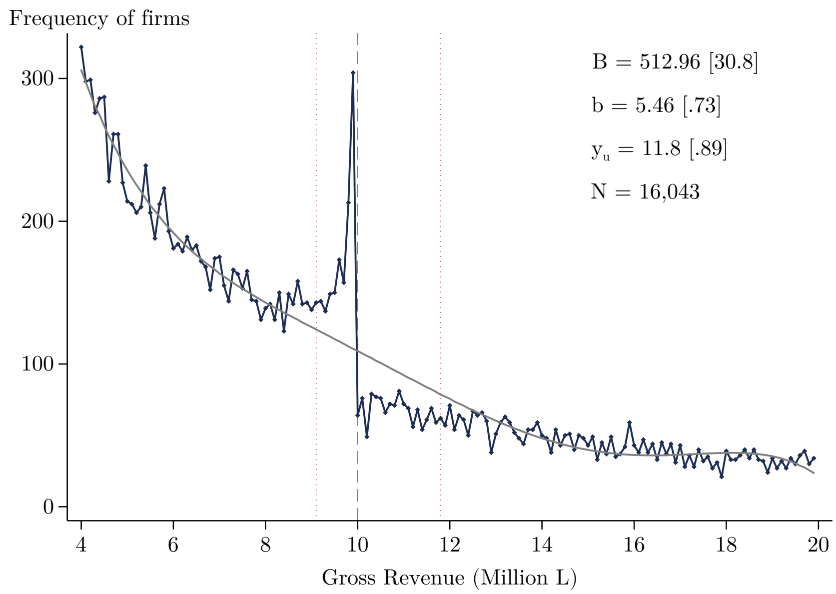
Figure 2.4: Empirical density of profit margins

(a) Empirical density of profit margins above L13 million - Pre and Post Minimum Tax

(b) Empirical density of profit margins in 2014-2017 - Below and above L10 million threshold

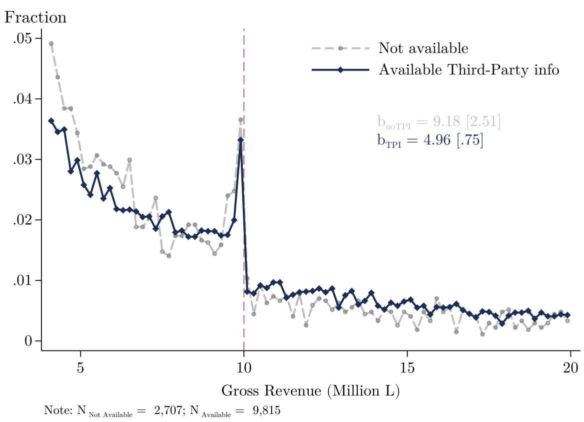
Note: These figures present the empirical density of positive reported profit margins. Panel A presents densities for firms with gross revenue above L13 million, before (2011-2013) and during (2014-2017) the existence of the minimum tax. Panel B present densities for the period of 2014-2017 of two groups of firms: those reporting gross revenue below L8 million (exempt from minimum tax) and those above L13 million (potentially liable for the minimum tax and infra-marginal to the bunching behavior at L10 million in revenue). Bins are 0.2 percentage points wide and the first bin starts at 0.1%, such that the 6% kink is the midpoint of a bin.

Figure 2.5: Empirical Density of Gross Revenue around L10 million threshold - Pooled Years (2014-2017)

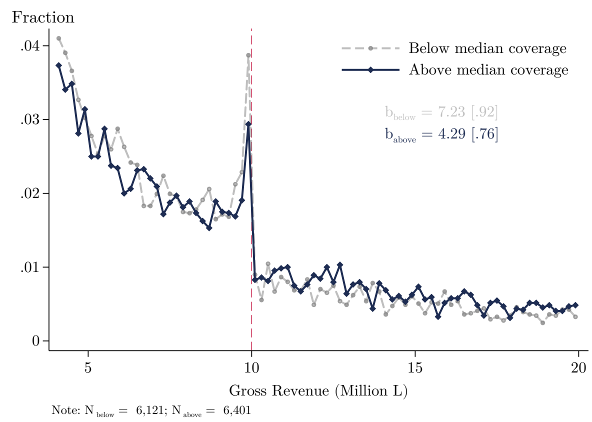


Note: This figure presents empirical and counterfactual densities of declared gross revenue for a pooled sample of firms (2014-2017). The dashed line marks the L10 million notch while the dotted lines mark the lower and upper bounds of the bunching region. We present the excess mass below the notch (B), the excess mass as a share of the predicted mass in the bunching region (b), the upper bound obtained from the convergence method (y_u) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are L100,000 wide.

Figure 2.6: Empirical gross revenue density by third-party status - pooled 2015-2017



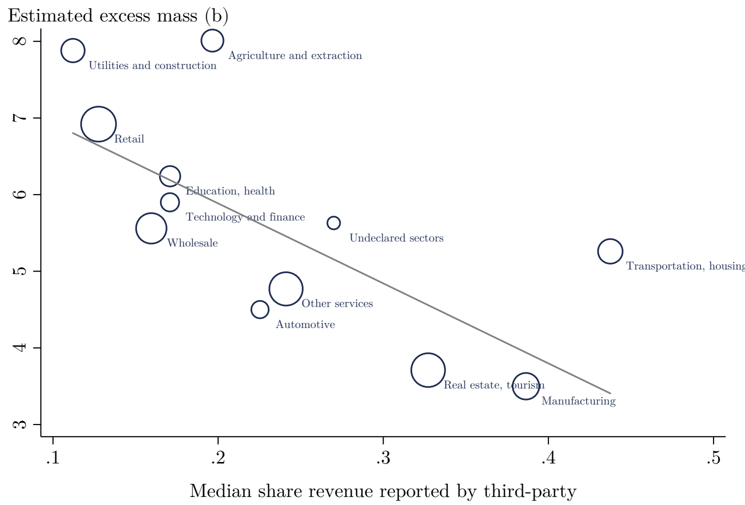
(a) Extensive margin



(b) Intensive margin

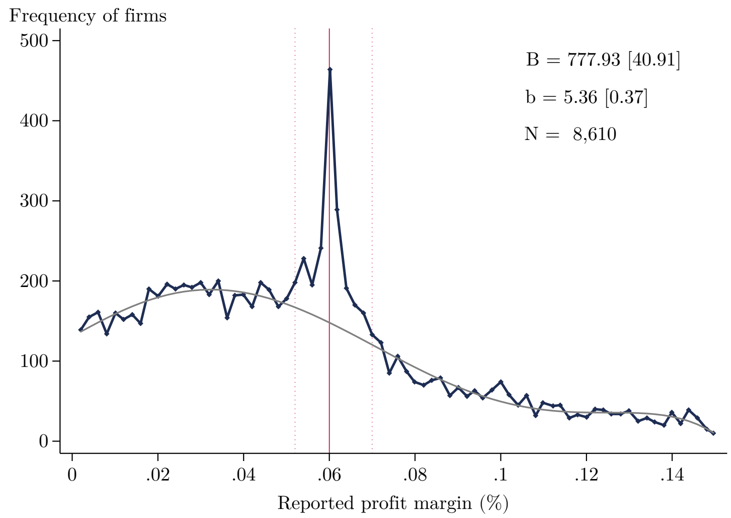
Note: These figure presents the empirical densities of declared gross revenue, pooled for the 2015-2017 period, exploring heterogeneity according to availability of third-party information on revenue. Panel A compares corporations for which no third-party information is available (gray line) with those for which some information is available (blue line). Panel B explores differences in the intensive margin of third-party information: it compares firms with below median (15%) share of declared revenue reported by third parties (gray line) with those above median (blue line). Bins are L200,000 wide.

Figure 2.7: Scatter plot of amount of bunching vs. revenue observability across industries



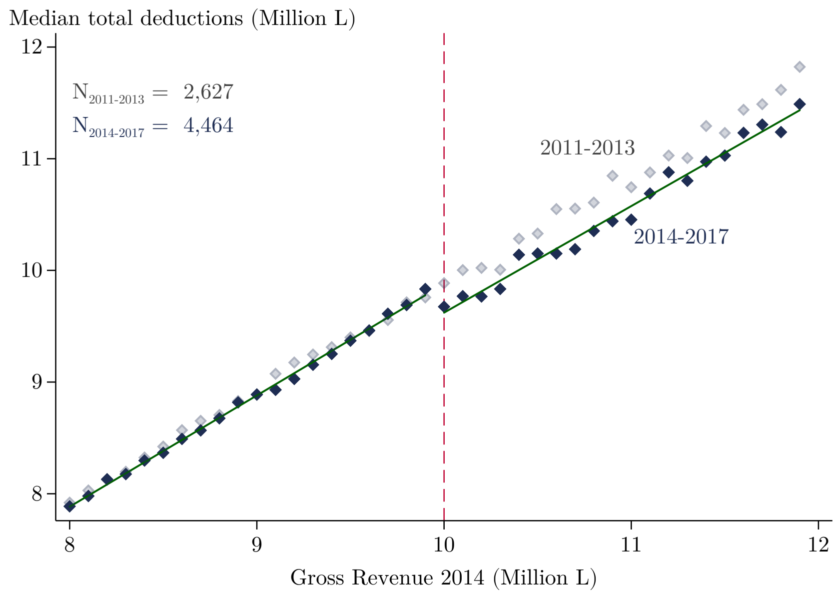
Note: This figure presents a scatter plot of estimated excess mass at the L10 million threshold and the median share of self-reported revenue also informed by third parties in each industry. Excess mass is defined as the excess number of firms bunching at the L10 million notch as a ratio of the predicted mass at the notch. The share of reported revenues is calculated in 2018, for firms declaring gross revenues in the interval L5-15 million. The size of markers is proportional to the reported sales in 2018 by industries.

Figure 2.8: Empirical Density around 6% profit margin threshold - Pooled Years (2014-2017)



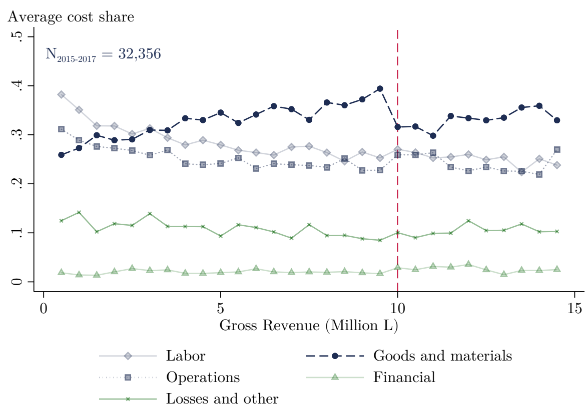
Note: These figures present the empirical and estimated counterfactual distributions of profit margins for a pooled sample of firms in the period period 2014-2017. The lower and upper bounds of the bunching region are determined visually. The solid red line marks the 6% kink while the dotted lines present the lower and upper bounds of the bunching region. We present the excess mass around the kink (B), the excess mass as a share of predicted density around the kink (b) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are 0.2 percentage points wide and the first bin starts at 0.1%, such that the 6% kink is the midpoint of a bin.

Figure 2.9: Median total deductions by gross revenue

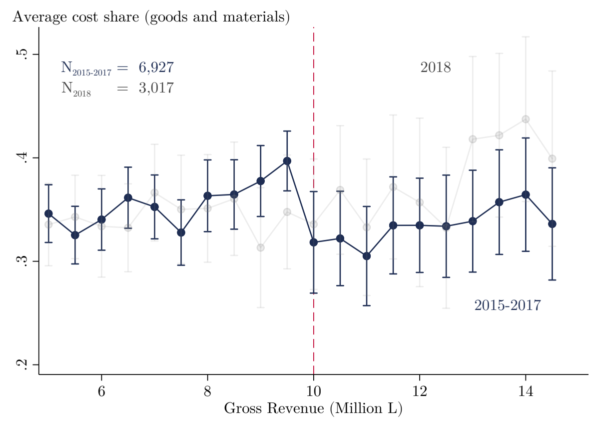


Note: This figure presents median reported total deductions by revenue bin for two groups: taxpayers in 2011-2013, before the introduction of the minimum tax, and 2014-2017, while the minimum tax was in place with a L10 million exemption threshold. Bins are L100,000 wide.

Figure 2.10: Cost line items as share of revenue



(a) All categories (2015-2017)



(b) Goods and materials (2015-2017 & 2018)

Note: These figures present cost line items as share of revenues in each bin. Panel A presents average shares in 2015-2017 for five cost categories: Labor, Goods and Materials, Operations, Financial, and Losses and other. Panel B focuses on Goods and Materials cost shares, separately for 2015-2017 and 2018. Bins are L500,000 wide in both panels. This sample only includes taxpayers using electronic declaration, for which we have detailed breakdown of cost items (approximately 80% of taxpayers per year) and excludes taxpayers with profit margins above the 99th and below 1st percentile of profit margin distribution.

Table 2.1: Share of revenue and taxes across gross revenue distribution

	2013		**2017**	
	(1) Revenue	(2) Taxes	(3) Revenue	(4) Taxes
Top 0.1%	28.1	32.2	28.5	34.3
Top 1%	63.0	68.6	63.4	67.2
Top 10%	91.0	91.9	90.8	93.2
Top 20%	95.8	96.2	95.6	97.1
Bottom 50%	0.6	0.9	0.5	0.7

Note: This table presents the share of total revenue and total taxes for corporations at the top 0.1%, top 1%, top 10%, top 20% and the bottom 50% of declared yearly gross revenues. Columns (1) and (2) refer to statistics in 2013, while columns (3) and (4) refer to 2017. Corporations exempt from all income taxes are excluded from the sample.

Table 2.2: Estimates by year for L10 million notch

Year	(1) Excess # Firms (B)	(2) Firms % counterfactual (b)	(3) y_u (upper bound)	(4) Δ Revenue (upper bound)	(5) ϵ_y (upper)	(6) ϵ_y (lower)
2014	84.63	4.21	12.10	2.10	1.33	0.20
	(11.14)	(0.86)	(0.96)	(0.96)	(1.53)	(0.06)
2015	120.54	6.12	13.00	3.00	2.61	0.40
	(10.12)	(0.90)	(0.92)	(0.92)	(1.53)	(0.08)
2016	142.05	5.55	11.40	1.40	0.61	0.40
	(18.63)	(1.28)	(1.00)	(1.00)	(1.44)	(0.13)
2017	144.54	5.22	11.40	1.40	0.61	0.35
	(11.21)	(0.82)	(0.90)	(0.90)	(1.30)	(0.06)
Pooled	512.96	5.46	11.80	1.80	0.99	0.35
	(30.80)	(0.73)	(0.89)	(0.89)	(1.40)	(0.05)

Note: This table presents estimates of change in reported revenue and elasticities for each year in the period 2014-2017 and also for all years pooled. The first column reports the estimated excess number of firms, defined above as $\sum_{b=y_L}^{y_N}(n_j - \hat{n}_j)$, while column 2 reports the ratio between excess mass and average counterfactual density in the bunching region. Column (3) presents the upper bound estimated using the convergence method and column (4) the change in revenue. Column (5) presents the upper bound estimates of reported revenue elasticity, defined in 2.10, while column (6) presents the lower bound estimates using the methodology presented in section 4.3. Bootstrapped standard-errors are presented in parentheses.

Table 2.3: Bunching at L10 million notch - by TPI and industries

	(1) Excess # Firms (B)	(2) Firms % counterfactual (b)	(3) Number Observations
Third-party information			
Below median TPI	253.33	7.23	6,121
	(20.77)	(0.92)	
Above median TPI	166.76	4.29	6,401
	(14.16)	(0.76)	
Industries			
Agriculture and extraction	45.75	8.01	865
	(3.62)	(0.97)	
Manufacturing	38.09	3.50	1,516
	(7.48)	(1.29)	
Utilities and construction	52.20	7.88	1,038
	(6.46)	(1.90)	
Automotive	16.70	4.50	650
	(6.08)	(2.07)	
Wholesale	65.11	5.56	1,880
	(8.93)	(0.91)	
Retail	71.64	6.92	1,884
	(13.01)	(1.69)	
Transportation, housing	31.65	5.26	1,174
	(10.09)	(2.31)	
Technology and finance	23.70	5.90	757
	(5.39)	(1.49)	
Real estate, tourism,other	48.30	3.71	2,530
	(9.40)	(0.67)	
Education, health, entertainment	37.00	6.24	1,050
	(10.72)	(2.15)	
Other services	62.23	4.77	2,298
	(10.74)	(1.51)	
Undeclared sectors	16.33	5.63	401
	(5.20)	(2.06)	

Note: This table presents estimates of bunching below the L10 million notch for firms with different levels of third-party information (TPI) (panel A) and in different industries (panel B). Column (1) presents the estimated excess mass of firms while column (2) presents the ratio between excess mass and average counterfactual density in the bunching region. Column (3) presents the number of firms for each exercise. Bootstrapped standard-errors are presented in parentheses.

Table 2.4: Estimated responses at the kink

Year	(1) Excess Mass (B)	(2) Bunching (b)	(3) Delta Profit ($\Delta\Pi$)	(4) Implied ϵ_y (no evasion)	(5) Estimated evasion (ϵ_y = 0.99)
2014	92.04	3.07	0.60	6.67	-8.52
	(10.60)	(0.42)	(0.10)	(0.99)	(1.48)
2015	192.76	5.18	1.00	11.11	-15.18
	(13.72)	(0.51)	(0.10)	(1.18)	(1.77)
2016	212.94	5.68	1.10	12.22	-16.85
	(14.98)	(0.55)	(0.10)	(1.24)	(1.85)
2017	212.68	4.57	0.90	10.00	-13.52
	(15.59)	(0.44)	(0.10)	(1.04)	(1.55)
Pooled	777.93	5.36	1.10	12.22	-16.85
	(42.00)	(0.38)	(0.10)	(0.91)	(1.36)

Note: This table presents estimates of change in reported profit margins and evasion estimates for each year in the period 2014-2017 and also for all years pooled. Column (1) reports the estimated excess number of firms while column (2) reports the ratio between excess mass and average counterfactual density in the bunching region. Column (3) presents estimated change in profit margins. Column (4) presents the implied revenue elasticity using the decomposition in 2.13 and considering no cost evasion. Column (5) computes the estimated cost evasion using the same decomposition and $\epsilon_y = 0.99$, our preferred estimate for the revenue elasticity upper bound. Bootstrapped standard-errors are presented in parentheses.

Table 2.5: Simulated impact of counterfactual increase in average profit tax

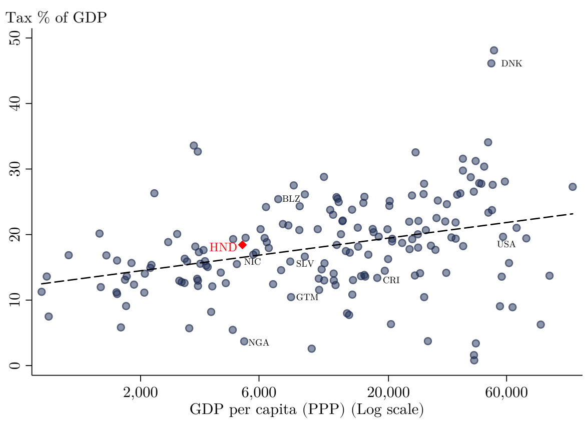
Average profit tax rate (%)	Tax revenue increase (%)	Change aggregate profits (%)
30	12.2	-6.9
35	22.2	-13.5
40	29.9	-20.0
45	35.9	-26.3
50	39.3	-32.5

Note: This table presents results of counterfactual policies where the average profit tax rate is increased for firms declaring gross revenue above L10 million, using the calibrated model. Columns (1) presents the average profit tax rate simulated in each scenario. Column (2) presents the total % increase in tax collection while column (3) presents aggregate profit losses.

For Online Publication

A Appendix Graphs and Table

Figure A1: Taxes as percentage of GDP across countries

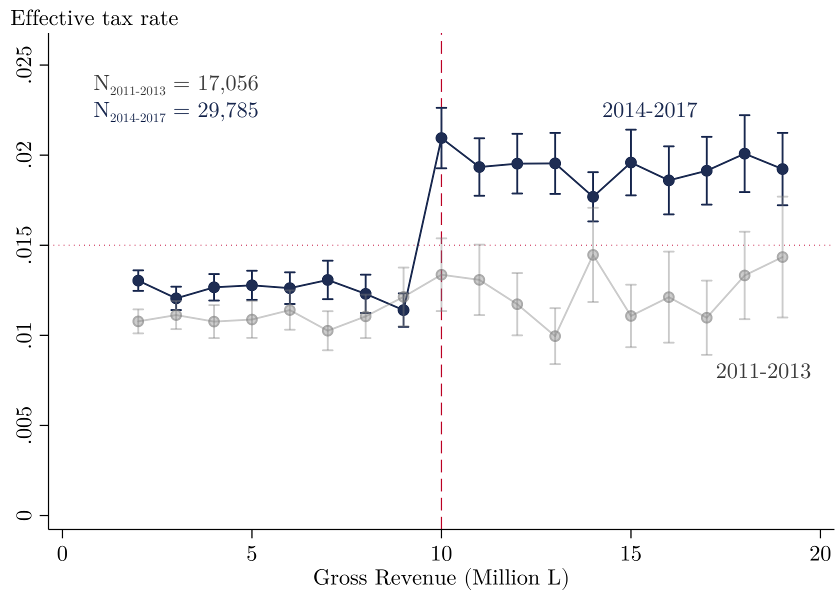


(a) Tax revenue

(b) Corporate income tax revenue

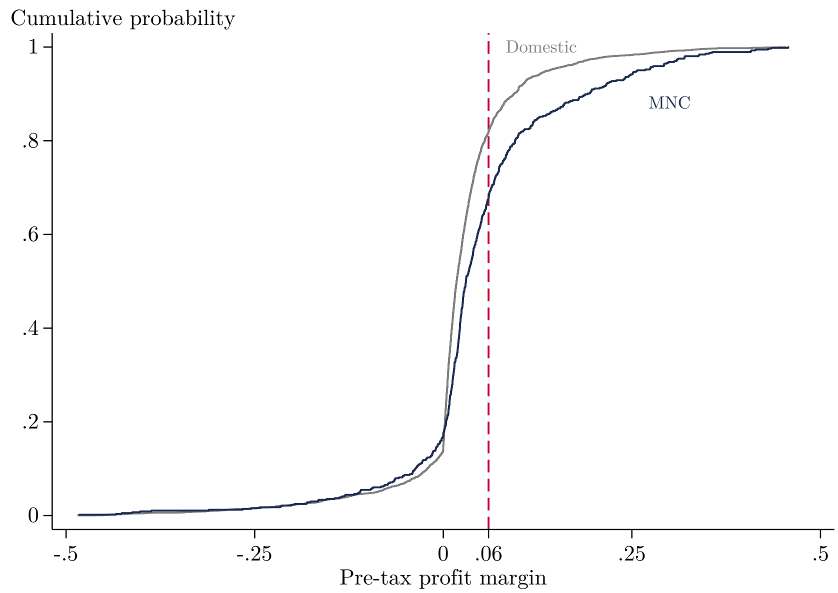
Note: These figures plot countries' tax revenue (Panel A) and corporate income tax revenue (Panel B) as percentage of GDP vs. (log) per capita GDP in 2016. Per capita GDP is expressed in PPP current dollars. Source: World Bank and International Monetary Fund (IMF) World Revenue Longitudinal Data.

Figure A2: Average effective tax rate across declared revenue distribution



Note: This figure presents mean and 95% confidence intervals of the effective tax rate, defined as the ratio between taxes due and gross revenue, for each bin of declared gross revenue. It documents that the minimum tax increased effective tax rates for corporations declaring more the L10 million: the average effective rate increases by approximately 1 p.p. around the threshold in 2014-2017, with no equivalent variation in 2011-2013, before the policy was introduced. Bins are L1 million wide. Sample is restricted to taxpayers declaring between L2-20 million and effective rate is trimmed at 99th percentile. The blue line refers to the pooled sample of taxpayers in 2014-2017, when the minimum tax was in place, while the gray line refers to the pooled sample of 2011-2013, before the introduction of the policy.

Figure A3: Pre-tax profit margin CDF - Domestic vs. Multinational corporations

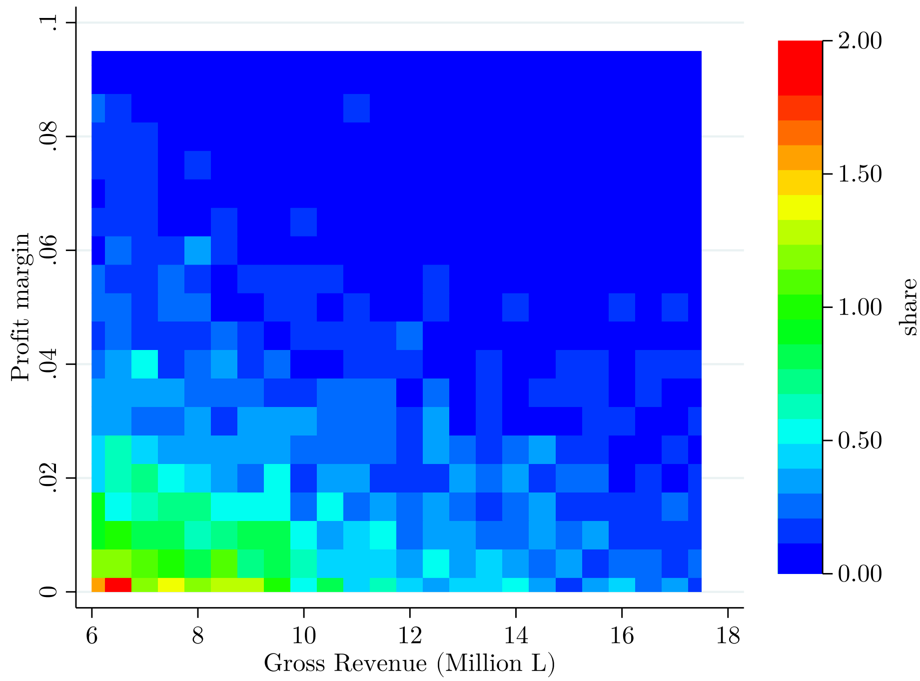


Note: This figure presents the cumulative distribution functions (CDF) of pre-tax profit margins by domestic and multinational firms in 2013, before the introduction of the minimum tax. The CDF of MNCs is shifted to the right, indicating higher declared profit margin across the distribution. In particular, approximately 30% of MNC declared profit margins above the 6% threshold that separates the minimum tax and profit regimes in 2014-2017, while this number is less than 20% for domestic corporations. MNCs are defined as taxpayers that present transfer pricing declarations at some point in 2014-2018. The sample is restricted to taxpayers declaring at least L8 million in gross revenue and the distribution is trimmed at the 1st and 99th percentiles.

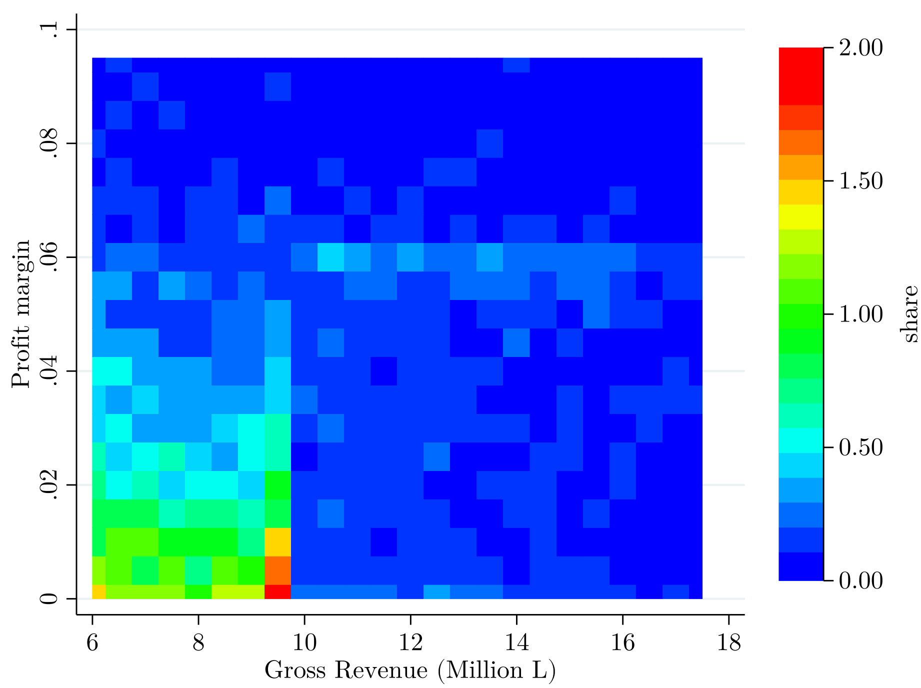
Figure A4: Total corporate tax liability and number of filers

Note: This figure presents, for each year in the period 2011-2018, the total number of corporate tax filers in our sample and the total tax liability. It documents the very significant increase in aggregate tax liability between 2013 and 2014, when the minimum tax was introduced. The sample excludes taxpayers exempt from all income taxes.

Figure A5: Heatmap of corporations on Revenue vs. Profit margin space



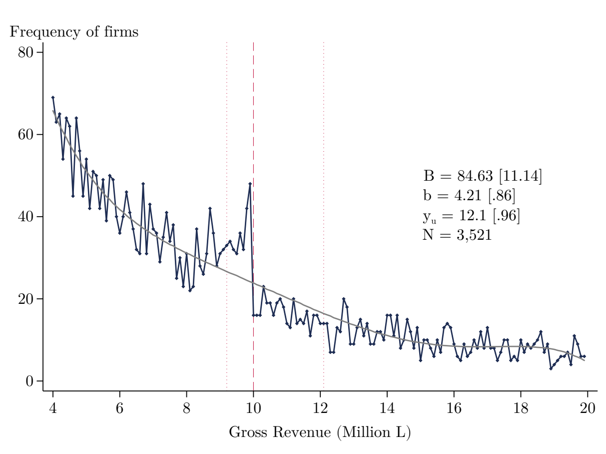
(a) 2011-2013 (before minimum tax)



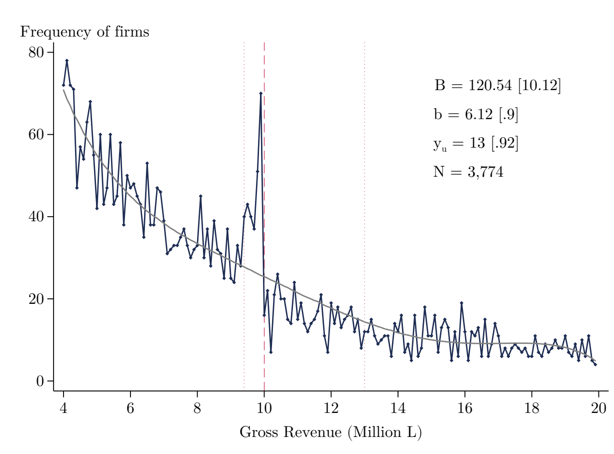
(b) 2014-2017

Note: These figures present heatmaps of the empirical distribution of corporations according to declared gross revenue (x-axis) and profit margin (y-axis). Panel A refers to the period 2011-2013, before the introduction of the minimum tax, while panel B refers to 2014-2017, while the minimum tax was in place with a L10 million exemption threshold. These figures summarized the response of firms to the

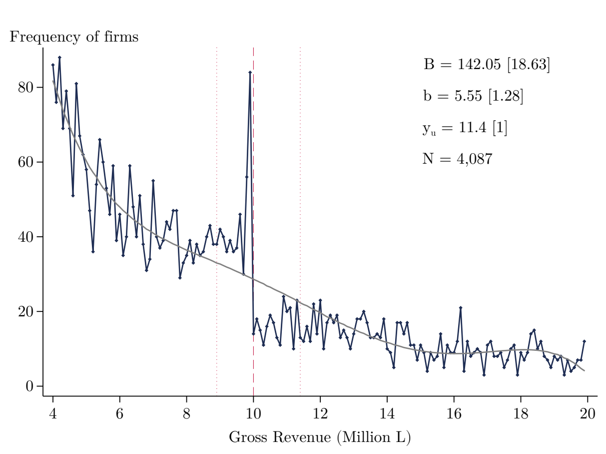
Figure A6: Empirical Density of Gross Revenue around L10 million threshold



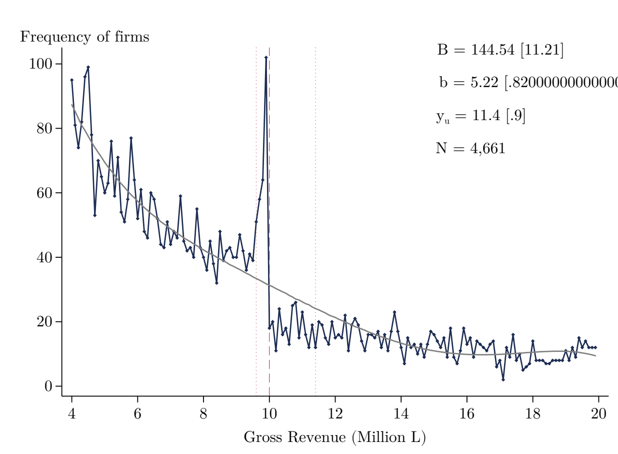
(a) 2014



(b) 2015



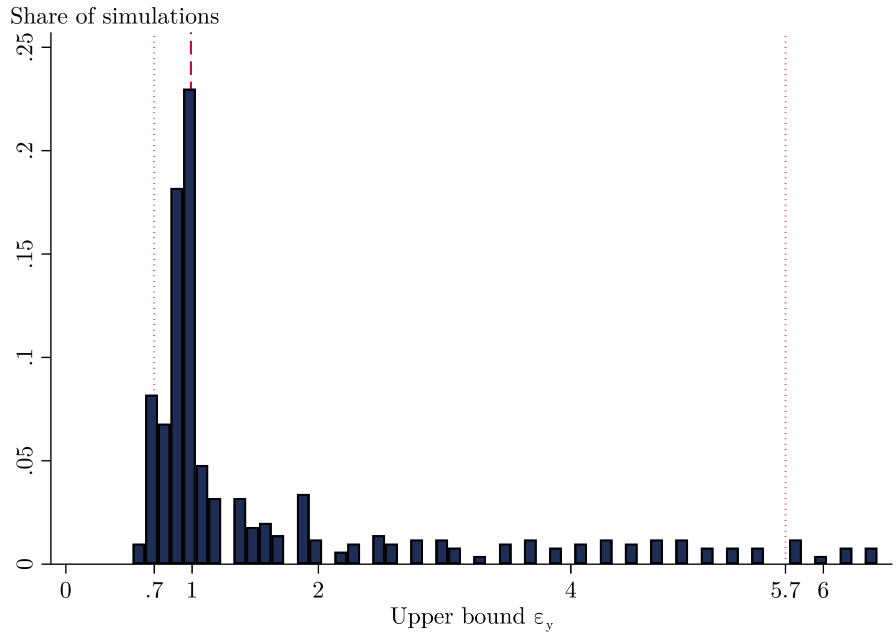
(c) 2016



(d) 2017

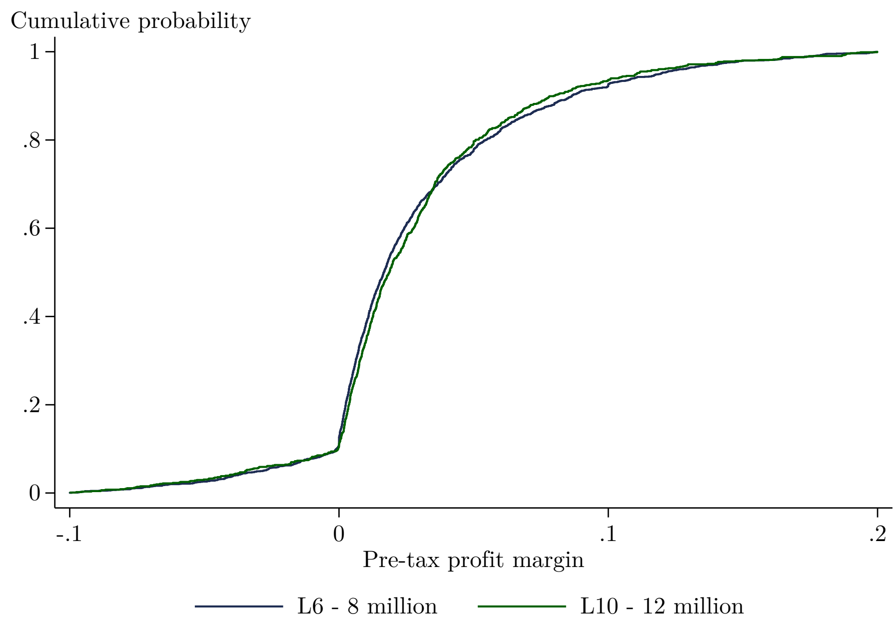
Note: These figures present empirical and counterfactual densities of declared gross revenue for each year in the period 2014-2017. The lower bound of the bunching region is chosen visually while the upper bound is obtained using the convergence method discussed in Section 4.3. The dashed line marks the L10 million notch while the dotted lines mark the lower and upper bounds of the bunching region. For each year we present the excess mass below the notch (B), the excess mass as a share of the predicted mass in the bunching region (b), the upper bound obtained from the convergence method (y_u) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are L100,000 wide.

Figure A7: Histogram of revenue elasticity bootstrap estimate for pooled sample (2014-2017)



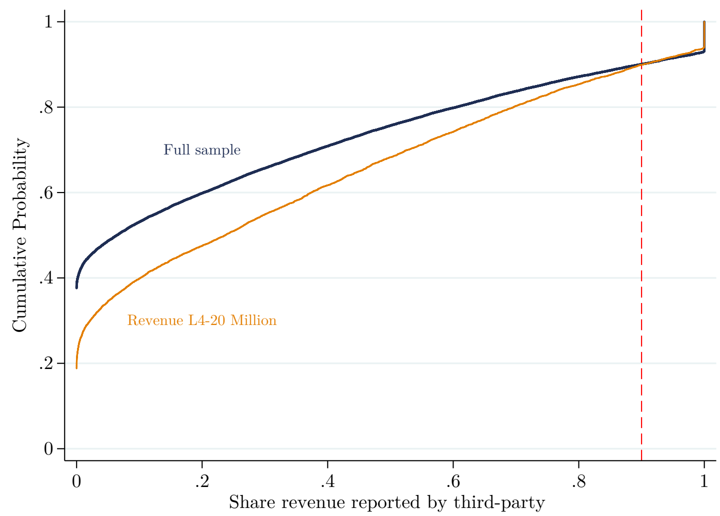
Note: This figure presents the histogram of 500 bootstrap estimates for the upper bound elasticity using the pooled sample of corporation filing in 2014-2017. The dashed line marks the point estimate of $\epsilon_y = 0.99$, while the two dotted lines mark percentiles 2.5 and 97.5 of the distribution. The empirical 95% confidence interval is $[0.7, 5.7]$. Bins are 0.1 wide.

Figure A8: CDF of profit margin for different revenue ranges



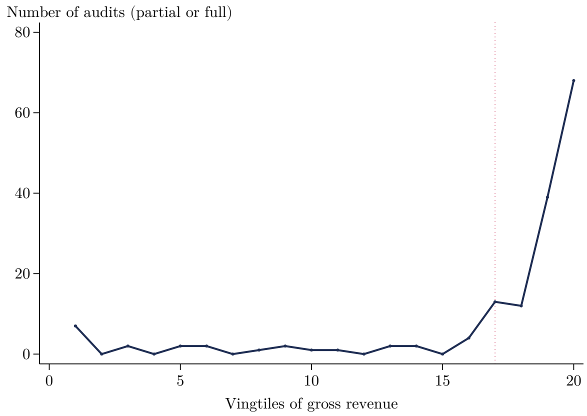
Note: This figure presents cumulative distribution functions (CDFs) of profit margins in 2011-2013, for corporations reporting gross revenues between L6 - 8 million and between L10-12 million. The distributions are trimmed at -10% and 20%. The profit margin distributions are similar across different revenue levels, suggesting the assumption used to estimate the lower bound revenue elasticity (using profit margin distribution below the L10 million notch as the counterfactual distribution above the notch) is reasonable.

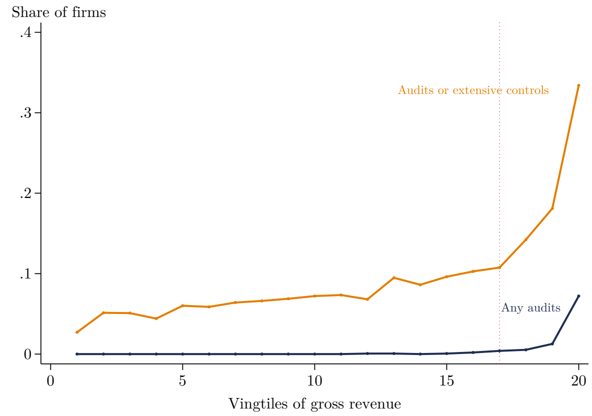
Figure A9: Share of revenue reported by third-parties



Note: This figure presents cumulative distribution functions (CDF) for the share of self-declared revenue that is also independently reported by third-parties. The sample is restricted to tax filers in 2018 and CDFs are presented separately for the entire sample (blue) and for those taxpayers declaring revenue in the vicinity of the L10 million threshold (L4 - 20 million) (orange). The dashed line shows that, in both samples, only 10% of taxpayers have 90% of more of their self-declared revenue independently reported by third-parties. For 40% of the total sample and 20% of the larger firms, no third-party reports are available.

Figure A10: Enforcement actions across revenue distribution



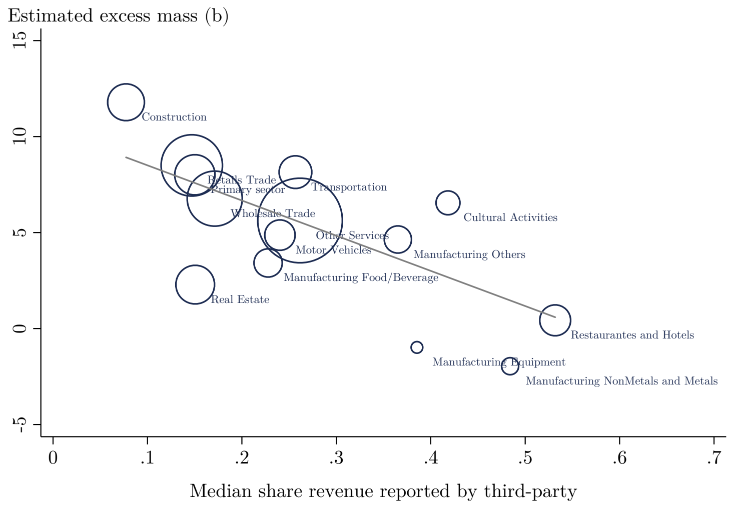


(a) Number of audits per vingtile of corporate revenue (2014)

(b) Share of firms receiving enforcement actions per vingtile of revenue (2018)

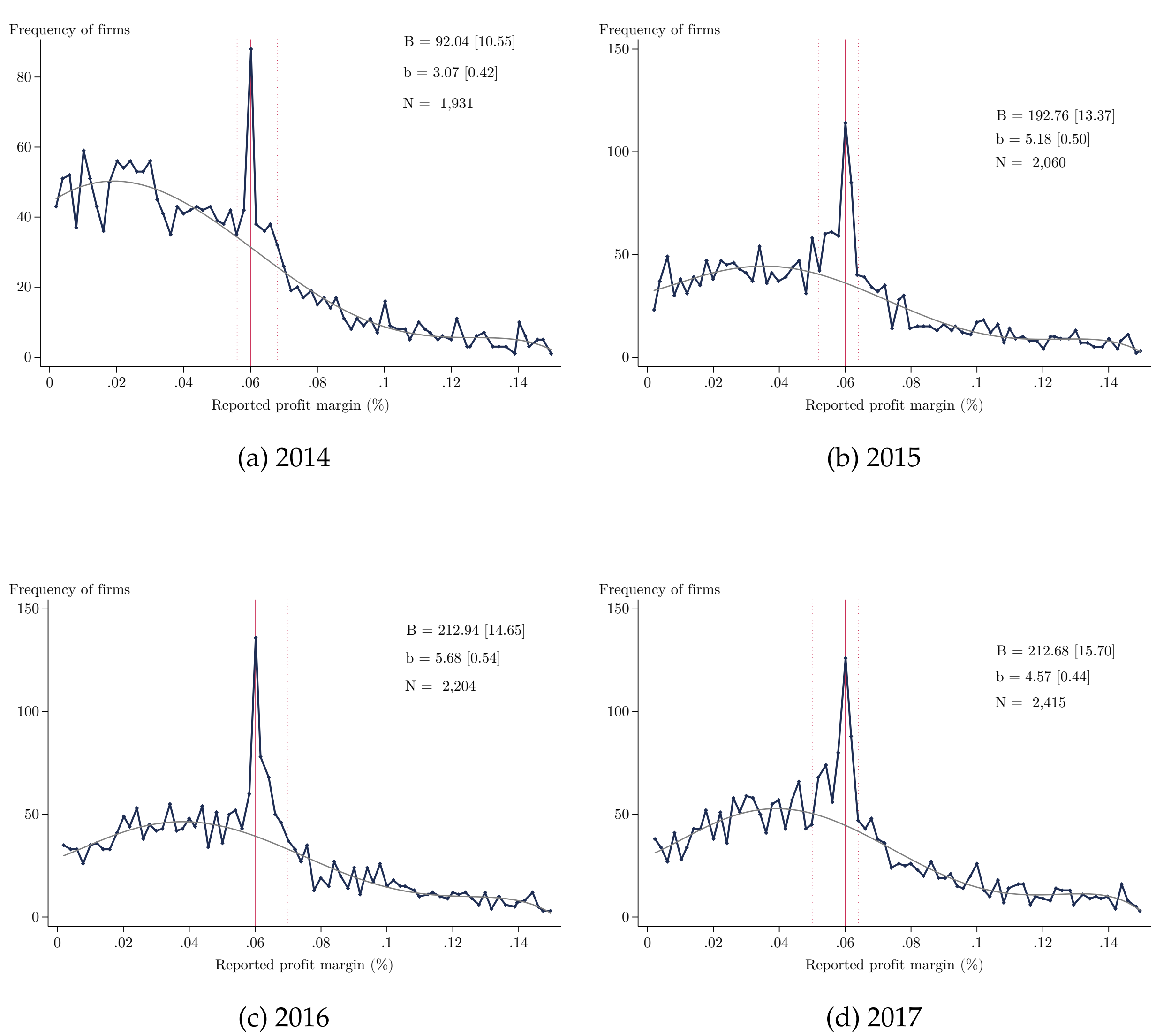
Note: These figures document the relationship between enforcement actions and firms' size. In panel (a) we compute the number of full or partial audits by gross revenue vingtile in 2014, while in panel (b) we compute the share of firms in each vingtile that faced an audit (blue line) or any kind of enforcement action (audit or extensive controls) (orange line) in 2018. The dotted line marks the 80th percentile of the size distribution, which approximately coincides with the L10 million exemption threshold for the minimum tax policy in 2014-2017.

Figure A11: Scatter plot of amount of bunching vs. revenue observability across industries - alternative sectoral definition



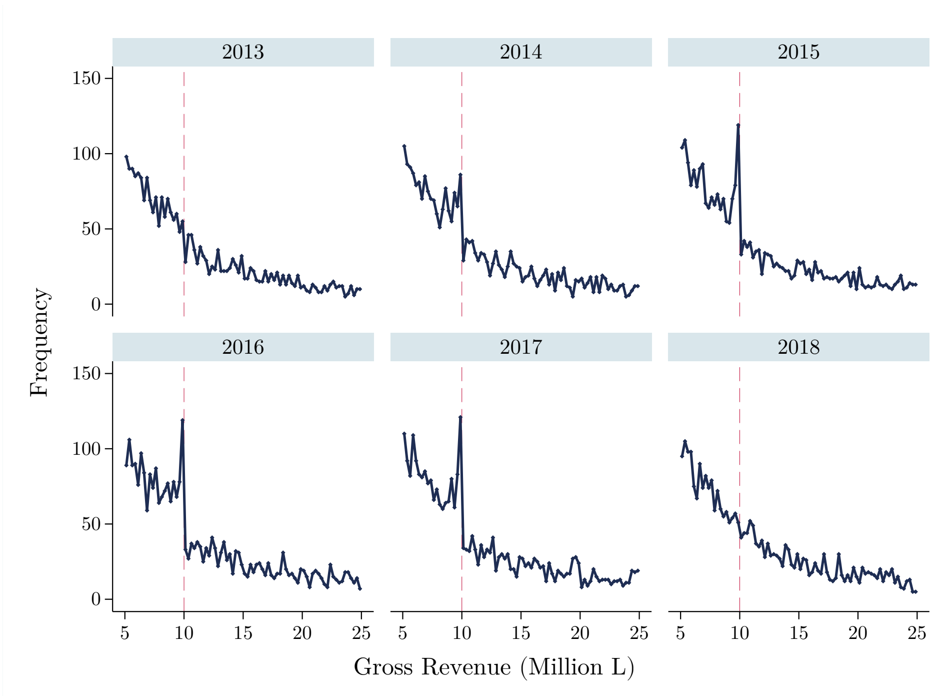
Note: This figure presents a scatter plot of estimated excess mass at the L10 million threshold and the median share of self-reported revenue also informed by third parties in each industry. Excess mass is defined as the excess number of firms bunching at the L10 million notch as a ratio of the predicted mass at the notch. The share of reported revenues is calculated in 2018, for firms declaring gross revenues in the interval L5-15 million. The size of markers is proportional to the reported sales in 2018 by industries. Industries are defined to approximate the same sectoral definition as in Almunia and Lopez-Rodriguez (2018).

Figure A12: Empirical Density of profits around 6% threshold

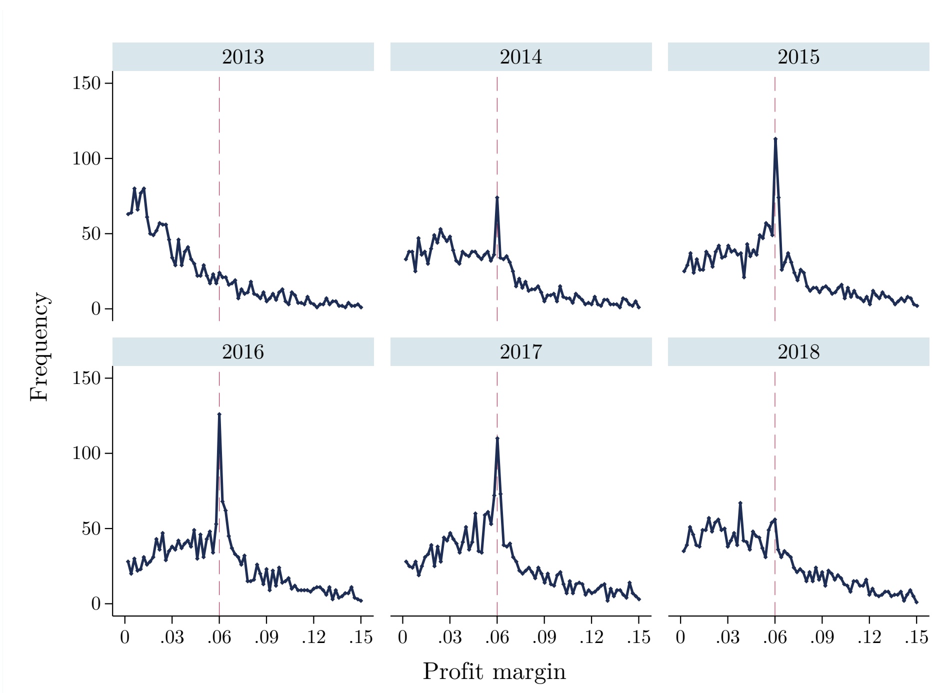


Note: These figures present the empirical and estimated counterfactual distributions of profit margins for each year in the period 2014-2017. The lower and upper bounds of the bunching region are determined visually. The solid red line marks the 6% kink while the dotted lines present the lower and upper bounds of the bunching region. For each year we present the excess mass around the kink (B), the excess mass as a share of predicted density around the kink (b) and the underlying number of taxpayers in each figure (N). Standard errors in brackets are obtained through bootstrapping. Bins are 0.2 percentage points wide and the first bin starts at 0.1%, such that the 6% kink is the midpoint of a bin.

Figure A13: Robustness: Balanced panel of corporations (2013-2018)



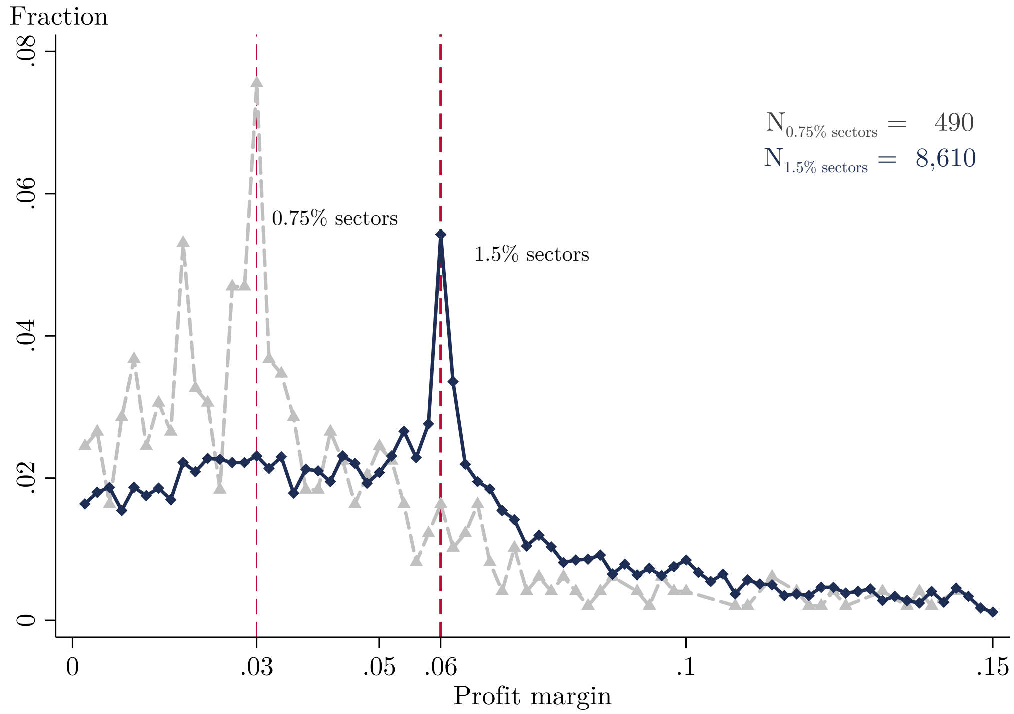
(a) Gross revenue empirical density



(b) Profit margin empirical density

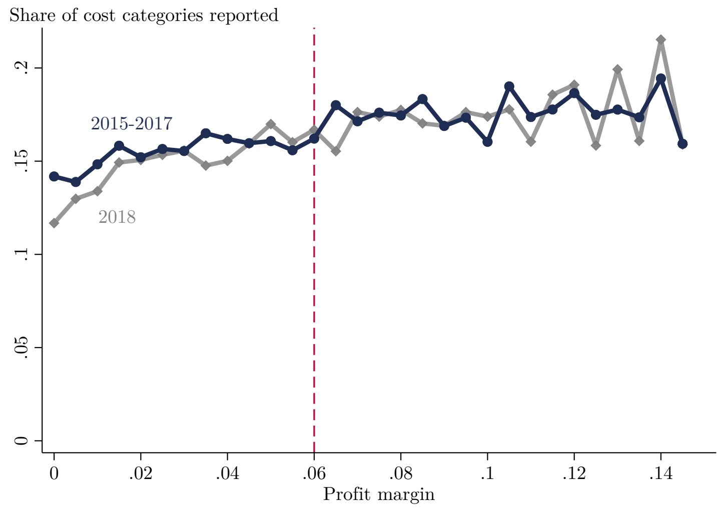
Note: This figure presents the empirical density of gross revenues (Panel A) and profit margins (Panel B) for a balanced panel of 12,172 firms, for each year in the period 2013-2018. It documents the same pattern observed for the full sample. Panel A shows a smooth distribution of gross revenue around the L10 million notch in 2013 and 2018, but significant excess mass between 2014-2017. This is evidence that taxpayers respond to the minimum tax by strategically bunching below the exemption threshold. Panel

Figure A14: Empirical Density around 6% profit margin threshold - 0.75% vs. 1.5% sectors (2014-2017)

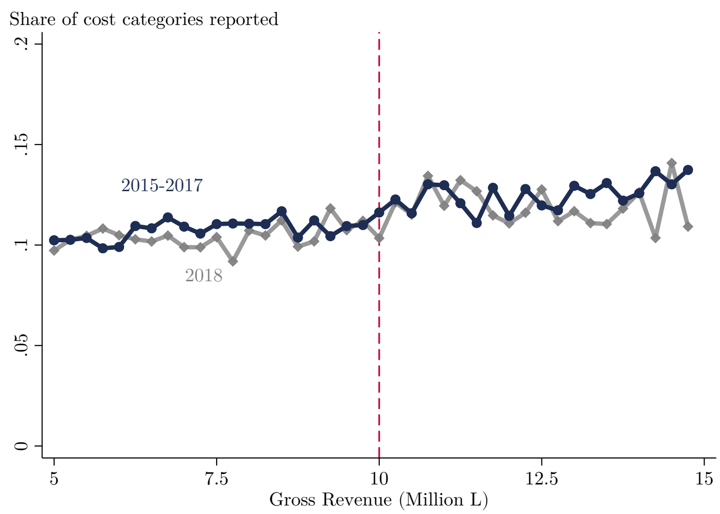


Note: This figure presents the empirical density of reported profit margins for firms subject to the 1.5% minimum tax (in solid blue) and those in sectors subject to the 0.75% rate (in dashed gray) for the period 2014-2017. The sample is restricted to firms reporting revenue above L13 million (infra marginal to revenue bunching). Bins are 0.2 p.p. wide and the first bins starts at 0.1% such that the 6% kink is the midpoint of a bin.

Figure A15: Average number of cost categories with positive deduction



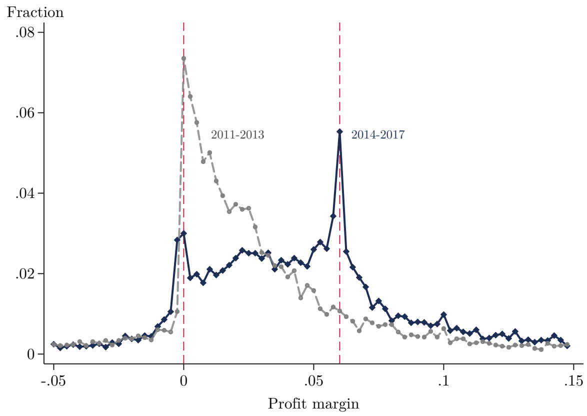
(a) Around 6% profit margin kink



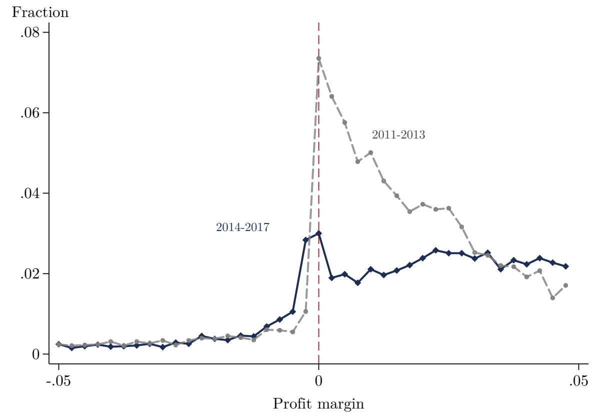
(b) Around L10 million notch

Note: This figure presents the average share of all cost categories reported by taxpayers in each bin. Panel (a) restricts the sample to taxpayers reporting revenue above L12 million and therefore infra-marginal to the revenue bunching behavior. Profit margin bins are 0.5% wide. The blue line represents declarations in the period 2015-2017, when the minimum tax affected a large number of taxpayers, while the gray line refers to declarations in 2018, when only a small subset of corporations were affected by the minimum tax. Panel (b) compares the usage of cost categories across the reported gross revenue distribution, for the period 2015-2017 (blue) and 2018 (gray). Both panels restrict the sample to taxpayers filing electronically, for which detailed cost categories are available.

Figure A16: Distribution of profit margins



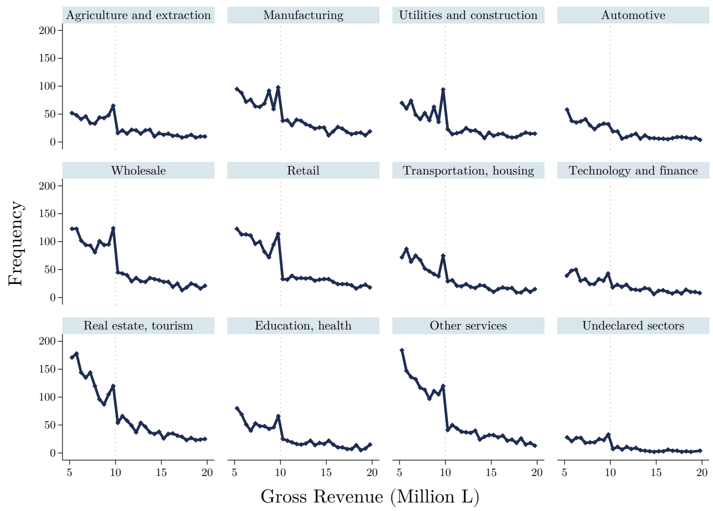
(a) Profit margin distribution - including losses



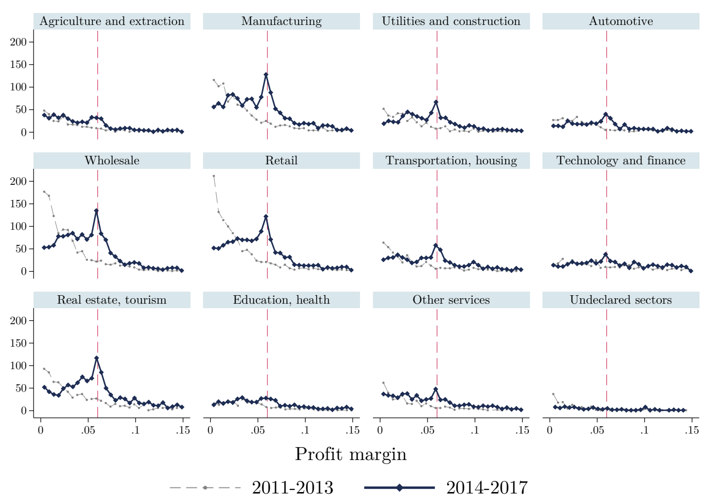
(b) Profit margin distribution - Zooming on losses

Note: These figures present the distribution of claimed profit margins for firms with revenue above L13 million, for the periods before (2011-2013) and after (2014-2017) the introduction of the corporate minimum tax.

Figure A17: Robustness: Behavioral responses by economic sector



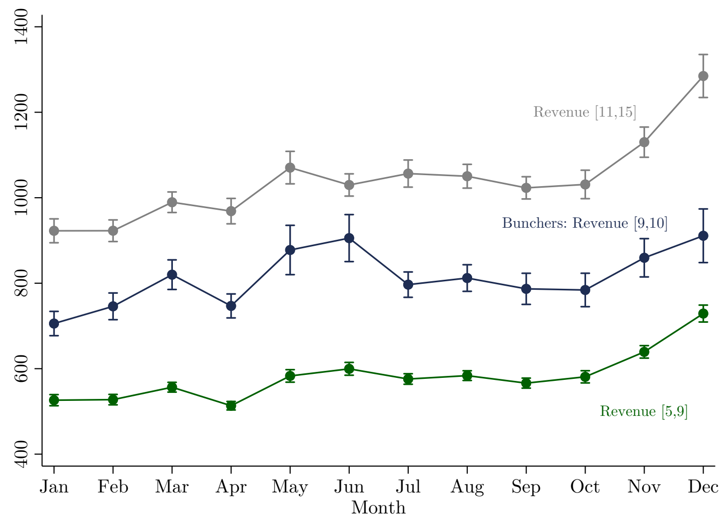
(a) Gross revenue empirical density



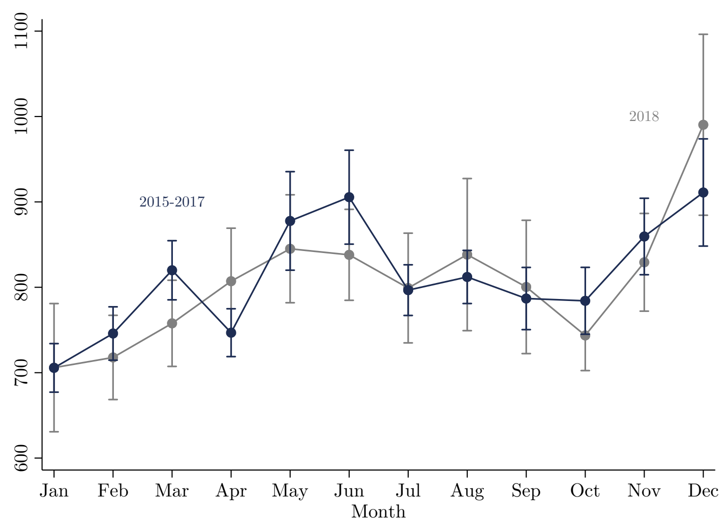
(b) Profit margin empirical density

Note: This figure presents the empirical density of gross revenues (panel A) and profit margins (Panel B) for firms in different economic sector for the period 2014-2017 pooled. Panel A documents that bunching below the notch is observed, in different degrees, for firms in the majority of sectors. Panel B shows that before the introduction of the minimum tax (2011-2013) the profit margin distribution is smooth around the 6% kink and presents a steep negative slope. With the introduction of the minimum taxation, the distribution shifts to the right and present excess mass around the kink. Bins are L500,000 wide in Panel A and 0.5 p.p. wide in Panel B. The sample in Panel B is restricted to firms reporting revenue above L13 million (infra marginal to the revenue bunching).

Figure A18: Monthly sales for firms with different yearly gross revenue



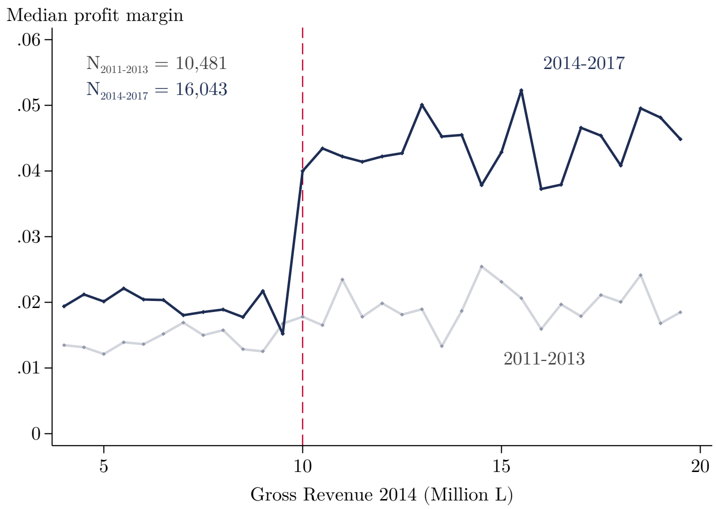
(a) 2015-2017 - Around L10 million notch



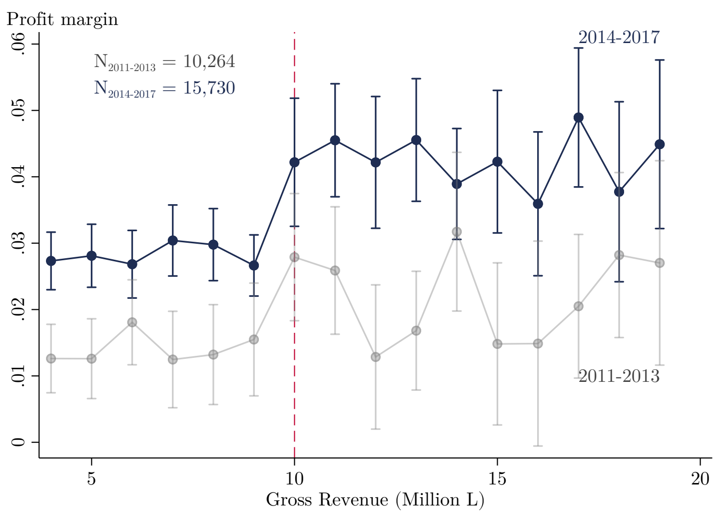
(b) 2015-2017 vs. 2018 - Below notch

Note: This figure presents average and 95% CI monthly sales separately for firms declaring gross revenue in L5-9 million, L9-10 million and L11-15 million bins on period 2015-2017 (Panel A), and for firms declaring gross revenue between L9-10 million in 2015-2017 and 2018. The sample is restricted to firms filing both monthly sales taxes and yearly income taxes and only include firm-year observations for which the total amount of monthly revenue falls within 5% of the total revenue declared in the yearly Income Tax Declaration,

Figure A19: Reported profit margin by gross revenue



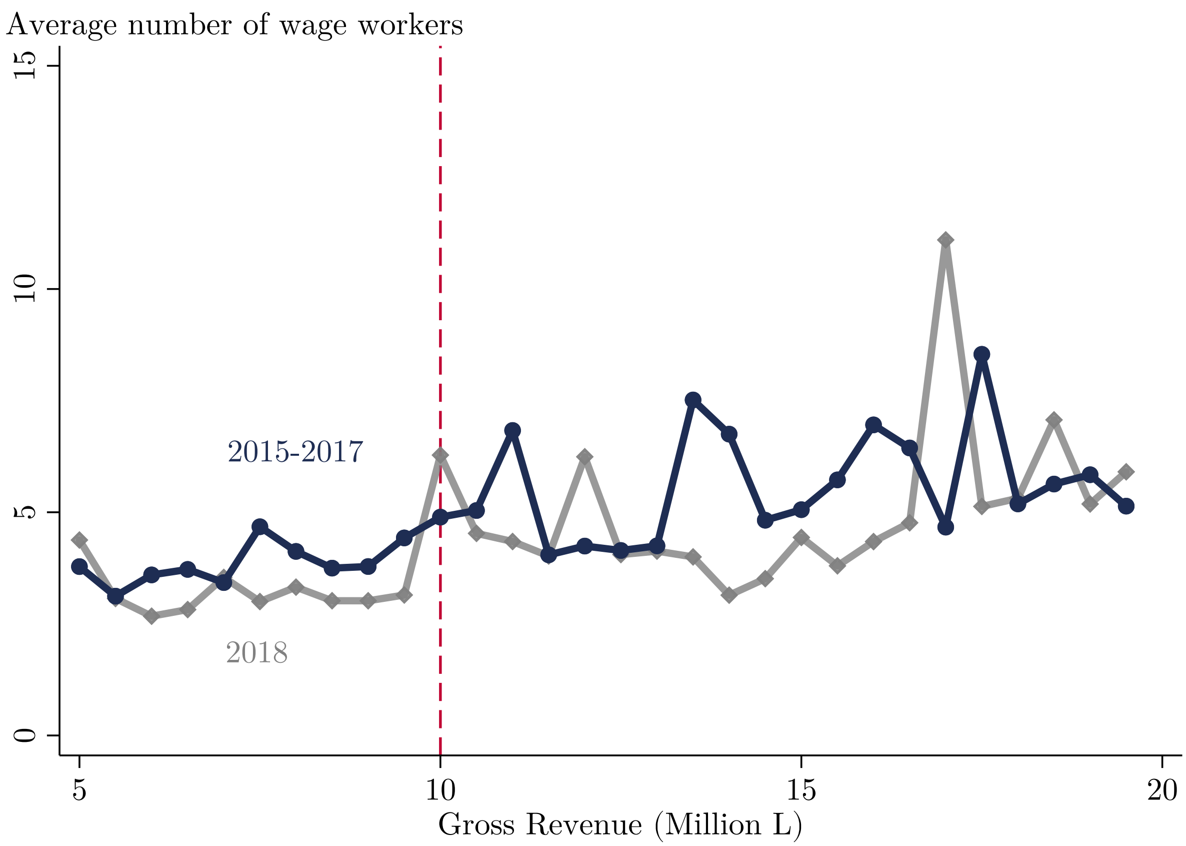
(a) Median profit margin



(b) Average profit margin

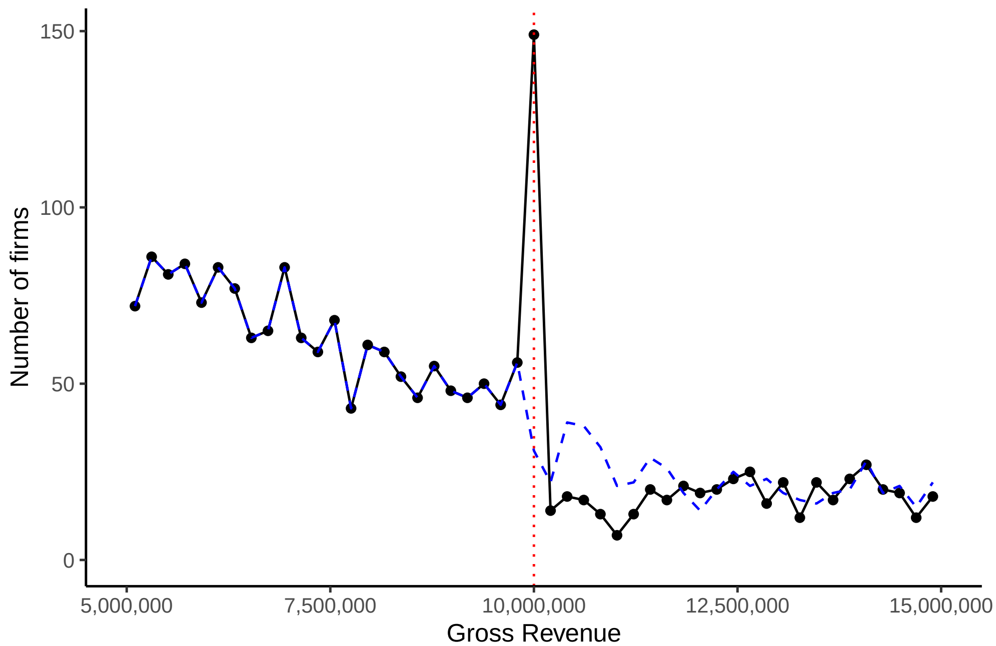
Note: This figure presents median (Panel A) and average with 95% CI (Panel B) reported profit margins by firms in two groups: 2011-2013, before the introduction of the minimum tax, and 2014-2017, then the minimum tax was in place for corporations with gross revenue above L10 million. The figure illustrates that corporations liable for the minimum tax increase their reported profit margins, consistent with the disappearance of the incentive to over report deductions in order to minimize tax liability. Bins are L500,000 wide in Panel A and L1 million in Panel B. Profit margins are trimmed at the 1st and 99th percentiles in Panel B.

Figure A20: Average number of wage workers by gross revenue (2015-2017 vs. 2018)



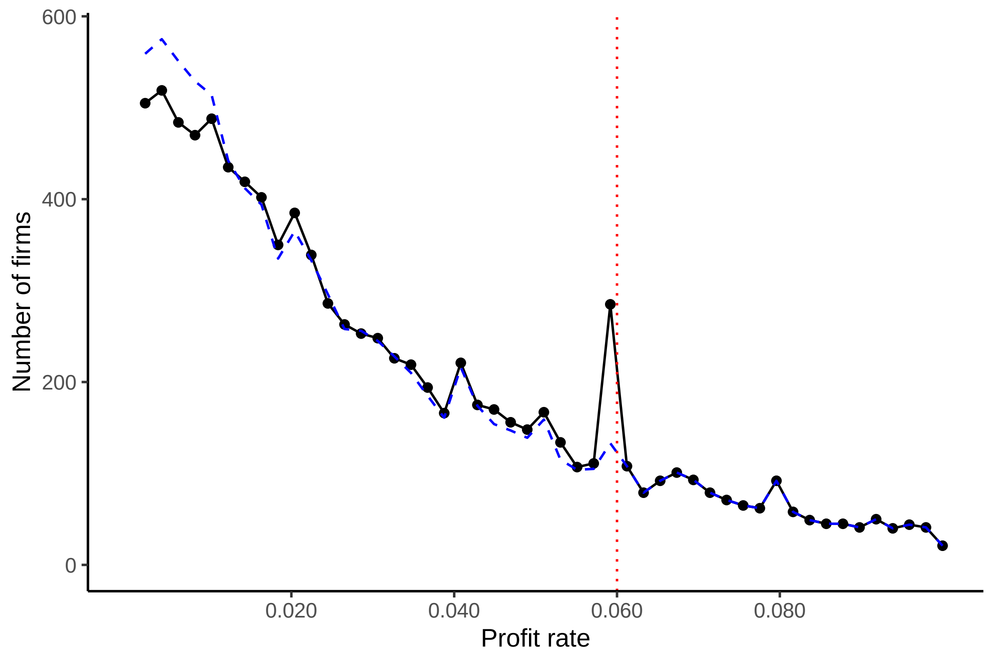
Note: This figure presents the average number of wage workers for firms in each gross revenue bin in 2015-2017 (when the exemption threshold was L10 million) and 2018 (when the threshold increased to L300 million). The number of wage workers is computed as the number of unique individuals for which the firm withheld taxes on wages. Firms are not required to withhold taxes if the total amount paid is below the exemption threshold for non-incorporated individuals, so these estimates of number of workers should be interpreted as lower bounds. The sample is limited to firms declaring at least one employee withholding (between 50-60% of firms declaring gross revenue above L5 million).

Figure A21: Calibrated model - bunching on L10 million revenue notch



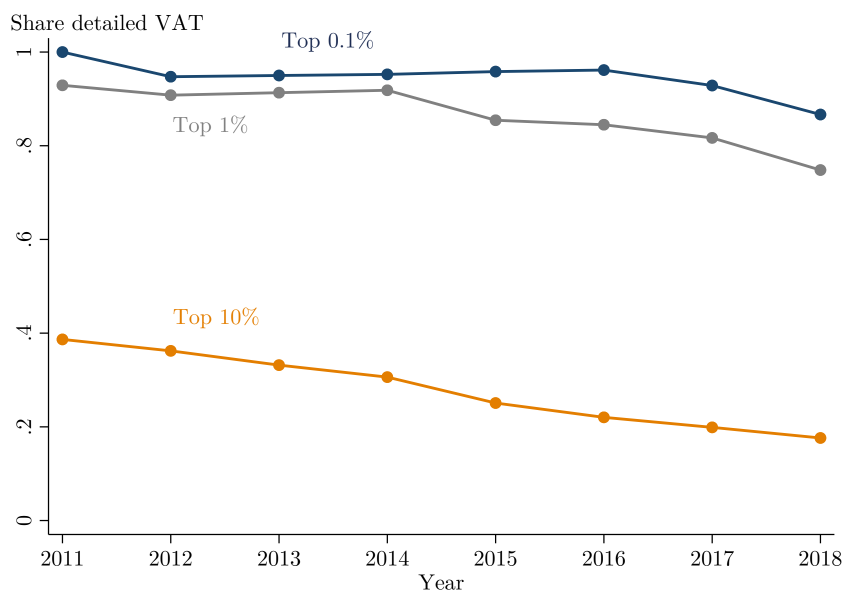
Note: This figure presents the density of simulated gross revenue using our calibrated model. The blue dashed line is the simulated density under profit taxation, while the solid black line presents the density under a Minimum Tax regime in which firms declaring above L10 million are subject to a minimum tax liability equivalent to 1,5% of their declared gross revenue.

Figure A22: Calibrated model - bunching on 6% profit margin kink



Note: This figure presents the density of simulated profit margin using our calibrated model. The blue dashed line is the simulated density under profit taxation, while the solid black line presents the density under a Minimum Tax regime in which firms declaring above L10 million are subject to a minimum tax liability equivalent to 1,5% of their declared gross revenue. We restrict the simulated sample to firms that choose to declared gross revenue above L12 million and are therefore infra-marginal to the bunching behavior at the notch.

Figure A23: Share of taxpayers mandated to file detailed VAT purchases



Note: This figure presents, for each year in the period 2011-2018, the share of taxpayers in each revenue group (top 0.1%, top 1% and top 10%) that are defined as medium or large. These are the taxpayers with an obligation to file individualized information on their purchases to claim VAT deductions, generating independent information on suppliers' revenues. The list of medium and large taxpayers was defined in 2011 and has not changed since. Groups are mutually exclusive, so the group defined as top 1% exclude taxpayers in the top 0.1% and the 10% group all those in the top 1% and 0.1%. The sample excludes taxpayers exempt from all income taxes.

Table A1: Alternative order of polynomial - gross revenue distribution

	(1) Excess # Firms (B)	(2) Firms % counterfactual (b)	(3) y_u (upper bound)	(4) Δ Revenue (upper bound)	(5) ϵ_y (upper)	(6) ϵ_y (lower)
Order p = 3	604.30	8.82	14.70	4.70	5.96	0.60
	(34.16)	(0.70)	(0.68)	(0.68)	(1.36)	(0.09)
Order p = 4	569.91	6.78	12.90	2.90	2.45	0.50
	(30.91)	(0.60)	(0.61)	(0.61)	(1.09)	(0.06)
Order p = 6	494.55	5.69	12.30	2.30	1.58	0.35
	(25.07)	(0.63)	(0.78)	(0.78)	(1.29)	(0.04)

Note: This table presents results from replicating the exercises performed in 2.2 using different order of polynomials to estimate the counterfactual distribution of gross revenue for the sample of pooled taxpayers in 2014-2017. The baseline specification uses polynomial regression of order five, while in this table we present results using polynomials of order three, four and six.

Table A2: Number of enforcement actions per year

	Partial Audit (1)	Full audits (2)	Extensive (3)	Any control (%) (4)
2011	0	140		
2012	3	108		
2013	0	41		
2014	1	157		
2015	2	66		
2016	0	5		
2017	15	12	1,039	3.8
2018	98	50	2,672	9.2

Note: This table presents the number of partial audits (1), full audits (2), extensive controls (3) and the share of taxpayers receiving any of those enforcement actions per year. The numbers refer to taxpayer in the sample of corporations used in this paper.

Table A3: Alternative order of polynomial - Profit margin distribution

Year	(1) Excess Mass (B)	(2) Bunching(b)	(3) Delta Profit	(4) Estimated evasion (ϵ_y = 0.99)
Order p = 3	779.64	5.38	1.10	-16.85
	(46.24)	(0.39)	(0.10)	(1.44)
Order p = 4	834.22	6.05	1.20	-18.52
	(40.69)	(0.38)	(0.10)	(1.39)
Order p = 6	788.99	5.49	1.10	-16.85
	(39.83)	(0.37)	(0.10)	(1.35)

Note: This table presents results from replicating the exercises performed in 2.4 using different order of polynomials to estimate the counterfactual distribution of profit margin for the sample of pooled taxpayers in 2014-2017. The baseline specification uses polynomial regression of order five, while in this table we present results using polynomials of order three, four and six.

Table A4: Estimated responses at the kink (Robustness - output evasion)

Year	(1) Delta Profit ($\Delta\Pi$)	(2) (ϵ_y = 0.5)	(3) Estimated evasion (ϵ_y = 0.99)	(4) (ϵ_y = 2)
2014	0.60	-9.68	-8.74	-6.81
	(0.10)	(1.57)	(1.57)	(1.57)
2015	1.00	-16.77	-15.83	-13.90
	(0.10)	(1.88)	(1.88)	(1.88)
2016	1.10	-18.55	-17.61	-15.67
	(0.10)	(1.98)	(1.98)	(1.98)
2017	0.90	-15.00	-14.06	-12.13
	(0.10)	(1.64)	(1.64)	(1.64)
Pooled	1.10	-18.55	-17.61	-15.67
	(0.10)	(1.40)	(1.40)	(1.40)

Note: Note: This table presents estimates of change in reported profit margins and evasion estimates for each year in the period 2014-2017 and also for all years pooled. Column (1) presents estimated change in profit margins while columns (2) through (4) computes the estimated output evasion under different real elasticity (ϵ_y) scenarios.

Table A5: Cost evasion responses across economic sectors

Year	(1) Excess Mass (B)	(2) Bunching(b)	(3) Delta Profit	(4) Estimated evasion (ϵ_y = 0.99)
Agriculture and extraction	38.35	6.06	1.20	-18.52
	(10.02)	(2.40)	(0.50)	(7.96)
Manufacturing	153.10	7.86	1.60	-25.18
	(13.45)	(1.16)	(0.20)	(3.87)
Utilities and construction	61.86	5.55	1.10	-16.85
	(7.52)	(0.89)	(0.20)	(3.02)
Automotive	49.72	7.91	1.60	-25.18
	(6.47)	(1.54)	(0.30)	(5.15)
Wholesale	132.19	5.66	1.10	-16.85
	(13.63)	(0.81)	(0.20)	(2.75)
Retail	85.16	3.71	0.70	-10.18
	(10.82)	(0.58)	(0.10)	(2.00)
Transportation, housing	69.39	8.09	1.60	-25.18
	(8.57)	(1.76)	(0.30)	(5.83)
Technology and finance	28.68	3.80	0.80	-11.85
	(7.22)	(1.18)	(0.20)	(4.07)
Real estate, tourism,other	93.89	4.15	0.80	-11.85
	(12.37)	(0.67)	(0.10)	(2.34)
Education, health, entertainment	31.71	4.59	0.90	-13.52
	(7.82)	(1.48)	(0.30)	(4.92)
Other services	34.21	4.04	0.80	-11.85
	(8.00)	(1.28)	(0.30)	(4.26)
Undeclared sectors	-1.93	-1.11	-0.20	4.82
	(4.98)	(2.64)	(0.50)	(8.71)

Note: This table presents estimates of change in reported profit margins and cost evasion for firms by economic sector, pooled for the 2014-2017 period. The first column reports the estimated excess number of firms (B) while column (2) reports the ratio between excess mass and average counterfactual density in the bunching region (b). Column (3) presents estimated change in profit margin, while column (4) present changes in cost misreporting using the decomposition in 2.13.

B Approximating the elasticity with notch

In this section we adapt the exercise of Henrik J. Kleven and Waseem (2013b) and Henrik J Kleven (2018) to obtain the elasticity formula when taxpayers face a notch instead of a kink. The intuition behind the derivation is that we try to recover what would have been the kink that would "replicate" the same behavior observed with the notch. We start by considering the average slope of the indifference curve of the marginal buncher: this IC is tangent to the threshold using the hypothetical kink with slope $(1-\tau^*)$ and has slope of $(1-t_0-\Delta t)$ at the point $y^t+\Delta Y$. In our case, $t_0=0$ since the effective marginal rate on revenue is zero below the threshold, and $\Delta t=\tau_y=0.015$. We can write

$$\frac{\int_{y^T}^{y^t+\Delta Y} I'(y)dy}{\Delta Y} \approx \frac{I'(y^T)+I'(y^t+\Delta Y)}{2} = \frac{(1-\tau^*)+(1-t-\Delta t)}{2} = \frac{(1-\tau^*)+(1-\tau_y)}{2}$$

The implicit tax rate faced by corporations is the change in tax liability when we change the reported revenue from above the threshold to exactly at the notch:

$$\begin{aligned} t^* = \frac{T(y^t+\Delta Y)-T(y^T)}{\Delta Y} &= \frac{\tau_y(y^t+\Delta Y)-\tau_\pi(y^T-\hat{c})}{\Delta Y} \\ &= \tau_y + \frac{\tau_y y^T+\tau_\pi(y^T-\hat{c})}{\Delta Y} \end{aligned}$$

Combining the fact that we have these two approximations to the slope of the IC in that region, and that $\Delta t=0.015=\tau_y$, we can write:

$$\begin{aligned} 1-t^* &= \frac{(1-\tau^*)+(1-\tau_y)}{2} \\ \tau^* &= \tau_y + 2\left(\frac{\tau_y Y^T+\tau_\pi\big(y^T-\hat{c}\big)}{\Delta Y}\right) \end{aligned}$$

Plugging in the expression for τ^* in the usual expression for obtaining revenue elasticity when facing changes in marginal taxes we obtain:

$$\begin{aligned} \epsilon_{y,(1-t)} = \frac{\frac{\Delta Y}{Y^T}}{\frac{\Delta\tau^*}{(1-\tau^*)}} &= \frac{\Delta Y}{Y^T}\left(\frac{1-\tau^*}{\tau^*-t_0}\right) \\ &= \frac{\Delta Y}{Y^T}\left(\frac{1-\tau^*}{\tau_y+2\left(\frac{\tau_y Y^T-\tau_\pi(Y_T-\hat{c})}{\Delta Y}\right)}\right) \\ &= \left(\frac{1}{\tau_y\big(2+\frac{\Delta Y}{Y^T}\big)-2\tau_\pi\frac{(Y_T-\hat{c})}{Y_T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2(1-\tau) \end{aligned}$$

Some things are worth noting from this expression. First, for a firm with zero reported profit at the notch ($y^T = \hat{c}$), than the expression above simplifies to

$$\epsilon_{y,(1-\tau)} \approx \left(\frac{\Delta Y}{Y^T}\right)^2 \left(\frac{(1-\tau)}{\Delta\tau}\right)\left(\frac{1}{2+\frac{\Delta Y}{Y^T}}\right)$$

which is exactly the same expression in Henrik J. Kleven and Waseem (2013b). This is the expression we use to calculate the upper bound of elasticities presented in the text, since the taxpayer with highest incentive to bunch has profits only marginally above zero.

Second, note that if profit margin is exactly 6%, then it's true that

$$\tau_y\Big(2+\frac{\Delta Y}{Y^T}\Big) - 2\tau_\pi 0.06 = 0.015\Big(2+\frac{\Delta Y}{Y^T}\Big) - 2(0.25)0.06 = 0.015 * \frac{\Delta Y}{Y^T}$$

and the elasticity becomes

$$\begin{aligned}
\epsilon_{y,(1-t)} &= \left(\frac{1}{\tau_y(2+\frac{\Delta Y}{Y^T}) - 2\tau_\pi\frac{(Y_T-c)}{Y_T}}\right)\left(\frac{\Delta Y}{Y^T}\right)^2(1-\tau) \\
&= \left(\frac{Y^T}{0.015\Delta Y}\right)\left(\frac{\Delta Y}{Y^T}\right)^2(1-\tau) \\
&= \left(\frac{\Delta Y}{Y^T}\right)\frac{(1-\tau)}{\tau_y} = \epsilon_{kink}
\end{aligned}$$

For a taxpayer with 6% reported profit margin, the exemption threshold represents a kink, not a notch, since their tax liability changes continuously around the cutoff.

C Estimation of revenue elasticity lower bound

Following Bachas and Soto (2021), we compute the lower-bound of average revenue elasticity considering that firms with different profit levels (generated by heterogeneity in fixed-costs) will face different incentives to bunch. First, recall that firms with counterfactual profits above 6% or below 0% will not decide to bunch, since they are not affected by the minimum tax. Second, for firms within that profit range, the incentive to bunch is directly proportional to their costs: firms with high costs (low profit margins) will have a strong incentive to bunch since their tax liability at the threshold will be small, while not bunching means a much larger tax liability based on their revenues.

Let $\Psi(y_0, c_0)$ be the joint distribution of revenue and costs. We can then express the amount of bunching taxpayers as

$$\begin{aligned}
B &= \int_c \int_{Y^T}^{Y^T+\Delta Y} \Psi(y_0, c_0) dy dc \\
&= \int_c \int_{Y^T}^{Y^T+\Delta Y} \phi_y(y_0)\phi(c_0) dy dc \\
&= \int_{Y^T}^{Y^T+\Delta Y} \phi_y(y_0) \int_{c_0} \phi(c_0) dc dy \\
&= \int_{Y^T}^{Y^T+\Delta Y} \phi_y(y_0) \int_0^{m(y_0)} \phi(m_0) dm dy
\end{aligned}$$

where in the second line we assume that the cost and revenue distributions are independent; in the third line we make it explicit that, for any given level of revenue, there is a cost region that will induce bunching; and in the last line we re-write the expression as a function of profit levels instead of cost, and make it explicit that, for any given revenue level, only low-profit taxpayers will bunch, the upper threshold of which depends on the revenue level. Intuitively, for taxpayers very close to the notch, all those potentially affected by the minimum tax will decide to bunch, whereas those farther from it will only bunch if the differential tax liability is large due to their low profits.

In order to connect the cost/profit levels that induce bunching at each revenue level, recall that we previously computed that, for the marginal buncher at revenue level $Y^T + \Delta Y$, we can compute the revenue elasticity as

$$\epsilon_{y,(1-t)} = \left(\frac{1}{\tau_y(2 + \frac{\Delta Y}{Y^T}) - 2\tau_\pi \frac{(Y_T - \hat{c})}{Y_T}} \right) \left(\frac{\Delta Y}{Y^T} \right)^2$$

We can rewrite this equality putting the reported cost $\hat{c}$ in evidence:

$$\hat{c}^* = Y^T \Big(1 - \frac{\tau_y}{\tau_\pi} \Big) - \frac{\tau_y}{\tau_\pi} \frac{\Delta y}{2} + \frac{(\Delta y)^2}{2\epsilon_y \tau_\pi Y^T}$$

For a given revenue level and elasticity, $\hat{c}^*$ is the cost at the threshold that would make a taxpayer indifferent between bunching and staying above the notch. Any taxpayer with costs above that level, i.e. a lower profit margin, would decide to bunch.

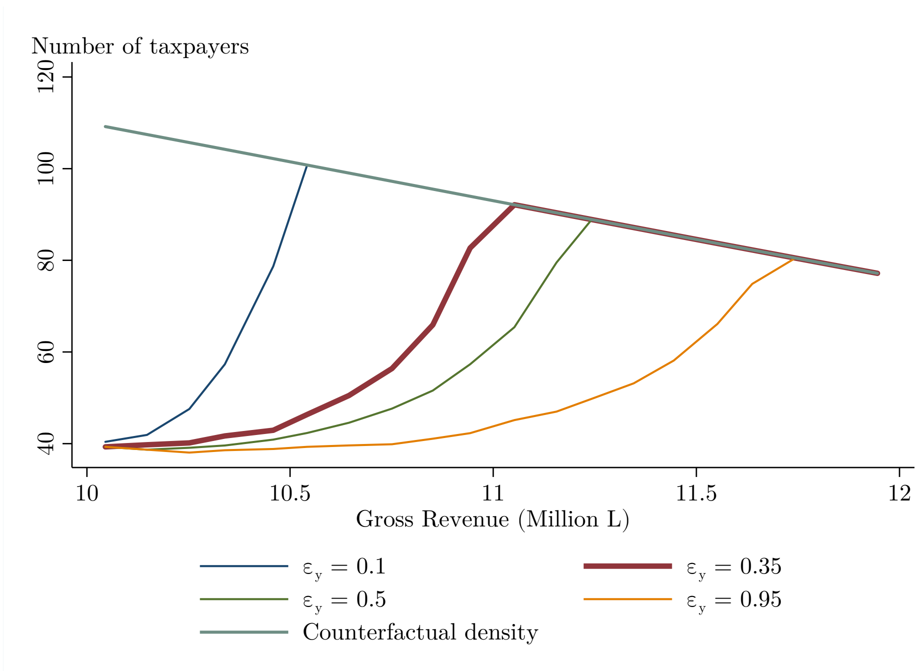
We implement the estimation of the revenue elasticity ϵ_y in the following steps. First, we need to consider the counterfactual profit distribution that would be observed in the absence of the notch. For each period in our sample, we take that to be the observed profit margin density for firms reporting revenue in the interval L6 - 8 million[29]. We then

[29]We show in A8 that the profit margin distribution is similar for the L6 - 8 million and L10-12 million range in the period before the introduction of the minimum tax.

proceed to compute, for each revenue bin (ΔY) and ϵ_y, what is the share of taxpayers with profit margin between 0 and the implied upper bound, and use the counterfactual density to obtain the number of taxpayers that bunch in each revenue bin. This allows us to obtain, for each potential revenue elasticity, the total number of predicted bunchers, which we compare to the estimated number of bunchers. The final elasticity, therefore, is the value that generates the same number of bunchers as the excess mass below the threshold.

We illustrate this procedure in A24 for the pooled sample of taxpayers in 2014-2017. Each of the curves is a simulated density that would prevail under a different revenue elasticity, according to our methodology.

Figure A24: Simulation to obtain average elasticity



Note: This figure presents the predicted density of gross revenues above the L10 million threshold and several simulations of what the density would have been given different revenue elasticities according to the model described above.

D Assessing dominated region with parametric model

As in Henrik J. Kleven and Waseem (2013b), we consider a parametric model to assess what is the dominated region in our notch setting, that is, the interval of revenue that is (potentially) strictly dominated for taxpayers to locate at. Consider a simple version of our iso-elastic cost model (with no possibility to overreport costs), where firms are defined by a productivity parameter θ and a fixed-cost parameter α and profits are given by

$$\hat{\Pi}(y,\alpha) = y - \alpha - \frac{\theta}{1+1/e}\Big(\frac{y}{\theta}\Big)^{(1+1/e)} - T(y,\alpha)$$

First, note that under a pure profit tax ($T(y,\alpha) = \tau_\pi(y - c(y))$), we have that $y^* = \theta$, so the revenue choice reveals the productivity parameter. Under revenue taxation, the optimal revenue choice is $y^* = \theta(1-\tau_y)^e$. Let the productivity of the marginal buncher be $\theta^T + \Delta\theta$. The marginal buncher is indifferent between reporting revenue exactly at the threshold or staying at their best interior solution. Their profit under each decision are given by

$$\Pi_{Bunch} = (1-\tau_\pi)\Big(y^T - \alpha - \frac{\theta^T+\Delta\theta}{1+1/e}\Big(\frac{y^T}{\theta^T+\Delta\theta}\Big)^{1+1/e}\Big)$$

$$\begin{aligned}
\Pi_{NotBunch} &= (1-\tau_y)y^* - \alpha - \frac{\theta^T+\Delta\theta}{1+1/e}\Big(\frac{y^*}{\theta^T+\Delta\theta}\Big)^{1+1/e} \\
&= (\theta^T+\Delta\theta)(1-\tau_y)^{1+e} - \alpha - \frac{\theta^T+\Delta\theta}{1+1/e}(1-\tau_y)^{1+e} \\
&= \frac{(\theta^T+\Delta\theta)(1-\tau_y)^{1+e}}{e+1} - \alpha
\end{aligned}$$

Finally, since the internal solution for the marginal buncher, had they not bunched, could be written as $y^T + \Delta Y = (\theta^T+\Delta\theta)(1-\tau_y)^e$, we can replace the terms involving the (unobserved) taxpayer type with the (observed) thereshold and the (estimable) change in revenue. We then have

$$\begin{aligned}
\Pi_{Bunch} &= \Pi_{NotBunch} \\
(1-\tau_\pi)\Big(y^T - \alpha - \frac{y^T+\Delta y}{(1-\tau_y)^e(1+1/e)}\Big(\frac{y^T(1-\tau_y)^e}{y^T+\Delta y}\Big)^{1+1/e}\Big) &= \frac{y^T+\Delta y}{(1-\tau_y)^e}\frac{(1-\tau_y)^{1+e}}{e+1} - \alpha \\
(1-\tau_\pi)(y^T-\alpha) - (1-\tau_\pi)(1-\tau_y)\frac{y^T+\Delta y}{1+1/e}\Big(\frac{y^T}{y^T+\Delta y}\Big)^{1+1/e} &= \frac{1-\tau_y}{e+1}(y^T+\Delta y) - \alpha
\end{aligned}$$

Let's consider what happens when taxpayers have $e = 0$. Taking the limit of the above equality as elasticity goes to zero we get:

$$\begin{aligned}
(1-\tau_\pi)(y^T-\alpha) - \frac{1-\tau_y}{1}(y^T+\Delta y) + \alpha = 0 \\
Lim_{e\to 0}\Delta y = \frac{\tau_y y^T - \tau_\pi(y^T-\alpha)}{1-\tau_y}
\end{aligned}$$

Some things to note. First, if $1 - \alpha/y^T = 0.06$, then $Lim_{e \to 0} \Delta y = 0$: for taxpayers with "profit margin" equal to 6% and zero elasticity, there exists no dominated region - the notch becomes a kink. For those with $y^T = \alpha$, so they report non-positive profits, $Lim_{e \to 0} \Delta y = \frac{\tau_y * y^T}{1-\tau_y} = L152,000$. These are the taxpayers with strongest incentive to bunch, and the region between L10 million and L10,152,000 is dominated. For those with taxable income rates between 0-6%, the dominated region lies between 0 and L152,000.

In our empirical estimation of elasticity we use bins of L100,000. According to the calculation above, no taxpayers with taxable income rate between 0 - 2% should locate in that region. Using the counterfactual taxable income rate distribution, this group represents approximately 30% of taxpayers, meaning that no more than 70% of taxpayers could be observed reporting revenue above the threshold. As can be seen in A24, for the first bin we observe less than 70 taxpayers while the counterfactual distribution predicts 110 taxpayers. So we cannot reject that, under 0 elasticity, all taxpayers that should bunch have actually bunched. Note that this is an extreme assumption, and we just cannot precisely explore the notch to recover "innatention" as in Henrik J. Kleven and Waseem (2013b) or Londoño-Vélez and Ávila-Mahecha (2019).

E Model calibration details

We modify firms' profit function by making explicit assumption about the cost and misreporting loss functions. Firms have isoelastic costs and also isoelastic loss function from misreporting costs:

$$\hat{\Pi}(y, \hat{c}) = (1-\tau)y + \tau\mu\hat{c} - \alpha_i - \frac{\theta_i}{1+1/e}\Big(\frac{y}{\theta_i}\Big)^{(1+1/e)} - \frac{B_i}{1+1/\gamma}\Big(\hat{c} - c(y)\Big)^{(1+1/\gamma)}$$

Each taxpayer is characterized by the vector $(\theta_i, \alpha_i, B_i)$ that define productivity, fixed cost and evasion ability, respectively. Given our functional forms, optimal vector of output and reported costs $(y^*, \hat{c}^*(y^*))$ are:

$$\begin{aligned} y^* =& \theta(1-\tau_E)^e \\ \hat{c}^*(y^*) =& c(y^*) + B_i\Big(\tau\mu\Big)^\gamma \end{aligned}$$

where $\tau_E = \tau\Big(\frac{1-\mu}{1-\tau\mu}\Big)$. Note that if we have profit taxation then $\mu = 1$ and $\tau_E = 0$, so firm size is undistorted.

In order to calibrate the model, we use data for the 2013, when no notches or kinks were in place. Under profit taxation, we have:

$$
\begin{aligned}
y^* =& \theta \\
c(y^*) =& \alpha + \frac{\theta}{1+1/e} \\
\hat{c}^*(y^*) =& \alpha + \frac{\theta}{1+1/e} + \left(\frac{\tau}{B_i}\right)^{\gamma}
\end{aligned}
$$

From the first-order conditions of an interior optimum, θ is simply the vector of reported output, which in this model coincides with real output. We also know the elasticity of output *e*, which we fix to be $e = 0.99$, the upper bound estimated for the pooled years. By using the upper bound of our elasticity estimate we are conservative in the case for using output taxation, since a higher elasticity will limit the potential benefit of the tax.

While we do not observe $c(y^*)$, the real costs, but only the reported costs $\hat{c}^*(y^*)$, we have estimated evasion as a share of profits using the 6% profit margin kink. Let that quantity be $\epsilon_{\hat{c}}$ .Using the fact that at the profit margin kink $(y-\hat{c})/y = \tau_y/\tau_\pi$ we can write:

$$
\frac{(\hat{c}-c)}{y} = \frac{(\hat{c}-c)}{(y-\hat{c})} * \frac{(y-\hat{c})}{y} = \epsilon_{\hat{c}}(\tau_y/\tau_\pi) = \epsilon_{\hat{c}} * 0.06
$$

Using the equations above, we have that

$$
\frac{(\hat{c}-c)}{y} = \frac{\left(\frac{\tau}{B_i}\right)^{\gamma}}{\theta} = 0.06\epsilon_{\hat{c}}
$$

In our setting, we do not have variation to identify γ, the elasticity of misreporting costs. Best, Brockmeyer, et al. (2015) explore different profit tax rates for different subset of firms, while Bachas and Soto (2021) use estimates of cost elasticity in two different thresholds. We calibrate our model using the estimate from Best, Brockmeyer, et al. (2015), which is approximately 1.5, which allows us to recover B_i as $B_i = \frac{\tau}{\left(\theta 0.06\epsilon_{\hat{c}}\right)^{1/\gamma}}$

Finally, given the previous we can just obtain the fixed cost vector α by computing

$$
\alpha = \hat{c}^* - \frac{\theta}{1+1/e} + \left(\frac{\tau}{B_i}\right)^{\gamma}
$$

F Social Contribution Tax and Net Asset Tax

Corporations face a 25% flat tax on yearly profits in Honduras. Three other provisions affect their potential tax liability. The first is the minimum tax studied in this paper, which was introduced in 2014 and started to phase out in 2018. Since 1994, corporations also face a net asset tax similar in nature to a minimum tax: if the tax liability under the asset tax is smaller than the profit tax liability, it can be used as a credit, meaning that in practice firms would only pay the profit tax. If the asset tax is larger, firms formally must pay the income tax and the additional difference between the two liabilities. In practice, the asset tax is also a tool to avoid that large corporations minimize their tax liability by inflating costs and driving down taxable income. In the period under study, the net asset tax was 1% of the net assets above L3 million.

The last provision is the Social Contribution (AS for the spanish *Aportación Solidaria*) tax, a surcharge on income tax applying to large firms. Established for the first time as a temporary measure in 2003, the AS tax rate varied between 5-10% in the period of this study and applied to declared taxable income above L1 million (USD 40,000)[30].

In A6 we present the distribution of firms by their tax status in each year of the sample. Both the AS and the asset tax existed throughout the analysis period, while the minimum tax was established in 2014. In each year, approximately one-quarter of tax filing corporations pay no income tax - this is often the result of generating no revenue in the period or, more frequently, registering losses (and not having enough assets to pay the Net Asset tax). Before the introduction of the minimum tax, around 63% of corporations were liable for income tax and 9% for the net asset tax. With the introduction of the minimum tax in 2014, the share of firms liable for asset tax does not change, but the share paying income tax falls by 8 percentage points as firms start being liable for the minimum tax. Between 1,400 and 1,700 firms were paying the minimum tax before 2018, when the number falls drastically to only 135 once the exemption threshold increases from L10 million to L300 million. The Social Contribution tax was payed by 8-10% of corporations every year, and it is a surcharge on those paying either income or minimum tax, but not the asset tax[31].

[30]A tax reform in 2010 established the AS tax rate at 10% for the first two years and then progressively declined to zero by 2015. With the 2014 tax reform, nonetheless, the tax was made permanent and the tax rate fixed at 5%.

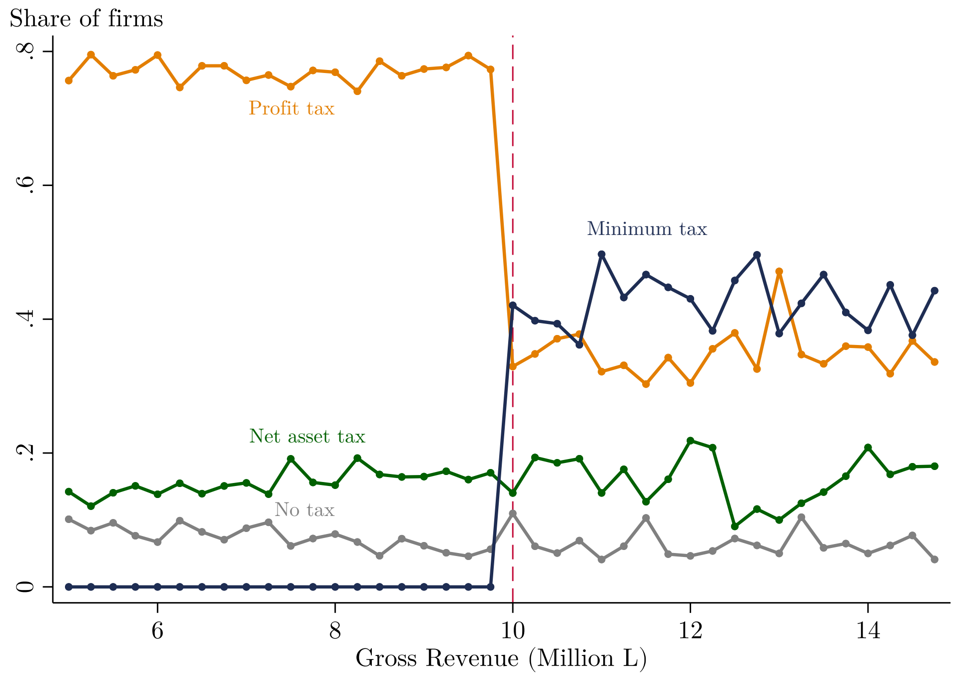
[31]In order to arrive at the final tax liability, the Tax Authority first calculates the maximum between the income tax and the minimum tax liabilities, and add the social contribution liability to that. This value is then compared to the asset tax liability, and the maximum of these two is the final tax liability.

Table A6: Taxpayer status by year

Year	Not taxed	Income Tax	Asset Tax	Minimum Tax	Total
2011	4,791	10,940	1,563	0	17,294
2012	4,763	11,548	1,798	0	18,109
2013	4,945	12,372	1,906	0	19,223
2014	5,397	11,566	1,891	1,610	20,464
2015	6,237	13,997	1,944	1,480	23,658
2016	6,641	15,553	2,057	1,478	25,729
2017	7,328	16,544	2,281	1,672	27,825
2018	7,946	19,080	2,783	135	29,944

Note: This table presents the distribution of corporate taxpayers each year, according to their tax liability status.

Figure A25: Share of firms liable for each type of tax (2014-2017)



Note: This figure presents the share of firms liable for each type of tax (profit, minimum, net asset or no tax), in each bin of gross revenue for the period 2014-2017 pooled. It shows that when crossing the L10 million exemption threshold the increase in the share of firms paying the minimum tax is mirrored by a decrease in the share of firms liable for profit tax, with little change observed in the share of firms paying the net asset tax or not paying any taxes. The sample excludes corporations exempt from the minimum tax due to sectoral exceptions and/or recent start of operations.

G Minimum taxes around the world

This section presents a summary of corporate minimum tax schemes across low and medium income countries. We highlight features that are common in several contexts. First, several countries exempt firms in the first 24-36 months of operations, a period where initial investment and set-up costs might legitimately generate low or negative profits Holland and Vann, 1998. Second, the tax rate applied to gross revenues often falls in the range of 0.5 - 2%, with reduced rates (or exemptions) applied to sectors such as pharmaceuticals, utilities and oil related industries. While this determines a floor for the effective tax rate (tax liability as share of gross revenues) corporations must pay, the implied minimum allowable profit margin (that is, the minimum profit margin reported such that firms are not paying the minimum tax rate) also depends on the corporate profit tax rate. In most countries the minimum allowable profit margin falls in the range of 1.5 - 5%, below the 6% level implied by the 1.5% gross revenue tax and 25% profit tax in place in Honduras in the period 2014-2017. Finally, in all but a few countries the minimum corporate tax provision apply to all firms, regardless of size.

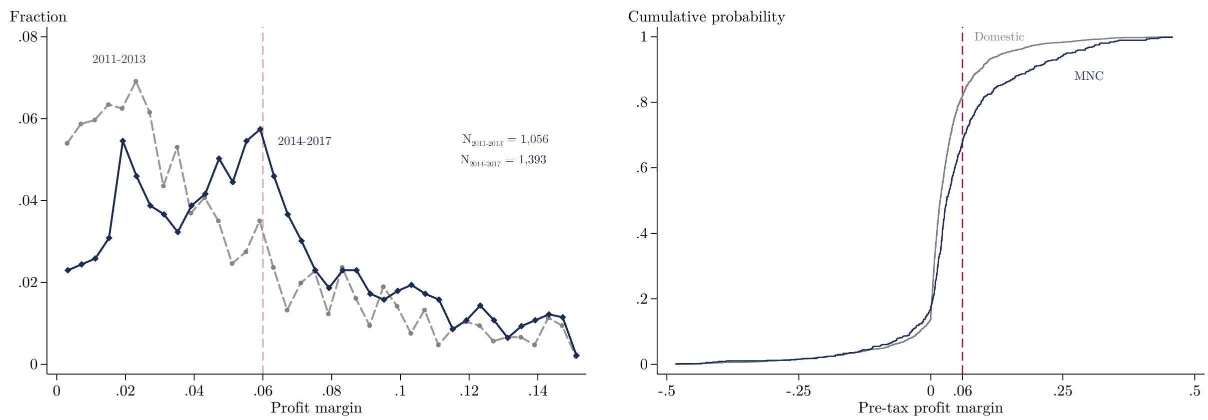
H How did multinational enterprises responded to the minimum tax?

Given the ongoing discussions on multinational taxation at the global level, we investigate whether the minimum tax in Honduras seem to have affected how multinational corporations (MNEs) use transfer pricing (i.e. transactions with related counterparts) costs to potentially reduce tax liability.

In A26, panel a, below we present preliminary evidence that, as observed in the sample of all corporations, MNEs also increased their reported profit margins when the minimum tax policy was introduced in 2014. The extent of adjustment, nonetheless, is much more muted than for the full sample. This is partly explained by the fact that, previous to the policy introduction, large MNEs were already declaring higher profit margins than their domestic counterparts (panel b).

Even though the overall adjustment by MNEs is smaller than by domestic firms, it is possible that the nature of this adjustment happens through the mechanism of transfer pricing. That is, it is possible that MNEs were more aggressively pricing transactions with foreign related parts before the introduction of the policy, in order to book profits on lower-tax jurisdictions, and changed their behavior in response to the introduction of the minimum tax policy.

Figure A26: Profit margin of Multinationals



(a) Empirical density of profit margins for MNEs

(b) Pre-tax profit margin CDF - Domestic vs. Multinational corporations

Note: Panel A in figure presents the empirical density of positive reported profit margins for multinational corporations (MNEs), before (2011-2013) and during (2014-2017) the existence of the minimum tax. It restricts the sample to MNEs declaring gross revenue above L13 million, significantly above the policy revenue exemption threshold. Bins are 0.4 percentage points wide. Panel B presents the cumulative distribution functions (CDF) of pre-tax profit margins by domestic and multinational firms in 2013, before the introduction of the minimum tax. The CDF of MNCs is shifted to the right (for positive values), indicating higher declared profit margin across the distribution. MNCs are defined as taxpayers that present transfer pricing declarations at some point in 2014-2018.

We obtained transaction level data on transfer pricing operations for all corporations operating in Honduras. Here we highlight some features of the data. First, corporations file transfer pricing declarations not only for transaction with foreign counterparts, but also with domestic partners that are under joint control. Almost 45% of total costs declared in TP declarations are with domestic partners (we define MNEs as firms with at least one TP transaction with a *foreign partner*). Corporations in the country file transactions with 94 other countries, with the majority of total volume concentrated in the United States (14%), Panama (13%) and Guatemala (5%) - all other countries combined make up 25% of claimed costs but with very fragmented shares. Among the top 15 trading partners, however, we observe countries widely recognized for offering ”low tax rates and favorable regulatory policies to foreign investors” Hines Jr., 2010: British Virgin Islands, Cayman Islands, Bahamas and Bermuda.

Multinationals are much larger than domestic firms both in terms of gross revenue and taxable income: over 80% of MNEs had revenue above the L10 million exemption threshold for minimum tax in place until 2017 and more than a quarter had revenue above L300 million, the new exemption threshold in 2018. We also show that in 2017 over 80% of

MNEs declared costs arising from a transaction with a related part[32] - 70% of them declared transactions with foreign partners and 45% with a domestic partner. Only a tiny share of domestic firms (2%) file a TP declaration informing of a transaction with a domestic related partner. Tax havens are also a popular source for foreign partners: 30% of MNEs declare at least one transaction with a related partner hosted in a tax haven (using the definition of Hines Jr. (2010)). We also show that, conditional on filing a TP declaration with a foreign partner, the (unweighted) average TP cost as a share of total costs is 32%, suggesting that costs arising from transactions with related parts are a meaningful share of the cost deductions used by MNEs. Finally, we should note that the number of MNEs filing income taxes every year is small (≈ 800).

One key limitation of the transfer pricing data for our exercise should be noted. That data is only available for the 2014-2018 period, meaning we cannot observe changes in behavior before and after the introduction of the minimum tax in 2014. As we document above, the sample of multinationals is rather small so in any case we cannot perform exercises relying on local variation around specific thresholds (e.g. there are only ≈ 120 firms with revenue between L8 - L12 million when pooling the entire 2014-2017 period).

For those reasons, we take a different approach to evaluate whether MNCs responded to the minimum tax policy. First, instead of considering the introduction of the policy in 2014, we will explore the variation generated by the phasing out of the policy in 2018. For that year, the revenue exemption threshold increased from L10 million to L300 million. Approximately 60% of MNEs (≈ 450 firms) declared revenue in that interval in 2017 and therefore were not exempt from the minimum tax that year but would be exempt in the following year if declaring the same revenue. Conversely, firms declaring revenue above L300 million in 2017 (≈ 200) were not affected by the increase in the exemption threshold in 2018, since they were still liable for the minimum tax. If, in the absence of the change in policy, the use of transfer pricing costs would have been similar among these two groups of firms, any differential behavior observed in 2018 could be attributed to the impact of the minimum tax.

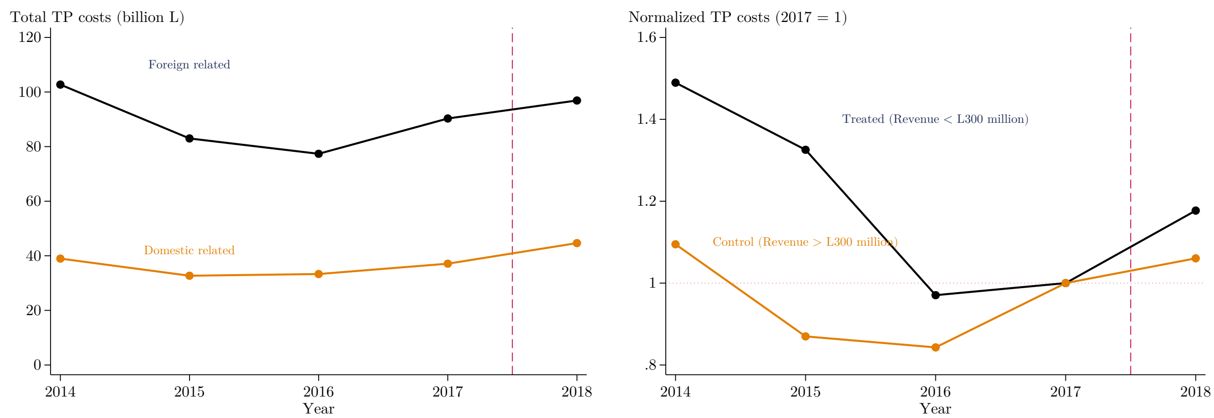
We start our analysis by presenting simple aggregate costs claimed through transfer pricing operations in each year (A27, panel a), separately for domestic and foreign counterparts. Costs claimed from transactions with foreign counterparts are 2 - 2.5 larger than those with domestic partners, but over time the pattern of aggregate costs is similar: they decreased in the period 2014 - 2016 then increased back to initial levels by 2018.

In panel (b) we present in graphical form the "differences-in-differences" approach we propose. We compare the amount of TP costs claimed from transactions with foreign parties for firms with revenue above L300 million in 2017 (186 firms) and those with revenue between L10 and L300 million (374 firms). We normalize the amount to one in 2017, so the graph presents the percentage change from that baseline year for each group. If affected MNEs reacted to the withdrawn of the minimum tax by significantly increasing

[32]Our definition of MNEs is that the taxpayer filed at least one TP declaration in the period 2014-2018, so in any given year some MNEs might not be filing any TP costs.

their costs (since now they would be taxed on profits), we should see a substantial increase in TP costs for that group in 2018, but not for those with revenue above L300 million. We do see that the total amount of TP costs claimed by the firms likely to be affected increased more (18%) when compared to those less likely to be affected (6%). However, the pre-trends of TP costs usage in these two groups are widely different. For those with revenue below L300 million, the total costs claimed were almost 50% higher in 2014 than 2017, and then increase again in 2018. For those with revenue above L300 million, costs were about 10% higher in 2014 when compared to 2017, then fell 10 - 20% below 2017 levels before recovering.

Figure A27: Transfer-pricing costs by multinationals



(a) Total TP costs - foreign vs domestic

(b) Normalized foreign TP costs - above vs. below L300 million

Note: This figure presents trends in the use of costs through transfer pricing operations by MNEs in the period 2014-2018. Panel (a) presents aggregate costs, separately for foreign and domestic partners, while panel (b) presents costs normalized to one in 2017, separately for firms with revenue above L300 million in 2017 and those below. In both figures we restrict the sample to a balanced panel of MNEs filing every year in the period and declaring revenue above L10 million.

The figure above is suggestive that any DiD approach will likely fail the pre-trends test. We formally estimate the following DiD regression for the same sample of MNEs:

$$log(1 + costs_{fy}) = \sum_{y=2014}^{2018} \beta_y \mathbb{1}\{treat = 1\} * \mathbb{1}\{year = y\} + \gamma_f + \delta_y + \epsilon_{yf} \tag{2.15}$$

where the outcome variable is $log(1 + costs)$ for firm f in year y; we include firm and year fixed effects; and our coefficients of interest are β_y, the differential use of TP costs by treated and control firms for every year.

We plot the resulting coefficients of interest in A29 below, where the outcome is the log of TP costs with foreign partners. The first feature of the results is that they are extremely noisy: in the pre-2017 period, we cannot reject that deviations of TP cost usage is similar among the two groups, but the 95% CI often cover the interval [-1,1.5] log-points, with point estimates of approximately [0.3, 0.5]. The point estimate for the difference in 2018 is much closer to zero, but still with very wide confidence intervals. That is likely partly driven by our sample size: we only have 450 multinationals in the balanced sample, so our sample size in the regression is 2,250 observations. The same pattern emerges when we estimate a similar regression using costs with domestic partners and costs with partners in tax havens as outcomes of interest: coefficients are very imprecisely estimated, and while all our estimates are not statistically different than zero we cannot rule out very large effects both before and after the 2017 phase-out of the minimum tax.

Our main takeway from these exercises is that while the effects of the minimum tax in Honduras on MNC are of much interest, we might not be able to precisely estimate them. That is in part because there are just not that many MNCs operating in the country, and also because our data has a limited time coverage and the phase out in 2018 only left out an even smaller number of very large MNCs unaffected by the change.

I Did the minimum tax lead to firm exit?

One key concern about minimum taxes specifically, and other distortive taxes in general, is that they might lead to firm exit. Some activities that might be worth pursuing when the tax base is profits – since the tax burden will be limited when profits are low or negative – become economically unfeasible if taxes are assessed on gross revenue. Here we provide more details on the exercises we perform to assess whether the introduction of the minimum tax in Honduras caused higher exit by affected firms.

We note the following. First, precisely because we show that firms manipulate their gross revenue in order to avoid the minimum tax threshold, we cannot use a regression discontinuity design to assess the policy impact, comparing firms just below and just above the exemption threshold. Second, both the behavioral response in terms of reported revenue and the fact that costs were overreported before the reform suggest that evasion responses might dampen any real economic responses.

The intuition behind our exercises is as follows. We determine groups of firms that were likely to be affected by the 2014 minimum tax based on **pre-reform characteristics**. Since the minimum tax only affected firms that would have declared gross revenue above L10 million and profit margins below 6% after its introductions, we use these thresholds to assign firms to the "treatment group": firms with revenue above L10 million and profit margin below 6% before the reform are more likely to be affected and potentially exit in response to the higher tax rate they will face.

We first define those groups by their characteristics in 2011, the first year in our panel dataset, and then follow firms until 2016 - we stop measuring firm exit before the end of our panel so we can assign firm exit only to those corporations that did not file in any subsequent period in the future. In A28, starting from the universe of filing firms, we construct four groups based on their revenue & profit margins in 2011 and follow their survival throughout the period. As we should expect, large firms (with revenue above L10 million in 2011) are more likely to survive over the entire period in comparison with smaller firms. Conditional on size, high-profit firms (declaring profits above 6% in 2011) are also more likely to survive than low-profit ones. But the figure does not suggest any **differential exit** by firms likely to be affected (high revenue & low profit) when compared to the other groups.

We implement a more formal testing of those differential exit rates in regression form. We consider a differences-in-differences setting, comparing the exit rate after the reform between firms with high- vs. low-revenue and those with high- vs. low-profitability. Formally, we estimate the following model using a cross-section of firms that file taxes in 2013:

$$Exit_{i,Y} = \alpha_i + \gamma_1 AboveL10_{i,before} + \gamma_2 Below6\%_{i,before} + \beta AboveL10_{i,before} * Below6\%_{i,before} + \epsilon_i \quad (2.16)$$

We are interested in the coefficient β, that presents the differential exit rate for firms likely to be affected by the reform: those with revenue above the L10 million threshold and profit margins below 6%. Since the reform was introduced in 2014, we present results for exit in different horizons: one, two and three-years after the reform. In our baseline specification[33], we define the groups by their declared revenue and profit margins in 2013, the year before the reform.

We present our results for the three exit horizons and considering two different samples. In columns (1)-(3), we use firms declaring gross revenue between L4 and L20 million in 2013, therefore restricting the sample to firms that were not too different in size but in both sizes of the minimum tax revenue threshold. In columns (1) and (2), the differential exit by 2014 and 2015 is very close to zero and not statistically different from zero. The estimate for differential exit by 2016, in column (3), is 2 percentage points - a larger effect in economic terms, considering the 10% general exit rate, but it is not statistically different from zero. Of course, our key results show that firms that would have declared revenue slighlty above L10 million after 2014 decide to bunch, so that is an important response margin that might mitigate any exit decisions.

For that reason, in columns (4) - (6) we lift the sample restriction and include all firms declaring income taxes in 2013. The sample increases five-fold (since most firms in 2013 declare revenue below L4 million), but now the comparison group includes firms vastly different in size. While the estimate for the first year is similar in size to the restricted sample, results for exit by 2015 (1.6 p.p.) and 2016 (3.6) are much larger in magnitude and statistically different than zero. They suggest that large firms with low profit margins were more likely to stop filing income taxes after the reform, which we use to proxy for firm exit.

[33]We control for economic sector of taxpayers in all regressions.

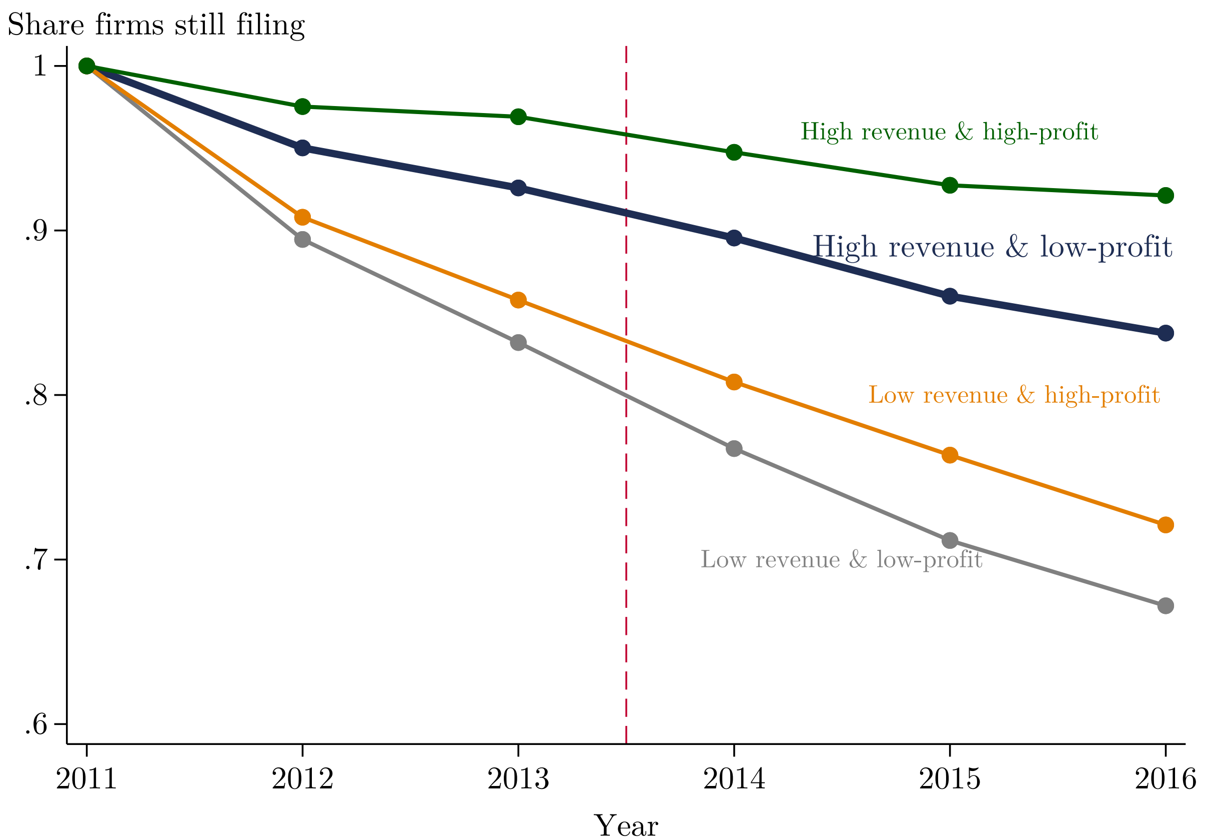
Since sample restrictions meaningfully affect the results, we provide a host of robustness tests in A30, where we plot the interaction coefficients for the regression considering exit by 2016, using different sample restrictions based on declared revenue in the base year. Here we show that restricting the sample to firms in a narrow band around the L10 million threshold in 2013 leads to small coefficients in magnitude but wide confidence intervals. As we expand the sample around the threshold, the coefficients increase from less than 1 p.p. to the range 2.5 - 3.5 p.p., with some estimates being significantly different from zero.

Under the assumption that we can attribute any changes in exit for high-revenue, low-profit-margin firms to the minimum tax, the previous result are suggestive that the reform might have increased firm exit by as much as 3.5 p.p. in the years following it.

Since our results are not quite robust across specifications and to stress our empirical specification, we also conduct a series of placebo tests. We implement the same specification used across samples in A30, but instead consider the base-year as 2011 and measure exit rates by 2013 - **before the introduction of the minimum tax**. We are not aware of any policies that might have affected the same group of firms, so our prior is that we should obtain null estimates. As we present in A31, nonetheless, for a range of samples we estimate negative coefficients that are economically and statistically significant: the group of firms with revenue above L10 million and low profit margins in 2011 were 3 - 5 p.p. less likely to exit by 2013.

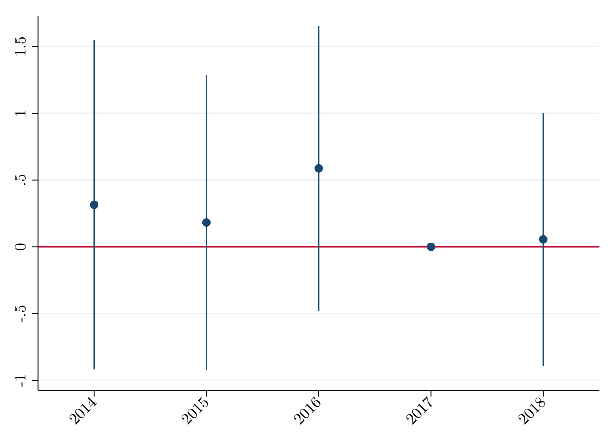
Given the sensitivity of our estimates to specification and the significant results estimated in the placebo regression, we avoid making claims about the impact of the introduction of a minimum tax on firm exit in our setting.

Figure A28: Firm survival using panel data



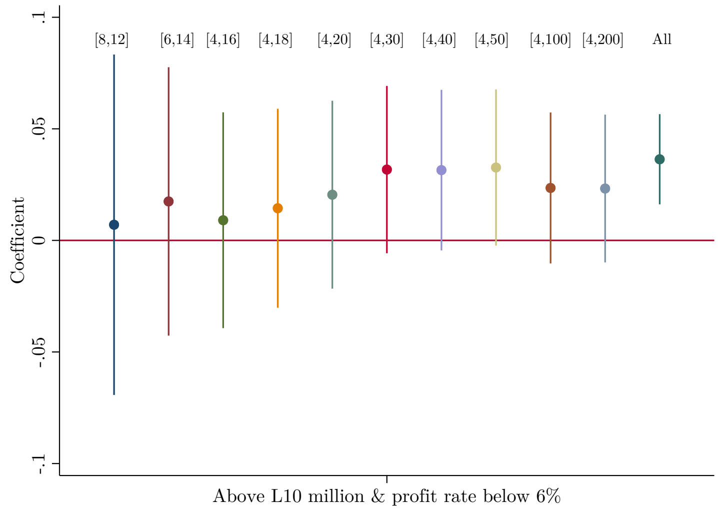
Note: These figures present the share of firms, in each year, that existed in 2011 and still file in each year. In both panels the sample is restricted to firms that presented a declaration in 2011. In panel A we restrict the sample to firms with gross revenue above L10 million in 2011 (and therefore likely to be affected by the minimum tax in the future) and present results separately for firms with low (below 6%) and high (above 6%) profit margins in 2011. In panel B, we restrict the sample to firms with low profit margins and present results for firms with low (below L10 million) and high (above L10 million) revenue in 2011.

Figure A29: Differences-in-differences estimates of TP costs



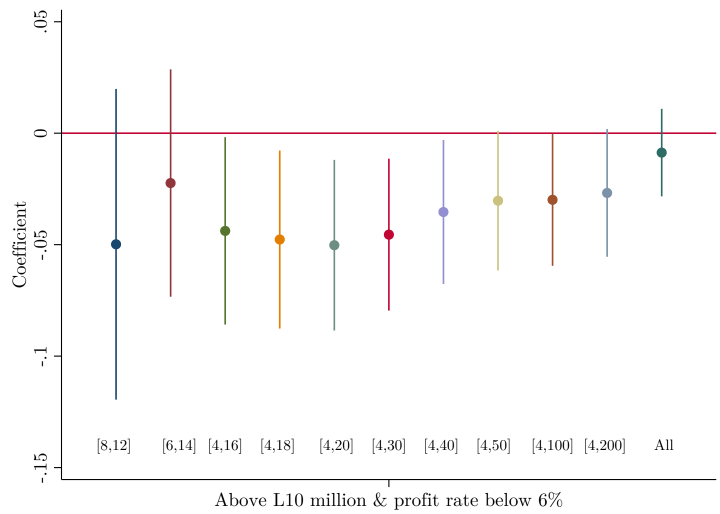
Note: This figure presents the point estimates and 95% CI of the coefficients of interest in 2.15. Standard errors were computed clustered at the taxpayer level.

Figure A30: Coefficients on exit (different revenue windows)



Note: This figure presents the coefficients on the interaction term for firms with revenue above L10 million and profit margin below 6% in 2013. The intervals indicated above each coefficient refer to the sample restriction related to declared gross revenue in 2013, the year before the introduction of the Minimum Tax. The first coefficient, for example, is estimated in a regression restricting the sample to firms with gross revenue between L8 - 12 million in 2013, while the last coefficient refers to a regression using all firms in 2013, regardless of revenue.

Figure A31: Coefficients on exit (different revenue windows) - Placebo test



Note: This figure presents placebo tests, where the coefficients on the interaction term for firms with revenue above L10 million and profit margin below 6% in 2011. The intervals indicated above each coefficient refer to the sample restriction related to declared gross revenue in 2011, while the dependent variable is exit by 2013, the year before the introduction of the minimum tax.

Chapter 3

Reducing Interference Bias in Online Marketplace Experiments using Cluster Randomization: Evidence from a Pricing Meta-Experiment on Airbnb

A Introduction

Many of the world's most highly valued and/or fastest growing technology firms (e.g., Airbnb, Uber, Etsy) are online peer-to-peer marketplaces. These platforms create markets for many different types of goods, including transportation, accommodations, artisanal goods, and even dog walking. Like almost all technology firms, online peer-to-peer marketplaces typically rely on experimentation, or A/B testing, to measure the impact of proposed changes to the platform and develop a deeper understanding of their customers. However, a randomized experiment's ability to produce an unbiased estimate of the total average treatment effect (TATE) relies on the stable unit treatment value assumption (SUTVA) (Rubin, 1974), one component of which is the "no interference" assumption (Cox, 1958). This assumption states that in any given experiment, each unit's outcome is a function only of their own treatment assignment, not the treatment assignments of others.

Bias in TATE estimates due to interference, which we refer to in this paper as "interference bias", is likely to occur in online marketplace settings because the buyers and sellers in marketplaces are inherently connected; different goods for sale in a marketplace are likely to complement or substitute for one another, and sellers are likely to make strategic decisions based on the actions of their competitors. Previous work (Blake and Coey, 2014; Fradkin, 2015) suggests that naive experimentation in online marketplace settings can lead to TATE estimates that are overstated by up to 100%, and as a result, a quickly emerging body of academic research (M. Liu, Mao, and K. Kang, 2021; Johari et al., 2022;

Bojinov, Simchi-Levi, and J. Zhao, 2022; Bright, Delarue, and Lobel, 2022; H. Li et al., 2022) focuses on how to properly account for interference bias specifically in the context of online marketplaces.[1] Both researchers and academics consider this an important problem to solve because decision-making based on experiment designs and analyses that fail to account for interference bias can have a non-trivial and negative financial impact for online marketplace firms.[2] However, there is still limited empirical work providing insight into the actual severity of interference bias, particularly in seller-side experiments.

Interference bias as a general phenomenon is not unique to online marketplaces, and has been well-studied in the research literature on unipartite social networks; in such settings, interference arises due to interactions between individuals, often referred to as peer effects (Manski, 2000; Moffitt et al., 2001). For instance, the observed behavior of one's peers can affect voting behavior (Bond et al., 2012), exercise habits (Aral and Nicolaides, 2017), and mobility levels (D. Holtz et al., 2020). One tool for reducing interference bias in social network experiments is graph cluster randomization (GCR) (Ugander et al., 2013; Eckles, Karrer, and Ugander, 2017), an experiment design technique in which the relevant network is clustered and treatment assignment is then randomized at the cluster-level, as opposed to the individual-level. While GCR is an established method in the network experimentation literature, it is unclear *ex ante* if cluster randomization will be an effective tool to reduce interference bias in online marketplaces. This is largely due to factors arising from the bipartite nature of online marketplaces: the mechanisms driving interference may be different than those in a social network setting,[3] and the appropriate mathematical model of interference in marketplaces may deviate from the one used to model the "self-reinforcing" spillovers seen in many unipartite network settings (i.e., positive (negative) direct effects lead to positive (negative) spillover effects).

In this paper, we use a randomized meta-experiment on Airbnb[4] to simultaneously 1) provide empirical evidence of interference bias in an online marketplace seller-side pricing experiment and 2) propose and assess the viability of utilizing cluster randomization to reduce interference bias in such settings. We test for interference bias in a pricing experiment in particular because pricing experiments are of special interest to online marketplace intermediaries; experiments related to prices help firms better understand the price elasticity of their customers, which consequently enables them to implement optimal pricing-related marketplace mechanisms such as fee structures and seller pricing suggestions. Understanding customer price elasticities can also be beneficial to sellers, who set their own prices. Results from our meta-experiment indicate that cluster ran-

[1]A working version of our paper predates and is cited by much of this research.

[2]In Appendix I, we use a simple economic model to explore the potential financial ramifications of misestimating price elasticities for an online marketplace intermediary.

[3]As a result, if an experiment designer were to try and create a "network" of sellers and perform GCR, it is not immediately obvious how edges between sellers should be defined.

[4]Airbnb is an online marketplace for accomodations and experiences. More than six million listings appear on Airbnb, and since the company's founding in 2008, over one billion guest arrivals have occurred on the platform (Airbnb, 2019).

domization *is* a viable tool for reducing interference bias in seller-side marketplace experiments, and that interference bias would have accounted for at least 19.76% of the "naive" TATE estimate produced by an individual-level randomized evaluation of the treatment intervention we study.

We begin by using a pre-existing linear model of interference to explore how online marketplace interference differs from social network interference, and the implications this has for experiment design. Interference in this model is captured by a matrix $\boldsymbol{B}$, which we refer to as the "interference matrix." In order to construct an appropriate interference matrix for online marketplace settings, it is necessary to understand the mechanism(s) that drive interference. One possibility is that interference in online marketplaces operates via the same mechanism as social network interference, i.e., it is driven by sellers observing each others' actions and/or interacting. To assess whether this is plausible, we use proprietary data from Airbnb to measure the frequency with which Airbnb hosts search in their own geographies and view the product detail pages (PDP) of other listings. We find that over the course of a month, only 13.3% of listing hosts searched for specific dates in their own geographies and only 21.3% of hosts had at least one PDP view in their own geography. These results suggest that it is unlikely social influence is a significant contributor to interference in online marketplaces. In contrast, a simple simulation of online marketplace dynamics that does not include any seller behavior (see Appendix J) produces results consistent with the existence of interference, suggesting that competitive dynamics are likely a contributor to marketplace interference. In other words, the amount of interference between listings is at least in part determined by the extent to which they co-occur within the consideration sets of shoppers.

Another difference between social network interference and online marketplace interference is that in most social network settings, positive (negative) direct effects beget positive (negative) spillover effects, whereas we expect positive (negative) direct effects to create negative (positive) spillover effects in online marketplaces. We extend a result from Eckles, Karrer, and Ugander (2017) and show that in the presence of both same-signed and opposite-signed spillovers, cluster randomization will always reduce the bias of the difference-in-means TATE estimator. In doing so, we derive a closed form expression for the expected amount of interference bias remaining under a given clustering; this expression is a function of interference matrix $\boldsymbol{B}$, and can be used to evaluate the "quality" of a given set of clusters.

Building on these insights, we present results from an in vivo meta-experiment, or "experiment over randomized experiments" (Saveski et al., 2017) conducted on Airbnb. The treatment intervention we study in this meta-experiment is a change to Airbnb's platform fee structure; more specifically, hosts in the treatment group were charged *higher* platform fees than hosts in the control group. The meta-experiment design randomly assigned clusters of Airbnb listings to one of two randomization schemes; 25% of clusters were randomized at the individual-level (i.e., treatment is randomly assigned to listings at the indvidual level), whereas the remaining 75% of clusters were cluster randomized (i.e., treatment is randomly assigned to listings at the cluster level). Using this design, we obtain separate TATE estimates in the individual-level and cluster randomized treatment arms, and then test for a statistically significant difference between the two. Results from the individual randomized meta-treatment arm (i.e., the “naive” experiment design) suggest that the treatment led to a statistically significant loss of .345 bookings per listing over the course of the experiment. However, when we compare this TATE estimate to the estimate produced by the cluster-randomized meta-treatment arm, we find that 19.76% of the individual-level TATE estimate is eliminated by cluster randomization and attributable to interference bias. We also find suggestive, non-statistically significant evidence that interference bias is more severe in demand-constrained geographies, and that the bias reduction from cluster randomization is larger in geographies with “higher quality” clustering.

Situating our work within the broader literature focused on interference bias in online marketplace experiments, we provide an estimate of the potential severity of interference bias in such settings, and evaluate the efficacy of cluster randomization at reducing said bias. We believe there is not a one-size-fits-all solution to interference bias in marketplace experiments, and that each proposed solution (including ours) has its strengths and weaknesses. Cluster randomization works well in marketplaces without centralized matching (in contrast to Bright, Delarue, and Lobel (2022)), for treatment interventions that must be randomized at the seller-level (in contrast to Johari et al. (2022)), and in marketplaces that are susceptible to intertemporal spillovers (in contrast to Bojinov, Simchi-Levi, and J. Zhao (2022)). Nonetheless, cluster randomization brings with it substantial reductions in statistical power, and many of our theoretical results apply only to treatment interventions that uniformly increase or decrease demand, but not a mixture of both. We consider both of these weaknesses promising avenues for future research.

B Related Literature

The research in this paper connects to three bodies of academic literature: one on interference bias in online marketplace experiments, one on interference in networks, and one on pricing-related interventions in online marketplaces.

Interference bias in online marketplace experiments

Our work is most closely related to an emerging body of research focused on the phenomenon of interference-related estimation bias in TATE estimates when conducting experiments in online marketplace settings. This issue was first identified by Blake and Coey (2014) and shortly thereafter by Fradkin (2015), who both report that naive marketplace experimentation can yield TATE estimates that are overstated by up to 100%. In the intervening years, a number of experiment design-based solutions to this problem have been proposed (M. Liu, Mao, and K. Kang, 2021; Bojinov, Simchi-Levi, and J. Zhao, 2022;

Johari et al., 2022; H. Li et al., 2022) including "two-sided randomization" (Johari et al., 2022) and "switchback" experimentation (Bojinov, Simchi-Levi, and J. Zhao, 2022).[5,6]

While each proposed solution to marketplace interference has appealing attributes, none of them offers a "silver bullet" solution. For instance, under two-sided randomization, both buyers and sellers are randomly assigned at the individual-level to treatment or control, and the treatment intervention is only delivered to buyer-seller pairs in which both the seller *and* the buyer have been assigned to the treatment. Two-sided randomization is especially well-suited to corporate experimentation settings, where existing experimentation tooling is often built specifically with individual-randomization in mind. Johari et al. (2022) show that this design reduces bias in TATE estimates due to interference without much loss of precision. However, not all treatment interventions can be delivered at the buyer-seller dyad level, e.g., a new tool for setting prices can only be delivered at the seller-level, and a new search algorithm can only be delivered at the buyer-level. In a switchback experiment design (Bojinov, Simchi-Levi, and J. Zhao, 2022), time is discretized and the experiment designer randomizes the treatment assignment that is delivered to the entire marketplace at each time step. While switchback experiments have appealing statistical properties, they can produce an inconsistent user experience for marketplace participants, and are difficult to implement when markets do not clear quickly, creating "carryover" or temporal spillover effects. This is the case in marketplaces such as Airbnb, where guests often visit the site multiple times over the course of days or weeks before making a booking.

Interference in networks

The aforementioned papers focus on solving the problem of interference bias in online marketplace experiments, which is uniquely difficult because of the bipartite nature of marketplaces. However, the problem of estimation bias in TATE estimates arising from SUTVA violations is well-studied in settings that are not bipartite. Researchers focused on this topic have developed statistical tests for the existence of interference (Rosenbaum, 2007; P. M. Aronow, 2012; Bowers, Fredrickson, and Panagopoulos, 2013; Athey, Eckles, and Imbens, 2018), techniques for conducting valid causal inference in the presence of interference (M. G. Hudgens and Halloran, 2008; Tchetgen and VanderWeele, 2012; P. M. Aronow and Samii, 2017; Sävje, P. Aronow, and M. Hudgens, 2021; Chin, 2018), and experiment designs that account for interference (Sinclair, McConnell, and Green, 2012; Imai, Tingley, and Yamamoto, 2013; Ugander et al., 2013; Lan Liu and M. G. Hudgens, 2014; Eckles, Karrer, and Ugander, 2017; Saveski et al., 2017; Baird et al., 2018; Basse and Feller, 2018; Ariel, Sutherland, and Sherman, 2019).

[5]Cluster randomization was first proposed as a solution to interference bias in online marketplaces in D. M. Holtz (2018), an unpublished master's thesis. The main results from D. M. Holtz (2018) now appear in Appendix J of this work.

[6]Analysis-based solutions to the problem have also been suggested, e.g., in Bright, Delarue, and Lobel (2022).

Our work is most closely related to that of Ugander et al. (2013), Eckles, Karrer, and Ugander (2017), and Saveski et al. (2017), which all focus on experiment designs that deliver cluster-randomized treatment to networks with the aim of obtaining less-biased TATE estimates. Ugander et al. (2013) propose graph cluster randomization (GCR), an experiment design in which, after clustering a network, treatment assignment is randomized at the cluster-level. The authors show that under certain conditions, GCR eliminates interference bias and produces unbiased TATE estimates. Eckles, Karrer, and Ugander (2017) build on this work by showing through simulation that in instances where the conditions outlined in Ugander et al. (2013) do not hold, GCR can still greatly reduce interference bias, although it does not eliminate it entirely.[7] Saveski et al. (2017) conduct a "meta-experiment" on LinkedIn that compares the TATE estimate obtained under individual-level randomization to that obtained under GCR. This paper makes two contributions to the literature: providing a method to test for interference bias in network settings, and reporting results that highlight the efficacy of GCR at reducing said bias.

In their totality, these papers provide a thorough exploration of GCR as a method for reducing interference bias in network settings. However, because of the bipartite nature of marketplaces, differences in the mechanisms driving interference, and differences in the appropriate way to mathematically model said interference, it is unclear ex ante if cluster randomization will be as effective in the marketplace setting. Thus, in this work we propose cluster randomization as a method to reduce interference bias in marketplace experiments, and test its efficacy using a Saveski-style meta-experiment.

Pricing-related interventions in online marketplaces

Finally, our research connects to the literature on pricing-related interventions in online marketplaces. It is important for both platform intermediaries and platform sellers to understand the price elasticity of their customers; sellers would like to price effectively, whereas intermediaries would like to implement effective fee structures (Choi and Mela, 2019) and pricing-related marketplace mechanisms. For instance, in recent years a growing number of online marketplaces have launched machine-learning based pricing interventions (Ifrach et al., 2016; Dubé and Misra, 2017; Filippas, Jagabathula, and Sundararajan, 2019; Ye et al., 2018). Many pricing interventions are tested and launched using randomized experiments, however, if the TATE estimates produced by these experiments are biased, marketplace designers may mis-estimate price elasticities and/or launch sub-optimal policies. For instance, in Appendix I, we use a simple economic model to show that setting platform fees based on biased elasticity estimates reduces firm profits. These losses have the potential to wipe out the positive impacts typically associated with A/B

[7]One drawback of assigning treatment at the cluster-level is that most treatment effect estimators will have less statistical power than under an individual-level randomized design. However, techniques such as regression adjustment (Gerber and Green, 2012) and pre- and post-stratification (Moore, 2012; Miratrix, Sekhon, and Yu, 2013) can be used in tandem with cluster randomization to mitigate the loss of statistical power.

testing (Feit and Berman, 2019; Azevedo et al., 2020). Our work confirms that interference *can* bias TATE estimates when conducting pricing-related experiments in online marketplaces and establishes that cluster randomization can be an effective tool to reduce this bias.

C Interference Bias in Online Marketplaces

Before presenting the results of our meta-experiment, we first explore the ways in which interference bias in marketplaces differs from interference bias in social networks, and the implications this has for experiment design. The basis for this exploration is the following linear parametric model of interference, which is studied in, e.g., Eckles, Karrer, and Ugander (2017) and Pouget-Abadie et al. (2018):

$$Y_i(\boldsymbol{Z}) = \alpha_i + \beta Z_i + \gamma \rho_i + \epsilon_i \tag{3.1}$$

where Y_i is the outcome of seller i, $\boldsymbol{Z}$ is the treatment assignment vector, β is the "direct" effect of the treatment, γ is the "indirect" effect of the treatment, ρ_i is the percentage of seller i's competitors/neighbors that are treated, and $\epsilon_i \sim N(0,1)$ is independent of ρ_i. The same linear outcome model can be represented in the following way:

$$E[Y_i(\boldsymbol{Z})] = \alpha_i + \sum_{j \in V} B_{ij} Z_j, \tag{3.2}$$

where Z_j indicates the treatment assignment of seller j, and $\boldsymbol{B}$ is an "interference matrix" capturing the strength of the interference between seller i and seller j.

Does "Seller Influence" Drive Interference?

The notation above makes it clear that in order to reduce interference bias through experiment design, it is helpful to have some idea how to construct an appropriate interference matrix, $\boldsymbol{B}$. In other words, it is helpful to understand the *mechanisms* that drive interference. Here, we investigate whether interference in online marketplaces operates via a similar mechanism to interference in social networks, i.e., it is driven by sellers observing the behavior of other sellers and changing their behavior in response. To do so, we reference the search and product detail page (PDP) view activity of Airbnb listing hosts in this paper's meta-experiment in the month prior to the meta-experiment's launch (February 16, 2019 to March 15, 2019). We find that the overwhelming majority of Airbnb hosts do not search in their own geographies or view the PDPs of competitors, suggesting that the "seller influence" mechanism is unlikely to play a major role in driving spillovers in our context. More specifically, in the month preceding our meta-experiment, only 22.7% of listing hosts searched at least once in their own geography, and only 13.3% searched at least once for specific dates in their own geography. Among hosts who ran at least one search in their own geography, the median host searched only 8 times. Furthermore, only 21.3% of hosts had at least one PDP view to a within-geography listing that wasn't their own. Among hosts that had at least one PDP view to a within-geography listing that wasn't their own, the median host carried out 4 PDP views across 3 distinct listings. More detailed data on search and PDP view activity in the month preceding our meta-experiment is shown in Figure A1. Given these results, in conjunction with the fact that 1) like our meta-experiment, many experiments run for much shorter periods of time than 30 days and 2) treatment interventions like the one we study in our meta-experiment are often subtle and unlikely to be noticed by hosts after just a few search sessions or PDP views, we consider it likely that interference in online marketplaces is driven not by "seller influence," but instead by the fact that sellers co-occur in the consideration sets of potential buyers and compete with each other for transactions.[8]

Modeling Interference in Online Marketplaces

Another point of contrast between interference in online marketplaces and interference in many social network settings is the nature of the interference between units. Many network experiments study treatment interventions with "self-reinforcing" spillovers, i.e., treatment interventions in which positive (negative) treatment interventions have positive (negative) spillovers (put differently, β and γ in Equation 3.1 have the same sign). For instance, a vaccination encouragement intervention might increase vaccination rates not only among those that are treated, but also among their peers. Similarly, in a social media setting we would typically expect an intervention that increases the posting activity of treated users to also increase the posting activity of treated users' peers.

In contrast, many potential marketplace treatment interventions act on seller outcomes in such a way that β and γ have opposite signs, since sellers and buyers compete with one another. For instance, if an intervention caused treated Airbnb hosts to raise (lower) their prices, this could lead to an decrease (increase) in demand for their listings, and, consequently, a increase (decrease) in demand for their competitors' listings.[9] This is exactly the pattern we observe in the fee meta-experiment results presented in Section E. While the TATE of increasing platform fees is negative (we estimate a TATE of -0.277 bookings per listing in the cluster-randomized meta-treatment arm), the bias we observe points in the opposite direction (we estimate a TATE of -0.345 bookings per listing in the individual-randomized meta-treatment arm). We hypothesize that this is because Airbnb customers are more likely to see a mixture of treatment and control listings under

[8] The notion that spillovers in online marketplaces are driven by competitive dynamics is consistent with the simulation results found in Appendix J.

[9] It is also possible that Airbnb hosts in a given geography could serve as complements to each other. For instance, guests may describe their positive (negative) experience with a given listing to their peers, which could increase (decrease) demand for similar listings. However, we consider it much more likely that accommodations on Airbnb are substitutes, and assume this to be the case throughout the rest of this work.

individual-level randomization, and customers who see such a mixture may shift their business from higher fee listings to lower fee listings.

Eckles, Karrer, and Ugander (2017) show that when β and γ have the same sign, i.e., when spillovers are "self-reinforcing," cluster randomization will always reduce the bias of the TATE estimator relative to individual-level randomization. However, they stop short of proving that this is true in cases where the direct and indirect treatment effects point in opposite directions, as is likely to be the case in online marketplace settings. We introduce the following proposition, which extends Theorem 2.1 from Eckles, Karrer, and Ugander (2017) and shows that cluster randomization is guaranteed to reduce the bias of TATE estimates, even in cases where the direct and indirect effects of a treatment intervention (captured by the interference matrix) have opposite signs.

Proposition 1. *Assume we have a linear outcome model for all sellers $i \in S$ that is a function of the form*

$$E[Y_i(\boldsymbol{Z})] = \alpha_i + \sum_{j \in V} B_{ij} Z_j, \tag{3.3}$$

where Z_j indicates the treatment assignment of seller j, and $\boldsymbol{B}$ is a matrix in which all of the diagonal entries have the same sign and all of the off-diagonal entries have the same sign. Then for any mapping of sellers to clusters $C(\cdot)$, the absolute bias of the difference-in-means TATE estimate under cluster randomization, $\hat{\tau}_{cr}$, is less than or equal to the absolute bias of the difference-in-means TATE estimate under individual-level randomization, $\hat{\tau}_{ind}$, with a fixed treatment probability p.

Proof. Proof. Given in Appendix K.

Proposition 1 establishes that cluster randomization will never increase TATE estimation bias, but does not provide any guidance on how to construct clusters. In any given marketplace setting, there will exist many different ways to cluster sellers. For instance, an experiment designer might cluster sellers based on seller-level attributes, observed rates of seller co-occurrence in search, or estimated cross-price elasticities, to name a few possibilities. However, not all clusterings will be equally effective at reducing TATE estimation bias. For instance, if a given approach to clustering produces clusters that are essentially random, bias reduction will be very close to 0, whereas if a given clustering does a very good job of capturing the relevant marketplace dynamics, bias reduction has the potential to be much larger. Given this fact, it is natural for an experiment designer to want to identify the clustering that will lead to the greatest reduction in estimation bias.

Unfortunately, there isn't a singular optimal method for clustering; the most effective clustering strategy will vary depending on the specific research context and the treatment intervention being studied. Considering this, it is necessary to develop a concept of 'cluster quality' that is adaptable to different contexts and takes into account the relevant interference matrix, $\boldsymbol{B}$, for a specific experiment. Thankfully, our proof of Proposition 1 provides a valuable resource. The left-hand side of the final inequality in this proof helps us quantify the bias of the difference-in-means TATE estimator within a given clustering.

This bias quantification can be used as an indicator of the quality for a defined set of clusters, represented as $C(\cdot)$.

Definition 1. *The quality of a given set of clusters, $Q_C(\boldsymbol{B})$, is defined as*

$$Q_C(\boldsymbol{B}) = \left| \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij} 1 \left(C(i) \neq C(j) \right) \right|. \tag{3.4}$$

Although in theory, Definition 1 provides a context-dependent measure of cluster quality, in practice, the relevant interference matrix $\boldsymbol{B}$ for a given research setting and treatment intervention is almost never observable to experiment designers. However, as long as the experiment designer is able to construct some proxy matrix $\boldsymbol{P}$ that is a monotonic transformation of $\boldsymbol{B}$, it follows directly from Proposition 1 that $Q_C(\boldsymbol{P})$ can still be used to determine which of two sets of clusters, C_1 and C_2, produces more biased difference-in-means TATE estimates.[10]

Proposition 2. *Suppose that $\boldsymbol{P}$ is a monotonic transformation of $\boldsymbol{B}$. Then,*

$$Q_{C_1}(\boldsymbol{P}) \leq Q_{C_2}(\boldsymbol{P}) \implies Q_{C_1}(\boldsymbol{B}) \leq Q_{C_2}(\boldsymbol{B}). \tag{3.5}$$

These results suggest that 1) for seller-side marketplace interventions that uniformly increase or decrease demand for treated sellers, cluster randomization should always reduce interference bias, regardless of cluster quality (although bias reductions will increase with cluster quality) and 2) after identifying a set of clusters, $C(\cdot)$, an experiment designer can assess their quality by calculating $Q_C(\boldsymbol{P})$.[11] In Section E, we investigate how cluster quality moderates the extent to which cluster randomization reduces interference bias in our meta-experiment. The measure of cluster quality used in this analysis is calculated using a proxy matrix $\boldsymbol{P}$ based on listing co-occurrence in searcher-level PDP view sessions. The intuition behind this choice is that in order for two Airbnb listings to compete with one another for bookings, they need to co-occur in searchers' consideration sets. In Appendix N, we provide more detail on how we calculated this particular $Q_C(\boldsymbol{P})$ using browsing data from Airbnb.

D Platform Fee Meta-Experiment

Although the theoretical results in the previous section suggest that cluster randomization should reduce interference bias in seller-side marketplace experiments, it is unclear

[10] Note that because $\boldsymbol{B}$ is typically not observable, the statement that a given proxy matrix $\boldsymbol{P}$ is a monotonic transformation of $\boldsymbol{B}$ will almost always rely on a set of modeling assumptions that are not empirically testable.

[11] Alternatively, Pouget-Abadie et al. (2018) propose a meta-experiment design that can be used to empirically compare the efficacy of different sets of clusters at reducing TATE bias.

if this is true in practice. Furthermore, even if interference bias in seller-side marketplace experiments is a theoretical concern, it may not be a practical one if the severity of interference bias is small. If the magnitude of interference bias is small and/or cluster randomization is not an effective bias reduction technique, cluster randomization may not be worth implementing; cluster randomization is more logistically complicated and many industry experimentation tools do not easily support cluster randomization.

In this section, we describe the design of an in-vivo meta-experiment conducted on Airbnb's platform in March 2019.[12] By analyzing this meta-experiment, we obtain an empirical lower bound on the severity of interference bias in a "naive" individual-level randomized pricing experiment on Airbnb, and also measure the extent to which cluster randomization reduces that bias.[13]

Treatment Intervention

The treatment intervention we study in our meta-experiment was a change to Airbnb's platform fees for guests. Airbnb's fees for guests were visible in three different locations throughout the booking process. First, guest platform fees were included in the total price shown to guests when a listing appeared in search (top panel of Figure A2). Second, if a guest opened the "price breakdown" tooltip on any search result, they were shown a price breakdown that separated out the nightly price and the guest platform fee (bottom panel of Figure A2). Finally, when viewing a listing's PDP, a detailed pricing breakdown (including fees) was displayed next to the "Request to Book" button (Figure A3).

Our meta-experiment targeted long-tenured listings (i.e., listings that had been listed on Airbnb as of a certain cutoff date). Listings in the treatment had their guest fees *increased* relative to the status quo, whereas listings in the control had their fees *decreased* relative to the status quo. Less-tenured listings (i.e., listings created after the cutoff date) did not have their fees changed relative to the status quo.[14,15]

[12]We roughly follow the meta-experiment design introduced by Saveski et al. (2017). Pouget-Abadie et al. (2018) propose a similar "experiment over experiments" design. Meta-experiment designs such as these can be thought of as special cases of the randomized saturation designs discussed in, e.g., Baird et al. (2018).

[13]This meta-experiment was motivated by the simulation-based work found in Appendix J. While simulation-based work is helpful for conducting preliminary analysis, we believe that our meta-experiment provides value above and beyond simulation based work, since any simulation-based study of interference in marketplaces (including ours) will rely on assumptions about consumer behavior, the nature of the interference between units, etc.

[14]Due to our NDA with Airbnb, we are unable to disclose the exact magnitude of the fee changes in this experiment, nor are we able to disclose the cutoff date used to determine whether listings were long-tenured. Furthermore, all of our outcome variables (bookings, nights booked, gross guest spend) are multiplied by a random constant.

[15]Because our meta-experiment only impacts fees for long-tenured listings, we restrict our analysis dataset to long-tenured listings. However, the clusters used in our experiment include all listings, regardless of tenure on the platform.

Experiment Design

Our meta-experiment design is extremely similar to the "experiment over experiments" design described in Saveski et al. (2017). First, Airbnb listings were sorted into clusters using the process described in Section D. Clusters were then randomly assigned to one of two meta-treatment arms: individual-level randomization (25% of clusters), or cluster randomization (75% of clusters). Within the individual-level randomized meta-treatment arm, treatment was randomly assigned to listings at the individual level. Within the cluster-randomized meta-treatment arm, treatment was randomly assigned to listings at the cluster level. The entire meta-experiment design is summarized in Figure A4.

Each meta-treatment arm can be analyzed as a standalone experiment that produces a TATE estimate, and then, by jointly analyzing the data from both meta-treatment arms, we are able to measure whether there is a statistically significant difference between these two estimates. In order to increase statistical power for this comparison, we arranged our clusters into strata and use post-stratification (Miratrix, Sekhon, and Yu, 2013) when analyzing our data. The process we used to generate those strata is described in Section D.

Generating Hierarchical Listing Clusters

The first step in the design of our meta-experiment was arranging listings into clusters. There are many different ways to sort listings into clusters (e.g., the simulation described in Appendix J takes a graph clustering approach to generating clusters: edges were drawn between listings that share observable traits, and the resulting graph was clustered using Louvain clustering (Blondel et al., 2008)). For our in-vivo meta-experiment, we took an approach to clustering that made use of technical infrastructure that already existed at Airbnb. The first step in the process of generating these clusters was generating a dense, 16-dimensional demand embedding for each listing. Listings were then arranged into hierarchical clusters based on their location in that 16-dimensional space. Finally, a maximum cluster size was chosen in order to determine which subset of the hierarchical clusters to use in our meta-experiment.[16]

We generated demand embeddings for each Airbnb listing using a process similar to the one described in Grbovic and Cheng (2018). The training data used to generate our demand embeddings consisted of sequences of listings that individual users viewed in the same search session. If, for instance, a user viewed listings L_A, L_B, and L_C in one search session, this would generate the sequence:

[16]We believe that providing guidance on cluster construction is beyond the scope of this paper, given that the "optimal" set of clusters for cluster randomization will vary depending on the research setting and the treatment intervention of interest. However, the cluster quality metric provided in Definition 1 can be a useful tool for adjudicating between two candidate sets of clusters. We also believe that the analyses and theoretical results in this paper provide a roadmap of sorts that other researchers can draw on when designing clusters for the purpose of a cluster-randomized marketplace experiment. We discuss this point further in Section F.

$$< L_A, L_B, L_C > . \tag{3.6}$$

We used a word2vec-like architecture (Mikolov, Sutskever, et al., 2013) to estimate a skip-gram model (Mikolov, Chen, et al., 2013) on this data. Given S sequences of listings, the skip-gram model attempts to maximize the objective function

$$J = \max_{W,V} \sum_{s \in S} \frac{1}{|s|} \sum_{i=1}^{|s|} \sum_{-k \leq j \leq k,\, k \neq 0} \log p\left(L_{i+j} | L_i\right), \tag{3.7}$$

where k is the size of a fixed moving window over the listings in a session, W and V are weight matrices in the word2vec architecture, and $p(L_{i+j}|L_i)$ is the hierarchical Softmax approximation to the regular softmax expression. The objective function above was augmented by including listing-level attributes (e.g., a listing's geography) in the search session sequences. The model was then trained using a geography-level negative sampling approach.

Once listing embeddings were generated using the aforementioned approach, a recursive partitioning tree (J. H. Kang, Park, and S. B. Kim, 2016) was used to arrange the Airbnb listings into hierarchical clusters. The algorithm starts from a single cluster containing all listings, and then recursively bisects clusters into two sub-clusters. The algorithm stops bisecting sub-clusters when the tree reaches a depth of 20, or when a new sub-cluster will contain less than 20 listings. Listings can then be assigned to clusters of arbitrary maximum size by applying a cut to the hierarchy of clusters generated by the recursive partitioning tree. Figure A5 depicts example clusters generated using this method in the San Francisco Bay Area. Using an ad-hoc approach, we chose a cluster size threshold of 1,000 for the fee meta-experiment. This ad-hoc approach is described in Appendix L.

Treatment assignment randomization

Once each Airbnb listing was assigned to a cluster, 75% of clusters were randomly assigned to the "meta-treatment" (cluster randomization) and 25% of clusters were randomly assigned to the "meta-control" (individual-level randomization). Within the meta-control arm, Bernoulli individual-level randomization was used to assign 50% of listings to the treatment and 50% of listings to the control. Within the meta-treatment arm, Bernoulli cluster randomization was used to assign 50% of clusters to the treatment and 50% of clusters to the control. Each listing in a meta-treatment cluster was assigned the treatment assignment corresponding to its cluster.

Strata for post-stratification

In our meta-experiment analysis, we use post-stratification (Miratrix, Sekhon, and Yu, 2013) to increase statistical power. The strata we use for this purpose were generated using a multivariate blocking procedure (Moore, 2012). As a first step, we collected pre-treatment listing-level data for the period running from January 16, 2019 to February 17, 2019. Across this period, we calculated cluster-level summary statistics: the average number of nights booked per listing, the average number of bookings per listing, the average gross guest spend per listing, and the number of non-experimental holdout listings in the cluster.[17] After centering and scaling each of these metrics, we calculated the Mahalanobis distance (Mahalanobis, 1936) between each pair of clusters. Finally, we used an optimal-greedy algorithm to arrange clusters into strata of maximum size $n = 8$.

Experiment Preliminaries

The meta-experiment was run from March 16, 2019 to March 21, 2019 on a sample of 2,602,782 listings.[18] Of those listings, 647,377 were assigned to the listing-randomized meta-control arm, and the remaining 1,955,405 were assigned to the cluster-randomized meta-treatment arm. Within the listing-randomized meta-treatment arm, 323,734 listings were assigned to the control and 323,643 listings were assigned to the treatment. Within the cluster-randomized meta-treatment arm, 2,981 clusters were assigned to the treatment and 2,979 clusters were assigned to the control, resulting in 979,015 listings assigned to the treatment and 976,390 listings assigned to the control. In total, across both meta-treatment arms, 1,300,124 listings were assigned to the control and 1,302,568 listings were assigned to the treatment. We check for balance on pre-treatment outcome variables between the meta-treatment and meta-control clusters, and between the control and treatment groups in both meta-treatment arms, we do not detect any statistically significant differences, indiczating our randomization procedure was sound.

E Results

In this section, we present results from the fee meta-experiment. We focus on a single outcome metric, bookings, but the results for two alternative outcome metrics, nights booked and gross guest spend, are qualitatively similar and can be found in Appendix M. Since relative to the control, the treatment *increased* fees, we expect the TATE on bookings to be negative.

We first present the results from separately analyzing the individual-level randomized and cluster randomized arms of the meta-experiment. While the individual-level randomized arm will have ample statistical power, we expect its TATE estimate to suffer

[17] At the time of our meta-experiment, experiments on Airbnb excluded listings in a long-term experiment holdout group, as well as listing in Airbnb's "Plus" tier.

[18] Shortly after the meta-experiment's conclusion, a "reversal experiment" was run from April 15, 2019 to April 22, 2019. In the reversal experiment, listings that had been assigned the treatment condition in the meta-experiment were assigned the control, and vice-versa. The purpose of the reversal experiment was to mitigate any potential negative impact of the meta-experiment on Airbnb hosts.

from interference bias. On the other hand, analysis of the cluster randomized arm should provide a less biased estimate of the TATE, since the amount of marketplace interference will be reduced, but will also have less statistical power. Simply comparing the point estimates obtained independently from the two meta-treatment arms is not sufficient to rigorously measure interference bias. In order to do so, we proceed to jointly analyze both the individual-level randomized and cluster randomized meta-treatment arms. Finally, we investigate the extent to which our results vary as a function of 1) the level of supply- or demand-constrainedness in an Airbnb marketplace and 2) the geography-level quality of our clusters.

Individual-level & Cluster Randomized Results

We analyze both the individual-level randomized and cluster randomized meta-treatment arms separately by estimating the following model on listing-level data,

$$Y_i = \alpha + \beta T_i + \sum_l \gamma_l 1(B_i = l) + \delta X_i + \epsilon_i \tag{3.8}$$

where Y_i is the number of bookings, T_i is the treatment assignment for listing i, B_i is a variable indicating which stratum listing i's cluster of belongs to, X_i is a vector consisting of listing i's pre-treatment bookings, nights booked, gross guest spend, calendar nights available, and geography-level number of searches per available night in the month prior to the meta-experiment, and ϵ_i is an error term. For the cluster-randomized meta-treatment arm, we cluster standard errors at the Airbnb listing cluster-level.[19]

Table 2 shows the TATE estimate for bookings in both the individual-level randomized (column 1) and cluster randomized (column 2) meta-treatment arms. In the individual-level randomized meta-treatment arm, the TATE is -0.345 bookings per listing, whereas in the cluster randomized meta-treatment arm, the TATE is -0.277 bookings per listing. Both of these TATE estimates are statistically significant at the 95% confidence level.

Joint Analysis

In order to determine whether the difference between the TATE estimates generated by the two meta-treatment arms is statistically significant, we estimate the model

$$Y_i = \alpha + (\beta + \nu M_i) T_i + \xi M_i + \sum_l \gamma_l 1(B_i = l) + \delta X_i + \epsilon_i, \tag{3.9}$$

[19] In order to increase statistical power, our preferred model specification is Equation 3.8, which utilizes post-stratification (Miratrix, Sekhon, and Yu, 2013) through the inclusion of stratum-level indicators. Results obtained from estimating a more straightforward model that regresses bookings only on treatment assignment can be found in Table P.6.

where Y_i is the outcome of interest, M_i is a binary variable set to 1 when listing i is in the individual-level meta-treatment arm and 0 when i is in the cluster-randomized meta-treatment arm, T_i is a binary variable set to 1 when listing i is exposed to the treatment, B_i is a variable indicating the stratum of clusters to which listing i belongs, X_i is a vector consisting of listing i's pre-treatment variables, and ϵ_i is the error term. Standard errors are clustered at the individual-level for listings in the individual-level randomized meta-treatment arm, and at the Airbnb listing cluster-level for listings in the cluster-randomized meta-treatment arm.[20]

In the above model, β measures the "true" effect of the treatment, and ν measures the difference between the estimated effect of the treatment in the individual-level randomized arm and the estimated effect of the treatment in the cluster randomized arm. In other words, ν should measure the extent to which cluster randomization reduces interference bias, and also provide a lower bound on the amount of interference bias in the individual-level randomized meta-treatment arm.[21] ξ measures any baseline difference between listings in the individual-level randomized arm of the meta-experiment and listings in the cluster-randomized arm of the meta-experiment; since clusters were randomly assigned to meta-treatment arms, we expect ξ to be zero. Once we have estimated Equation 3.9, our estimate of the interference bias is

$$\Omega = \frac{\hat{\nu}}{\hat{\nu} + \hat{\beta}}, \tag{3.10}$$

i.e., the percentage of the listing-randomized meta-treatment arm TATE estimate that does *not* appear in the cluster-randomized meta-treatment arm TATE estimate. We calculate standard errors on this quantity using the delta method (we use the `deltamethod` function in the R library `msm`).

Column 1 of Table 3 and Figure A6 show the results from estimating Equation 3.9 on our entire sample. We estimate that the "true" TATE is -0.277 bookings per listing, whereas -0.068 bookings per listing of the TATE measured in the listing-randomized meta-treatment arm is due to interference bias. Plugging these point estimates into Equation 3.10, we estimate that 19.76% ($\pm 9.06\%$) of the TATE estimate achieved through the individual-level randomized experiment is due to interference bias, and was eliminated through cluster randomization.

[20] In order to increase statistical power, our preferred model specification is Equation 3.9, which utilizes post-stratification (Miratrix, Sekhon, and Yu, 2013) through the inclusion of stratum-level indicators. Results obtained from estimating a more straightforward model that regresses bookings only on meta-treatment assignment, treatment assignment.

[21] Recall that even when using cluster randomization, TATE estimates will likely remain biased to some extent, since any given clustering will do an imperfect job of capturing every pair of listings that interfere with one another.

The Moderating Effect of Supply and Demand Constrainedness

We hypothesize that the extent to which the TATE estimate under listing-level randomization suffers from interference bias will depend on marketplace conditions. More specifically, we expect that interference bias will be *larger* in geographies that are demand constrained, and *smaller* in geographies that are supply constrained. The intuition for this is as follows: in an extremely supply-constrained geography, all listings will eventually get booked, which will push the interference bias to zero, whereas in an extremely demand-constrained geography, only "more appealing" listings (i.e., only those in the treatment or control, depending on the treatment intervention) will be booked, maximizing interference bias. Simulation-based evidence motivating this hypothesis can also be found in Johari et al. (2022).

To test this hypothesis, we re-estimate Equation 3.9 separately for listings that are above/below the median listing in terms of the supply-constrainedness of their geography. Our measure of "supply constrainedness" is relatively crude, but effective: we divide the number of searches occurring in a given geography in the month prior to our meta-experiment by the number of calendar nights available in the geography at the outset of the month prior to our experiment. Columns 2 and 3 of Table 3 display our results for supply-constrained and demand-constrained geographies, respectively; these results are also visualized in Figure A7. We estimate that 12.05% ($\pm 11.55\%$) of the listing-level randomized TATE estimate in supply-constrained geographies can be attributed to interference bias, whereas 28.65% ($\pm 14.91\%$) of the listing-level randomized TATE estimate in demand-constrained geographies can be attributed to interference bias. While these results are consistent with both our hypothesis and the results reported in Johari et al. (2022), the difference between these two point estimates is not statistically significant (see Column 1 of Table P.8), and hence these results should only be considered suggestive.

The Moderating Effect of Cluster Quality

We also hypothesize that geographies with higher quality clusters (as defined in Definition 1) should see a greater reduction in interference bias. Using a process described in Appendix N, we construct a geography-level measure of cluster quality. Under this measure, which uses a proxy for the "true" interference matrix $\boldsymbol{B}$ based on user-level PDP view sessions, a given clustering is considered "higher quality" if listings tend to co-occur with listings from the same cluster in user-level PDP view sessions. We proceed to split listings into those that are above or below the median listing in terms of geography-level clustering quality, and separately estimate Equation 3.9 on these two samples. Columns 4 and 5 of Table 3 display our results for low-quality and high-quality clustering, respectively; these results are also visualized in Figure A7. We find that clustering reduces the TATE estimate by 25.92% ($\pm 15.14\%$) in geographies with high-quality clusters, and reduces the TATE estimate by 14.98% ($\pm 11.69\%$) in geographies with low-quality clusters. As was the case for our heterogeneity analysis with respect to supply-constrainedness, although these results are consistent with our hypothesis, we consider them suggestive since the difference between these two estimates of interference bias reduction is not statistically significant (see Column 2 of Table P.8).[22]

F Discussion

In this paper, we have highlighted the ways in which interference bias in online marketplaces differs from interference bias in social networks, and presented results from an in vivo meta-experiment conducted on Airbnb. Results from this meta-experiment provide empirical evidence that interference has the potential to cause substantial statistical bias in online marketplace seller-side experiment TATE estimates, and establish that cluster randomization is a promising tool for reducing said bias. More specifically, we find that at least 19.76% of the TATE estimate obtained from our individual-randomized meta-treatment arm was due to interference bias. We also find suggestive, non-statistically significant evidence that interference bias is more severe in demand-constrained geographies, and that higher-quality clusters lead to greater bias reduction in TATE estimates.

While our results show that there *can* be a sizable amount of interference bias in online marketplace experiments, it is possible that different treatment interventions in different marketplaces would be less (or more) prone to estimation bias. Although we are unable to make evidence-based claims on this topic, we believe that the analyses described in this paper provide something of a roadmap for researchers and firms hoping to assess the potential severity of interference bias in their setting and/or use cluster randomization to mitigate it. For instance, researchers might begin by estimating the potential financial impact of interference bias in their setting (Appendix I), conducting observational analysis to better understand the potential mechanisms driving interference in their setting (Section C) and/or running simulated experiments (Appendix J).

When interference bias seems worth accounting for, an appropriate next step would be to weigh the pros and cons of cluster randomization relative to other proposed solutions such as two-sided randomization (Johari et al., 2022) and switchback experimentation (Bojinov, Simchi-Levi, and J. Zhao, 2022). In general, both two-sided randomization and switchback experimentation will reduce TATE estimation bias relative to the individual-level randomized baseline. The extent to which this bias reduction comes at the price of reduced statistical power depends on the amount of supply-demand imbalance (in the case of two-sided randomization) or the strength of temporal "carryover"

[22] We conduct the same analysis with an alternate definition of cluster quality that is based on observable listing attributes, as opposed to consumer search data. To construct this alternative measure, we classify two listings as "substitutable" if they are in the same geography-level decile for the following three variables: share of 5 stars trips, person capacity, and price. At the geography-level, we then calculate the average percentage of a listing's "substitutable" listings (including itself) that are in the same cluster. Table P.9 shows our results using this alternative cluster quality measure; they are qualitatively similar to those found in Table 3.

effects (in the case of switchback experimentation). There are also some treatment interventions for which switchback experimentation and/or two-sided randomization may not be viable (for instance, data-driven decision-making aids cannot be assigned at the buyer-seller dyad level, as is required for two-sided randomization). Beyond relying on domain knowledge and intuition, managers and researchers may find it informative to run simulated experiments that make reasonable assumptions about, e.g., the strength of carryover effects or the types of sellers that might interfere with one another, and compare the bias and statistical power of different experiment designs and treatment effect estimators in these simulations. As previously mentioned, relative to alternatives, our belief is that cluster randomization is well-suited to seller-side interventions that are susceptible to intertemporal spillovers.

In cases that are best suited to cluster randomization, researchers can consider many different sets of clusters and either calculate and compare the "quality" of said clusters (Appendix N) or conduct a meta-experiment using the design described in Pouget-Abadie et al. (2018) to identify which clustering will provide the greatest bias reduction. Having chosen a set of clusters, one can imagine either running a straightforward cluster randomized experiment to obtain a TATE estimate, or conducting a meta-experiment similar to ours (Section D) to obtain a lower-bound on the actual amount of interference bias present.

We believe our work leaves open multiple promising avenues for future research, the most pressing of them being the development of methods to increase the statistical power of cluster-randomized experiments in online marketplaces. Even in cases where cluster randomization is well-suited to the treatment intervention under evaluation, one major barrier to the adoption of cluster randomization in online marketplaces is the fact that clustering greatly reduces the precision of TATE estimates. Loss of statistical power due to clustering can also make it difficult to estimate the severity of interference bias. This is evidenced by the fact that the confidence interval around our interference bias estimate is still quite wide, despite our meta-experiment including over 2 million Airbnb listings.[23] Future work might focus on, e.g., using meta-experiments to estimate underlying structural parameters of marketplaces (such as price elasticities), and subsequently using those structural parameter estimates to optimize the design of future experiments and/or predict the amount of interference bias associated with other potential treatment interventions.

Furthermore, the results we present in Section C are somewhat specific to treatment interventions that lead to uniform increases/decreases in demand. However, many treatment interventions of interest, including algorithmic pricing interventions (Ifrach et al., 2016; Dubé and Misra, 2017; Filippas, Jagabathula, and Sundararajan, 2019; Ye et al., 2018) increase demand for some sellers while decreasing demand for others. Future research

[23]To further emphasize this point, let us provide an explanatory anecdote: prior to the meta-experiment reported in this paper, we conducted a different pricing-related meta-experiment on Airbnb with a milder treatment intervention. Because the treatment intervention was milder, this meta-experiment was underpowered to detect interference bias, despite having a sample size in the millions.

might explore theoretical guarantees around cluster randomization in marketplaces when treatment interventions are more complicated than those considered in this paper and/or conduct meta-experiments similar to ours to assess the efficacy of cluster randomization when the treatment intervention under evaluation is more complex.

G Figures

Figure A1: Panel A shows the distribution of product detail page (PDP) views to within-geography listings, whereas Panel B shows the distribution of unique within-geography listings with at least one PDP view. Panel C shows the distribution of within-geography searches, whereas Panel D shows the distribution of within-geography searches with dates. Searches with dates are generally considered to be higher intent to book.

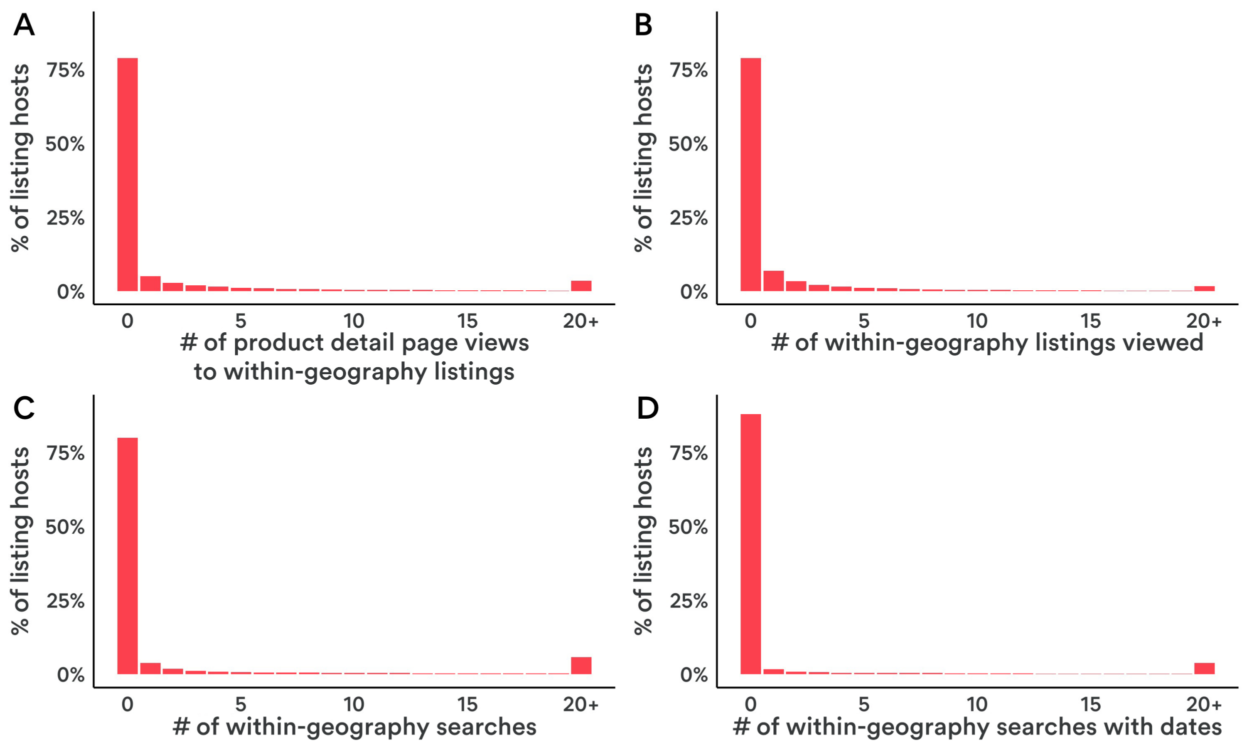


Figure A2: The top panel shows a typical search result on Airbnb at the time of the experiment. In this case, the guest platform fee was included in the total price of $508. The bottom panel shows what was displayed to guests after clicking the "price breakdown" tooltip: the guest platform fee (listed here as a service fee of $58) was broken out from the total nightly price.

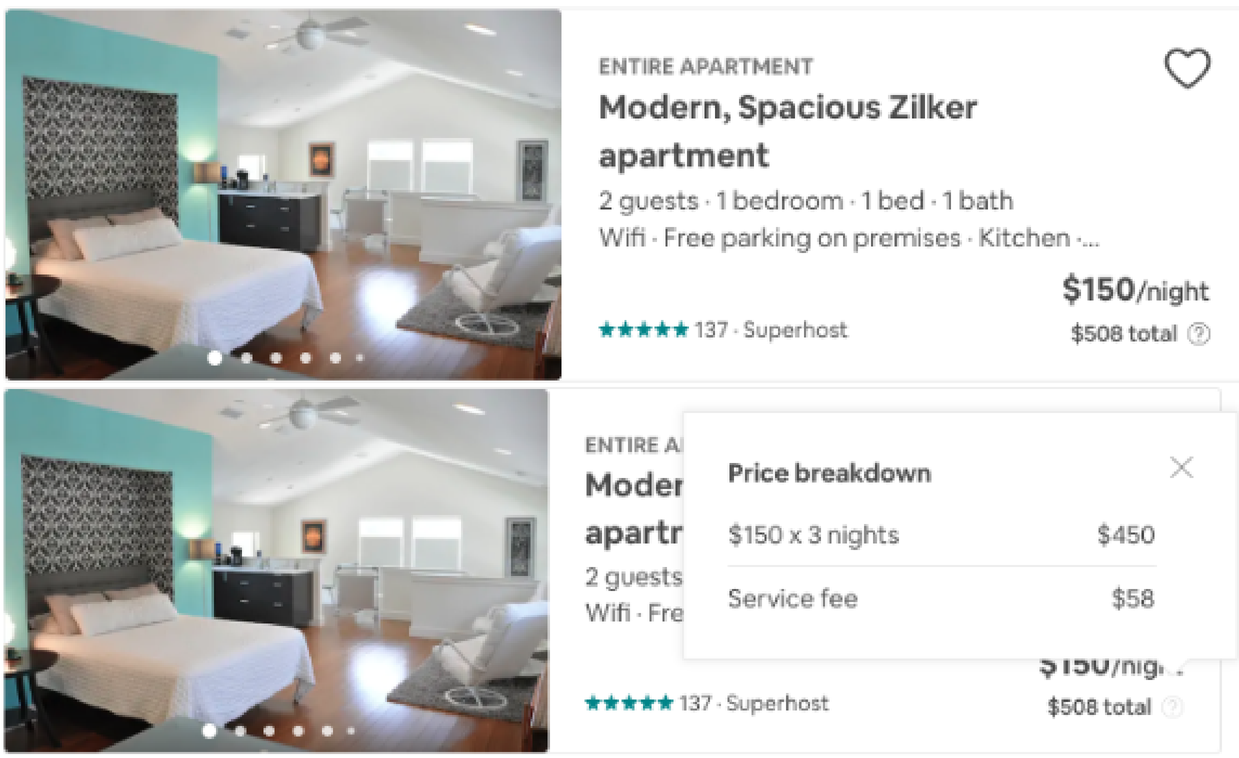


Figure A3: The section of the Airbnb product detail page that provided a full pricing breakdown for would-be guests. In this pricing breakdown, the guest platform fee (listed here as a service fee) is $58.

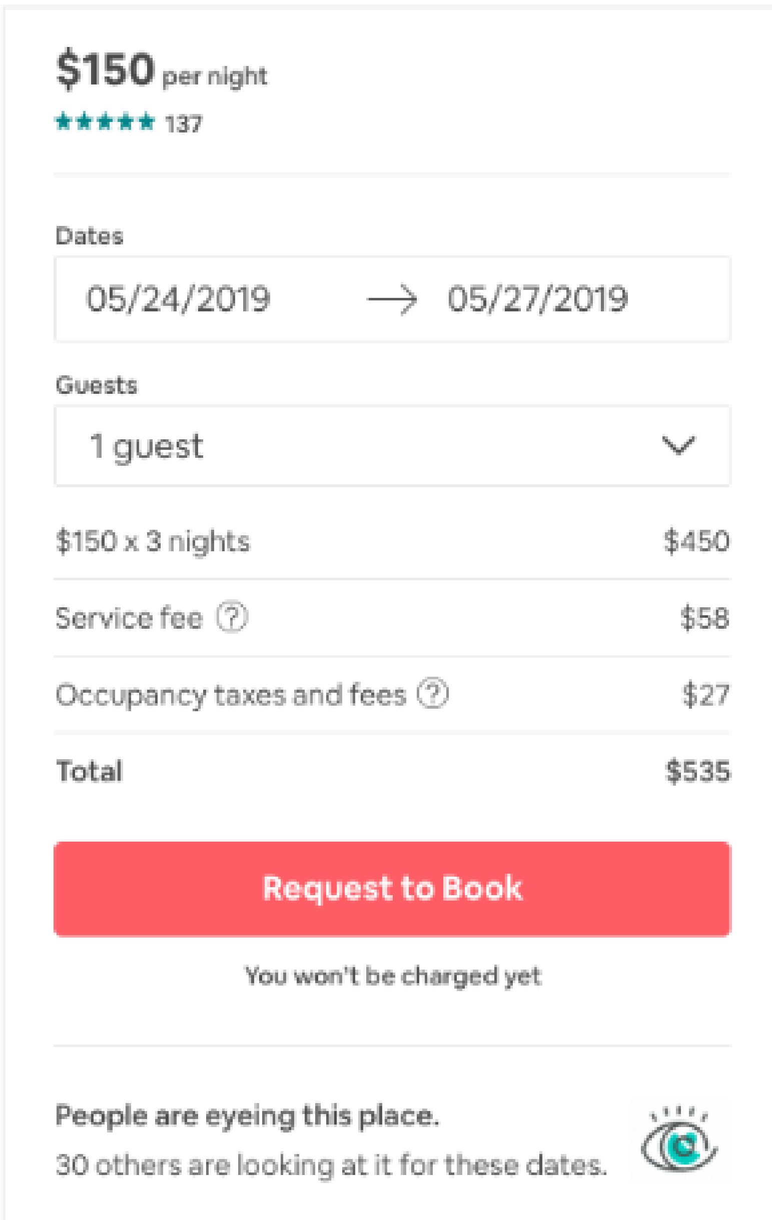


Figure A4: This figure depicts the experiment design process. We use listing-level co-occurrence in search (a) in order to learn "demand embeddings" (b). A hierarchical clustering algorithm is then applied to those embeddings in order to generate clusters (c). Clusters are randomly assigned to meta-treatment or meta-control (d); within meta-control, treatment is assigned at the individual-listing level, whereas in meta-treatment, treatment is assigned at the cluster-level (e). We arrange clusters into strata after treatment assignment to facilitate post-stratification (Miratrix, Sekhon, and Yu, 2013).

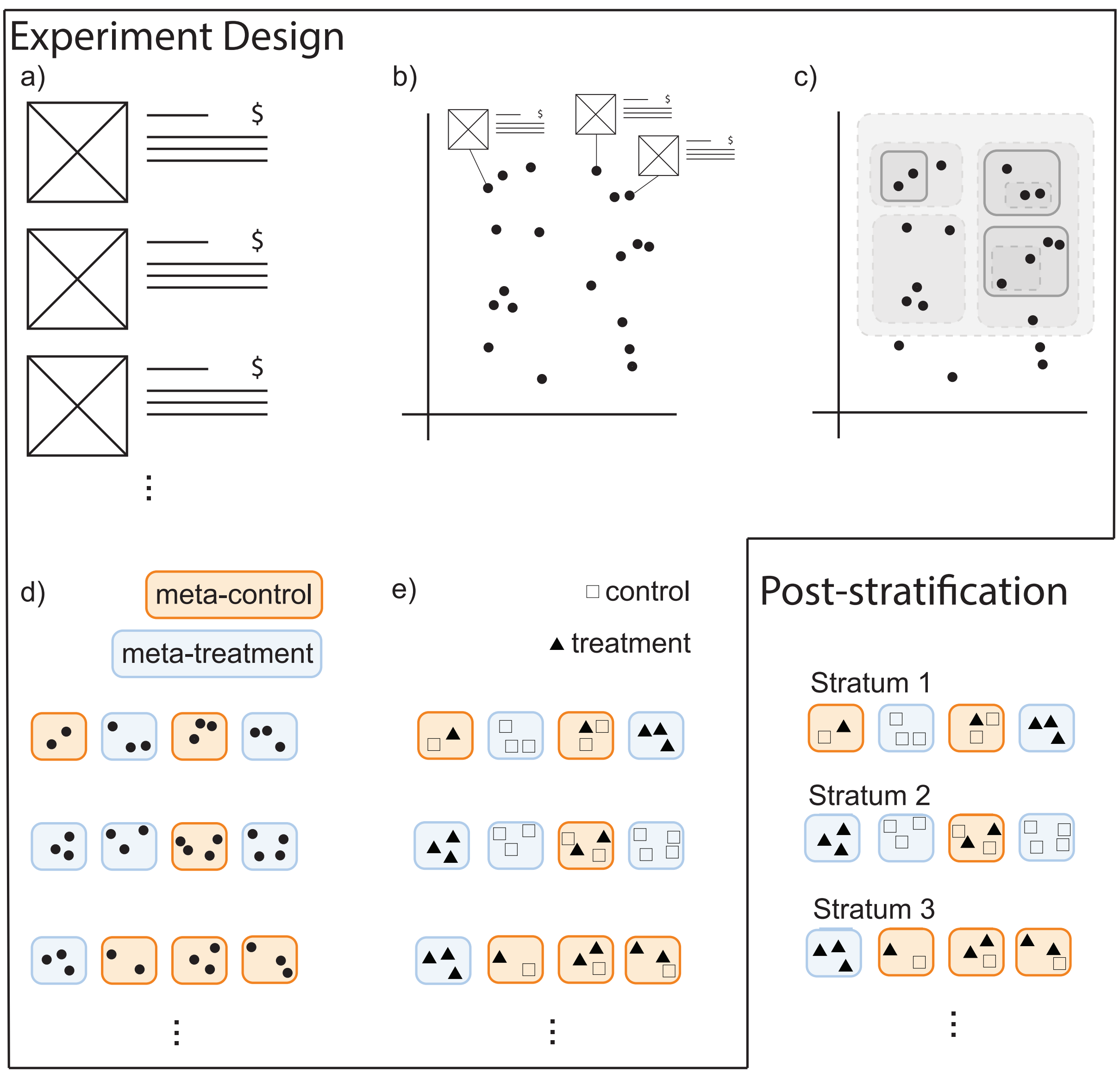


Figure A5: These maps illustrate clusters generated using the hierarchical clustering scheme described in this paper. Image from Srinivasan, 2018.

Figure A6: Coefficient estimates for the joint analysis of the fee meta-experiment. Error bars represent 95% confidence intervals. The dotted blue line correponds to a treatment effect of 0 bookings per listing. The red shaded area corresponds to values that are below the MDE (80% power, 95% confidence).

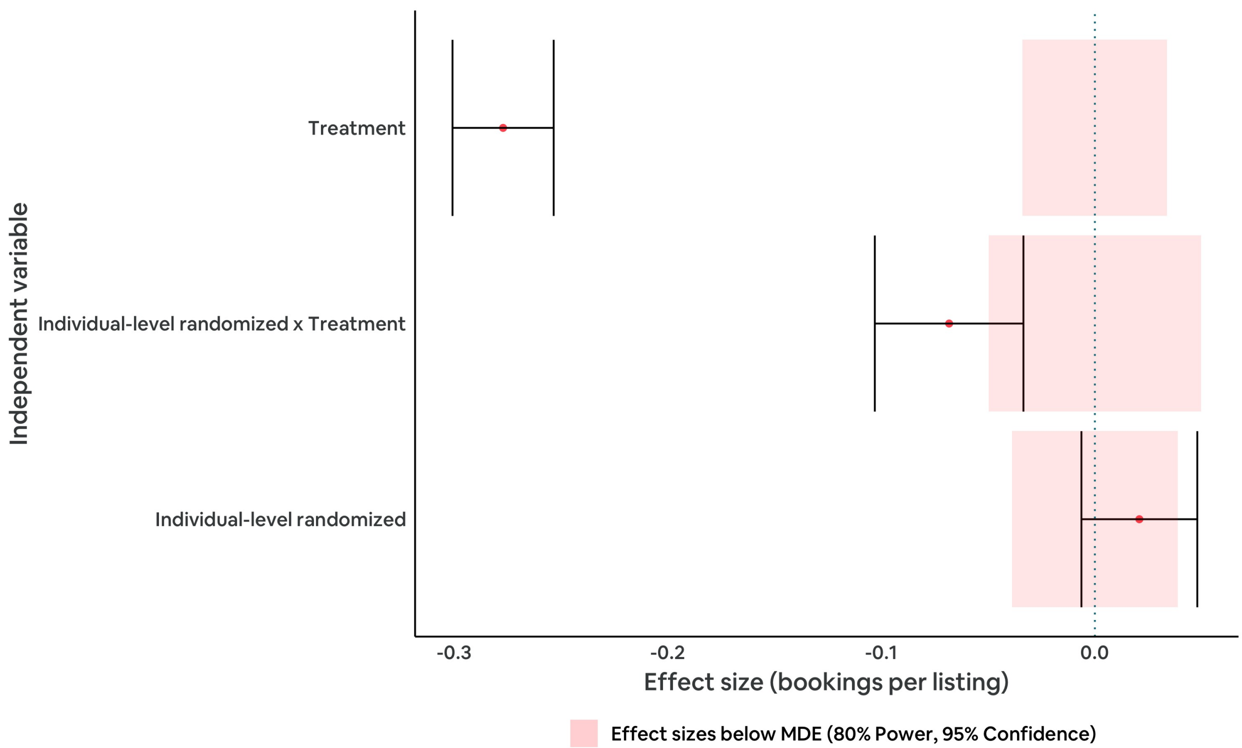
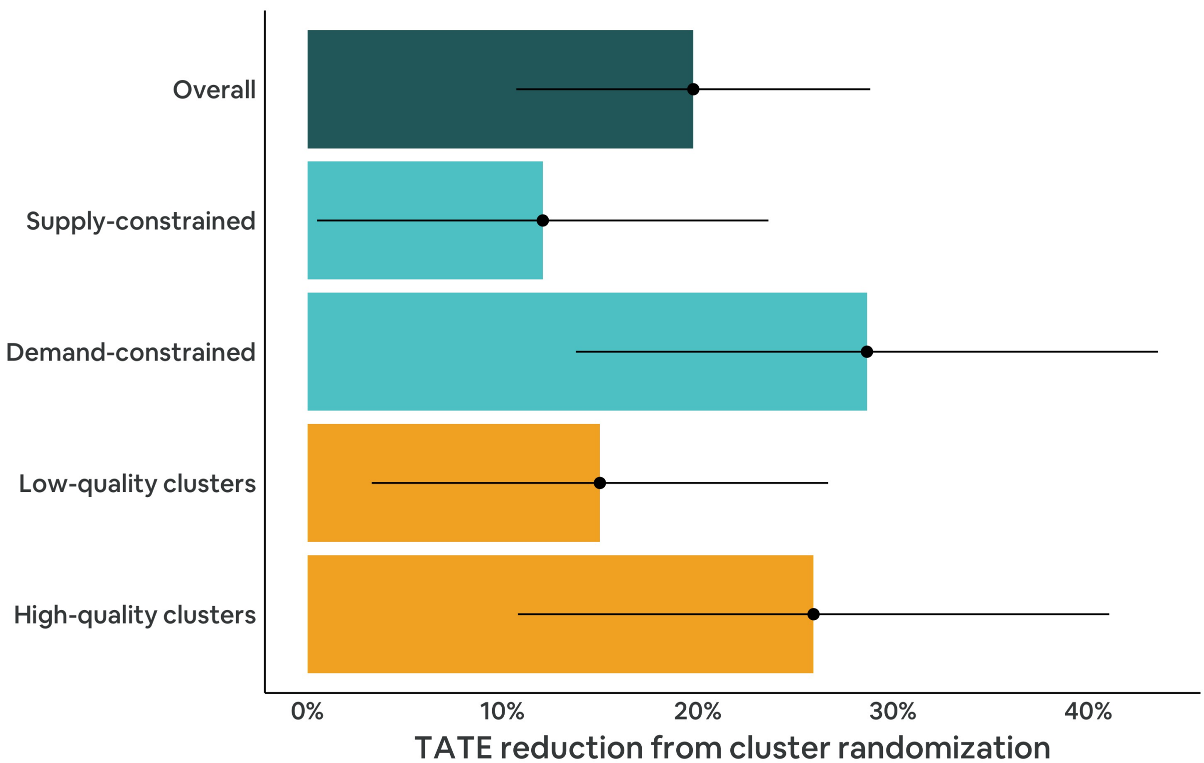


Figure A7: This graph visualizes the reduction in interference bias from cluster randomization that we estimate across different samples: overall, listings in supply-constrained geographies, listings in demand-constrained geographies, listings in geographies with low-quality clusters, and listings in high-quality clusters.



H Tables

Table A1: This table tests for statistically significant differences in pre-treatment outcomes between treatment and control in the individual-level randomized meta-treatment arm, treatment and control in the cluster-randomized meta-treatment arm, and meta-treatment and meta-control. Each comparison uses a two-sided t-test. Analysis is conducted at the individual level within the meta-control arm, at the cluster level within the meta-treatment arm, and when comparing the two meta-treatment arms.

	Individual-randomized			Cluster-randomized			Meta-experiment		
	Control	Treatment	p-value	Control	Treatment	p-value	Meta-control	Meta-treatment	p-value
Pre-treatment statistics									
Bookings	11.864	11.882	0.78	11.760	11.572	0.49	11.790	11.666	0.65
	(26.275)	(26.174)		(10.559)	(10.256)		(10.664)	(10.408)	
Nights Booked	44.984	44.953	0.90	43.288	42.497	0.37	43.195	42.893	0.73
	(101.570)	(102.677)		(34.339)	(33.646)		(34.517)	(33.994)	
Gross Guest Spend	5,920.370	5,934.694	0.72	5,554.392	5,399.833	0.37	5,587.642	5,477.087	0.53
	(15,751.420)	(15,824.250)		(6,764.090)	(6,412.172)		(6,953.921)	(6,590.321)	
$N_{\text{individuals}}$	323,734	323,643							
N_{clusters}				2,979	2,981		1,987	5,960	

Table A3: This table reports the TATE results obtained by analyzing the two meta-treatment arms separately. Individual-level randomized results are found in Column (1), and cluster randomized results are found in Column (2).

	Dependent variable: Bookings	
	Individual-level randomized	Cluster randomized
	(1)	(2)
Treatment	-0.345^{***}	-0.277^{***}
	(0.013)	(0.012)
Pre-treatment bookings	0.174^{***}	0.175^{***}
	(0.001)	(0.001)
Pre-treatment nights booked	-0.003^{***}	-0.003^{***}
	(0.000)	(0.000)
Pre-treatment gross guest spend	-0.000^{***}	-0.000^{***}
	(0.000)	(0.000)
Pre-treatment nights available	0.002^{***}	0.001^{***}
	(0.000)	(0.000)
Pre-treatment searches/night	0.267^{***}	0.033^{**}
	(0.027)	(0.015)
Stratum F.E.	Yes	Yes
Robust s.e.	Yes	Yes
Clustered s.e.	No	Yes
R^2	0.408	0.405
Adjusted R^2	0.407	0.405

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

I The potential impact of interference bias on platform profit

In this appendix, we use an example toy model to quantify the potential profit loss associated with interference bias.[24] In the model, a firm chooses a price to maximize profits given a fixed demand curve. The demand function depends on the output elasticity with respect to price, which is ex-ante unknown by the firm. Pricing experiments like the one presented in this paper are one tool available to firms in order to pin down the demand elasticity of consumers. If the firm estimates the wrong elasticity due to interference bias, then the firm's optimization procedure will lead to suboptimal profit.

To be more concrete: Define P as price and Q as quantity. In this context it is important to differentiate the actual demand elasticity from the estimated one, which can suffer from interference bias. Denote the latter as the observed elasticity (η'). Assume the demand function is iso-elastic, $Q = P^{-\eta}$, and the cost function is linear with a slope 1 for simplicity: Q. The profit is defined as a function of P: $\pi(P) = PQ - Q$. The firm equates marginal cost and marginal benefit based on their assessment of demand elasticity η', which does not align with η in the presence of interference. Therefore the price chosen is such that:

$$(1-\eta')P^{\eta'} = \eta P^{-\eta'-1} \Rightarrow P = \left(\frac{\eta'}{\eta'-1}\right).$$

Based on this choice of price, the quantity is defined based on the real (unobserved) elasticity η that drives demand,

$$Q = \left(\frac{\eta'}{\eta'-1}\right)^{-\eta}.$$

Therefore, the firm's profit is given by,

$$\pi(\eta'|\eta) = \left(\frac{\eta'}{\eta'-1}\right)^{1-\eta} - \left(\frac{\eta'}{\eta'-1}\right)^{-\eta}. \tag{3.11}$$

We define b as the elasticity bias ($b = \frac{\eta'-\eta}{\eta}$). We can restate the profit as a function of the true elasticity and the bias:

$$\pi(b,\eta) = \left(\frac{\eta(b+1)}{\eta(b+1)-1}\right)^{1-\eta} - \left(\frac{\eta(b+1)}{\eta(b+1)-1}\right)^{-\eta}. \tag{3.12}$$

[24] The inclusion of this toy model in our paper does not imply or suggest that Airbnb sets guest fees or make any other platform design choices to maximize profits.

In our setting, the parameter η could be estimated using a cluster-randomized experiment, the parameter η' could be estimated using an individual-level randomized experiment, and the bias b could be estimated by taking the difference of the two. Finally, to assess the loss due to interference bias we could calculate:

$$\Delta(b|\eta) = \frac{\pi(b|\eta) - \pi(0|\eta)}{\pi(0|\eta)}. \tag{3.13}$$

Figure O.1 shows the profit loss of a hypothetical firm for a particular estimated demand elasticity given different levels of bias.[25] We can see that the profit loss is increasing in the size of the bias, and occurs both for under- and over-estimates of the actual demand elasticity (η). As expected, profit is maximized when bias is zero.

[25] Recall that we estimate that interference bias accounts for at least 19.76% of naive TATE estimates for pricing applications on Airbnb.

J Interference Simulation

In this section, we design a simulation of booking behavior for one calendar night in a single Airbnb geography (Miami). We use this simulation framework to determine whether individual-level randomization yields biased TATE estimates, and perform a preliminary investigation into the viability of cluster randomization at reducing that bias.

Data & Network Construction

Our simulation framework is built on top of a dataset scraped by Slee (2015), which describes all of the Airbnb listings in and around Miami as of February 13, 2016. This dataset details the room type, number of reviews, average "overall satisfaction" rating, guest capacity, number of bedrooms, number of bathrooms, price per night (USD), minimum length of stay, latitude, and longitude of 8,855 Airbnb listings. Figure O.2 depicts the geospatial distribution of the listings by room type, and Table P.1 provides information about the distribution of listing-level covariates across the sample of Airbnb listings.

Before using the dataset for our analyses, we impute missing values in a number of fields: missing guest capacity, bedroom, and bathroom values are imputed using the modal value for each variable. Minimum length of stay values are capped at 30, and missing minimum length of stay values are imputed using the modal value for minimum length of stay. Missing overall satisfaction values are imputed using the mean value of non-empty entries. We also assign each listing j in our dataset an unobservable quality component,

$$\xi_j \sim N(0,1), \tag{3.14}$$

which is kept constant across all simulations. This unobserved quality component is observable to searchers, but not observable to the search algorithm or the platform. Depending on the quality of a given platform's data, factors that contribute to a listing's unobservable quality might include the quality of its photos, the responsiveness of the seller, and/or the text content of the listing's reviews

We proceed to build a "product network" for listings in this dataset. Each listing in the dataset constitutes a node in the network, and an edge between two listings implies that the listings are likely to substitute for one another when searchers are making purchase decisions. We generate an edge between two listings when the following three criteria are satisfied:[26]

1. The listings are within 1 mile of each other

[26] One could imagine using a subset of these criteria (e.g., all listings within 1 mile of each other are substitutes), or a totally unrelated criteria (e.g., listings must have co-occurred in search more than x times). For instance, in the main body of this paper and in Srinivasan (2018), items in an online marketplace are clustered based on how often they co-occur in search results.

2. The listings have the same room type
3. The difference between the guest capacity of the two listings is not greater than 1 in absolute magnitude

Using the edge heuristic described above, we produce a network that has 1,538,637 edges, and a clustering coefficient of 0.74. The average degree of nodes in the network is 173.76.

In order to simulate cluster randomized experiments, we need to divide this network into clusters. We do so using the Louvain clustering algorithm (Blondel et al., 2008). Louvain clustering attempts to maximize modularity, which is defined as

$$Q = \frac{1}{2E} \sum_{ij} \left(A_{ij} - \frac{d_i d_j}{2E} \right) 1(C_i = C_j), \tag{3.15}$$

where E is the total number of edges in the graph, A_{ij} is a $\{0, 1\}$ variable that indicates whether or not an edge exists between nodes i and j; d_i and d_j are the degrees of nodes i and j, respectively, and $1(C_i = C_j)$ is an indicator function that is equal to 1 only when i and j belong to the same cluster. At a high level, Louvain clustering attempts to maximize the density of links inside communities relative to links between communities. After running the algorithm on our listing network, the network is partitioned into 169 clusters, which have an average size of 52.40 listings.

As noted in the main body of this paper, cluster randomization can increase the variance of TATE estimates. In order to counteract this increase in variance, our simulated cluster randomization experiments use block random assignment, with blocks of size $b = 2$, to assign cluster-level treatment. To arrange clusters into pairs that will be used in that block random assignment procedure, we first calculate the average number of reviews, the average overall satisfaction score, the average number of beds, the average number of bathrooms, the average minimum stay, the average latitude, the average longitude, the percentage of private room listings, and the percentage of shared rooms for each cluster. After concatenating these metrics into a vector representing each cluster, we calculate the Mahalanobis distance (Mahalanobis, 1936) between every possible pair of clusters, and select pairs of clusters using a greedy algorithm that attempts to minimize the sum of the Mahalanobis distances between each chosen pair.

Simulation Process

In order to estimate the true TATE under different treatment interventions, as well as the bias and sampling variance of the TATE estimator under different experiment designs and analysis approaches, we create a framework for simulating the Airbnb booking process for one calendar night. Each set of simulated outcomes is generated using the following procedure.

First, a "search algorithm," $\boldsymbol{\delta}$, is drawn, with each element of $\boldsymbol{\delta}$ being generated by first drawing from the uniform distribution over the interval $[0,1]$ and then normalizing so that the sum of the elements of $\boldsymbol{\delta}$ is one, i.e.,

$$\begin{aligned}
&\delta_{k0} \sim U[0,1] \text{ for } k = 1,2,3,...,9, \\
&\delta_k = \frac{\delta_{k0}}{\sum_j \delta_{k0}}.
\end{aligned} \tag{3.16}$$

The nine elements of δ correspond to the weight that the algorithm puts on normalized versions of the following listing-level attributes: number of reviews, average satisfaction score, number of bedrooms, number of bathrooms, minimum stay, price, whether the listing is for an entire home/apt, whether the listing is for a private room, and whether a listing is a shared room. The "search algorithm" can then determine a "score" for each listing by taking the inner product of $\boldsymbol{\delta}$ and $\mathbf{x_j}$, the full vector of the listing i's centered and scaled attributes, i.e.,

$$\text{Search Score}_j = \delta \cdot \mathbf{x}_j. \tag{3.17}$$

Conditional on being issued a query by a searcher with certain geographic or attribute constraints, the algorithm will return to the searcher the ten unbooked listings with the highest search score. In cases where ten listings meeting the searcher's criteria are not available, the algorithm will return all of the listings satisfying the searcher's criteria. This allows for the possibility that the algorithm returns no listings if there are none that satisfy the searcher's requirements.

Then, $n_{searchers}$ "searchers" sequentially arrive at Airbnb and look for an available listing in our marketplace, i.e., Miami. Each searcher randomly draws a region of interest in latitude/longitude space. The locations of the box edges are drawn with uniform probability from the interval spanning from the .25th percentile of the latitudes (longitudes) belonging to listings in the geography to the 99.75th percentile of latitudes (longitudes) belonging to listings in the geography.[27] The searcher also draws a minimum guest capacity from a uniform distribution over {1,2,3,4}. The geographic boundaries and minimum guest capacity constitute the searcher's "query," and only listings that satisfy the searcher's geographic and capacity requirements will be returned by the search algorithm.

Searcher i's utility from booking listing j is given by the following equation, which is chosen so that our simulation framework is comparable to models used in the demand estimation literature (e.g., Berry, Levinsohn, and Pakes, 1995 and Nevo, 2000):

$$u_{ij} = \alpha_i(y_i - p_j) + \tilde{\mathbf{x}}_{\mathbf{j}}\boldsymbol{\beta}_i + \xi_j + \epsilon_{ij}, \tag{3.18}$$

[27]This is done to account for the potential that there are listings in our dataset that are geographic outliers.

where $\tilde{x}_j$ is the vector of listing j's attributes *besides* price, and

$$\begin{aligned} y_i &\sim N(0,1) \\ \alpha_i &\sim N(0,1) \\ \beta_{ik} &\sim N(0,1)\,\forall\, k \\ \epsilon_{ij} &\sim f(x) = e^{-x}e^{e^{-x}} \text{ (the Type I extreme-value distribution)}. \end{aligned} \tag{3.19}$$

Searcher i uses the above utility function to determine which of the up to 10 listings provided by the search algorithm they would like to book. If none of the listings have a utility greater than 0 (representing the outside option), or if the search algorithm does not return any listings meeting the searcher's query parameters, the searcher chooses not to book and exits the marketplace. Otherwise, the searcher "books" the listing that provides the highest utility to them. After this point, that listing cannot appear in future searchers' consideration sets.

Although this simulation framework simplifies the marketplace dynamics of a platform like Airbnb, we believe it can still provide insight into the degree to which interference may bias TATE estimates in online marketplace experiments, and can help determine the extent to which cluster randomization reduces that bias. We conduct simulations of marketplace activity both under marketplace-wide policy regimes (i.e., 100% treatment and 100% control), as well as under different experiment designs. We then compare the ground truth TATEs generated by contrasting outcomes under marketplace-wide policy changes to the TATE estimates produced by different experiment designs, and calculate the bias and root mean square error (RMSE) of the TATE estimates produced under different approaches to experiment design. In each of our simulations of marketplace activity, we are interested in two different outcomes. The first is whether or not a listing was booked. The second is the amount of revenue earned by a listing. We also consider two different types of treatment intervention. The first is a price reduction of .75 standard deviations for treated listings. The second is an increase of .75 standard deviations in the unobserved quality of listings.

Simulating Ground Truth

We first use our simulation framework to simulate the distribution of marketplace-level average outcomes in the case in which 100% of listings receive the treatment, and the case in which 100% of listings receive the control. For the control, as well as both the price reduction treatment and the unobserved listing quality treatment, we conduct 500 simulations of one night of booking activity in which 1,000 searchers visit Airbnb. Figure O.3 compares the sampling distributions of the rate of listings being booked and the average listing revenue under all three conditions.

A two-sided t-test between the distribution of booking rates under the control and the distribution of booking rates under the price reduction treatment yields a t-statistic of $t = 17.27$ ($p \leq 2.2 \times 10^{-16}$), with an average TATE of 0.002, whereas a two-sided t-test between the distribution of average listing revenue under the control and the distribution of average listing revenue under the price reduction treatment yields a t-statistic of $t = 1.63$ ($p = 0.10$), i.e., at the 95% level, we are unable to reject the null hypothesis that the average TATE is equal to zero. This pair of results is somewhat intuitive: when sellers lower prices, the rate at which listings are booked increases, because a greater share of listings dominate the outside option. However, that increase in booking rate does not translate into an increase in revenue, since those listings are being booked at a lower price.

A two-sided t-test between the distribution of booking rates under the control and the distribution of booking rates under the unobserved listing quality treatment yields a t-statistics of $t = 21.63$ ($p \leq 2.2 \times 10^{-16}$), with an average TATE of 0.003, whereas a two-sided t test between the distribution of average listing revenue under the control and the distribution of average listing revenue under the unobserved listing quality change treatment yields a t-statistic of 2.17 ($p = 0.03$), with an average TATE of 0.612. This pair of results is also intuitive: when the unobservable quality of listings increases, the rate at which listings are booked increases, again because a greater share of listings dominate the outside option. Because this increase in booking rate does not come hand in hand with a reduction in price, this increase in booking rate translates into an increase in revenue.

Measuring bias and RMSE

Having simulated the distribution of marketplace-level outcomes under both 100% treatment and 100% control for both our price reduction treatment and our unobservable listing quality treatment, we can now use our simulation framework to estimate the bias and RMSE of different experiment designs for both treatments. We first use our framework to simulate 500 individual-level randomized experiments, in which treatment effects are estimated using a difference in means treatment effect estimator, and then use the simulation framework to simulate 500 blocked cluster randomized experiments. Under this design, we calculate the difference in means estimator and also estimate the treatment effect using a linear regression with clustered standard errors.

Table P.2 shows the bias and RMSE of each experiment design for the booking outcome, under both the price reduction treatment and the unobserved listing quality treatments. Table P.3 shows the same information for the listing revenue outcome under both treatments. Relative to the difference in means estimator under the individual-level randomized experiment, we find that the difference in means estimator under blocked cluster randomization reduced bias by as much as 64.5%, across both metrics and both types of treatment. However, this came at the cost of increasing RMSE by as much as 204%. In other words, although the TATE estimates are on average closer to the ground truth TATE, the variance of the distribution of those estimates is much higher, i.e., statistical power is much lower.

Statistical Inference

In addition to measuring the true bias and RMSE of different experiment designs, we also assess the coverage probability associated with the 95% confidence interval that each of these approaches yields. For our difference in means estimators, we calculate the variance of the treatment effect estimate using the following expression,

$$\hat{\sigma_\tau^2} = \sigma^2(Y_{iT}) + \sigma^2(Y_{iC}), \tag{3.20}$$

where $\sigma^2(Y_{iT})$ and $\sigma^2(Y_{iC})$ are the variance of outcomes in the treatment group and control group, respectively. We also calculate the variance of the blocked cluster randomized TATE estimate when analyzed with a linear model that clusters standard errors at the level of the cluster. This approach to analyzing the data better takes into account the design of the experiment, and should lead to 95% confidence intervals with a coverage probability closer to the nominal level.

The coverage probabilities corresponding to our 95% confidence intervals are found in the rightmost columns of Tables P.2 and P.3. We find that the coverage probability of the difference in means estimator when used with the individual-level randomized design is below the nominal 95% coverage in all cases, and can be as low as 6%. The blocked cluster randomized design, when used in conjunction with the difference in means estimator, tends to move the coverage probability closer to the nominal coverage probability for the price reduction treatment, but negatively impacts the coverage probability for the unobserved quality change treatment. Regression analysis of the blocked GCR design with clustered standard errors produces coverage probabilities that are greater than the nominal 95% coverage probability, ranging from 95% to 100%.

K Proof of Proposition 1

Proof. Proof.

Under independent randomization,

$$\begin{aligned}\hat{\tau}_{ind} &= \frac{1}{N}\sum_{i=1}^{N}(E[Y_i|z_i=1]-E[Y_i|z_i=0]) \\ &= \frac{1}{N}\sum_{i=1}^{N}([B_{ii}+p\sum_{j\neq i}B_{ij}]-[p\sum_{j\neq 1}B_{ij}]) \\ &= \sum_{i=1}^{N}B_{ii}.\end{aligned} \tag{3.21}$$

Under cluster randomization,

$$\begin{aligned}\hat{\tau}_{cr} &= \frac{1}{N}\sum_{i=1}^{N}\left(E\left[Y_i|z_i=1\right]-E[Y_i|z_i=0]\right) \\ &= \frac{1}{N}\sum_{i=1}^{N}([B_{ii}+\sum_{\substack{i\neq j\\ C(j)=C(i)}}B_{ij}+p\sum_{\substack{j\neq i\\ C(j)\neq C(i)}}B_{ij}]-p\sum_{\substack{j\neq 1\\ C(j)\neq C(i)}}B_{ij}) \\ &= \frac{1}{N}\sum_{i=1}^{N}([B_{ii}+\sum_{\substack{i\neq j\\ C(j)=C(i)}}B_{ij}]) \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}B_{ij}1[C(i)=C(j)].\end{aligned} \tag{3.22}$$

If the bias of the treatment effect under graph cluster randomization is less than the bias under independent randomization, then $|\tau(1,0)-\hat{\tau}_{cr}| \leq |\tau(1,0)-\hat{\tau}_{ind}|$, which implies that

$$\frac{1}{N}\left|\sum_{i=1}^{N}\sum_{j=1}^{N}B_{ij}-\sum_{i=1}^{N}\sum_{j=1}^{N}B_{ij}1\left(C(i)=C(j)\right)\right| \leq \frac{1}{N}\left|\sum_{i=1}^{N}\sum_{j=1}^{N}B_{ij}-\sum_{i=1}^{N}B_{ii}\right|. \tag{3.23}$$

This expression can be simplified to

$$\left|\sum_{i=1}^{N}\sum_{j=1}^{N}B_{ij}1\left(C(i)\neq C(j)\right)\right| \leq \left|\sum_{i=1}^{N}\sum_{j=1}^{N}B_{ij}1(i\neq j)\right|. \tag{3.24}$$

Since the set of sellers not equal to i is a superset of the sellers not in the same cluster as i, and since all of the off-diagonal elements of $\boldsymbol{B}$ have the same sign, this will always hold true.

$\square$

L Determining Cluster Size

Previously attempted pricing meta-experiments at Airbnb had used clusters of minimum size 250, so this was considered the "status quo" cluster size. We also decided based on statistical power considerations that a cluster size threshold of 1,000 was the maximum feasible threshold. Given these facts, the choice of cluster size threshold became a direct comparison between a minimum size of 250 and a minimum size of 1,000. In choosing a cluster size threshold, the fundamental trade-off is between statistical power and capturing Airbnb demand. While smaller clusters yield more statistical power (since there are more of them), they will also do a poorer job of capturing demand, since a given user search session is more likely to contain listings from many different clusters. As a consequence, cluster quality and bias reduction will both be lower. On the other hand, larger clusters will provide less statistical power, but will do a better job of capturing demand and reducing bias. Power analysis suggested that without taking into account differences in cluster quality, our fee meta-experiment would have a minimum detectable effect (MDE) for interference bias that was 1.17 times as large if clusters of minimum size 1,000 were used as opposed to clusters of size 250. In order to determine whether this degradation in "ideal" MDE was worthwhile, we needed to measure differences in the extent to which the two sets of clusters captured demand on the platform.[28,29]

In order to make a principled decision between the two different minimum cluster sizes, we assumed that the "ideal" MDEs obtained via our power calculations would be reduced due to poor demand capture according to the relationship below:

$$MDE_{actual} = \frac{MDE_{ideal}}{\text{Demand capture}}. \tag{3.25}$$

In other words, as a given set of clusters' demand capture moved closer to 1, the MDE would approach the ideal MDE. Given this assumed relationship between actual MDE, ideal MDE, and demand capture, we determined that the 1,000 listing threshold clusters would be preferable to the 250 listing threshold clusters if

$$\frac{\text{Demand capture}_{1,000}}{\text{Demand capture}_{250}} > \frac{MDE_{ideal_{250}}}{MDE_{ideal_{1,000}}} \rightarrow \frac{\text{Demand capture}_{1,000}}{\text{Demand capture}_{250}} > 1.17 \tag{3.26}$$

Table P.4 shows the ratio of demand capture for clusters with a threshold of 1,000 listings to the demand capture for clusters with a threshold of 250 clusters according to five

[28]The analysis we describe below was originally conducted using data and clusters from February 2019, however, we present analyses using clusters generated on January 5, 2020, PDP views occurring between January 5, 2020 and January 12, 2020, and bookings occurring between January 5, 2020 and January 26, 2020. The results we report and the corresponding conclusions are qualitatively similar to those obtained using 2019 data.

[29]The meta-experiment design process occurred prior to the creation of the cluster quality metric introduced in Definition 1. Moving forward, we would recommend others use the cluster quality metric found elsewhere in this paper, as opposed to any of the demand capture metrics described below.

different demand capture measures calculated across one week of PDP views: the average share of PDPs belonging to a given cluster, the average user-level PDP Herfindahl-Hirschman index across clusters, and the percentage of users for which one cluster accounts for at least 67%, 75%, and 90% of listings viewed. Across all five of these demand capture metrics, and across different user subpopulations, the demand capture ratio is consistently above 1.17. Based on this calculation, we determined that clusters with a size threshold of 1,000 listings were preferable to those with a size threshold of 250.

M Interference bias for nights booked and gross guest spend

In this appendix, we present the results of our analyses for two additional outcomes: nights booked per listing and gross guest spend per listing. Qualitatively, our results for nights booked per listing and gross guest spend per listing are extremely similar to our results for bookings per listing.

We show that the estimated effect of the fee treatment in both the individual-level randomized meta-treatment arm and the cluster randomized meta-treatment arm on both nights booked per listing and gross guest spend per listing. We estimate in the individual-level randomized meta-treatment arm that the treatment led to a statistically significant loss of 0.308 nights booked per listing and $29.92 in gross guest spend per listing, whereas we estimate in the cluster randomized meta-treatment arm that the treatment led to a statistically significant loss of 0.257 nights booked per listing and $26.56 in booking value per listing.

In order to test whether or not there is a statistically significant difference between the TATE estimates in the two meta-treatment arms, we conduct a joint analysis of both meta-treatment arms simultaneously. Our results are displayed in Table P.5 and Figure O.4. We find statistically significant evidence of interference bias in the individual-level randomized TATE estimate for nights booked per listing, but do not find statistically significant evidence of interference bias in the individual-level randomized TATE estimate for gross guest spend per listing. Our point estimates suggest that interference accounts for 15.26% of the Bernoulli TATE estimate for nights booked per listing (stat sig.) and 9.98% of the Bernoulli TATE estimate for gross guest spend per listing (not stat. sig).

N Estimating cluster quality using browsing data

In this appendix, we describe the process used to estimate a geography-level version of the "cluster quality" metric found in Definition 1 for our clusters. In Section E, we use this metric to estimate heterogeneity in the amount of interference bias with respect to geography-level cluster quality.

Because this paper focuses on pricing-related interventions, the "true" interference matrix $\boldsymbol{B}$ that we wish to construct in order to assess cluster quality is likely the matrix of listing cross-price elasticities. Unfortunately, the full set of cross-price elasticities on Airbnb is extremely difficult, if not impossible to estimate. However, given that cross-price elasticities are at least partially driven by co-occurrence in searchers' consideration sets, we argue that search or PDP view data can be used to construct an appropriate proxy matrix for a pricing experiment on Airbnb.

In place of constructing a full proxy matrix, we use a procedure similar to the one described by Rolnick et al. (2019) to calculate the quality score $Q_C(\boldsymbol{P})$ for a given set of clusters by "folding" the underlying bipartite graph between searchers and/or PDP viewers and clusters. More specifically, let s_{ik} be the number of search impressions or PDP views by searcher i to listings in cluster C_k. We calculate the normalized folded edge between clusters k and k':

$$F_{kk'} = \sum_i \frac{s_{ik}s_{ik'}}{\sqrt{\sum_i s_{ik} \sum_i s_{ik'}}\sqrt{\sum_k s_{ik} \sum_k s_{ik}}}. \tag{3.27}$$

It follows that $Q_C(\boldsymbol{P})$, i.e., the total edge weight not captured by the clustering C consisting of M clusters is

$$Q_C(\boldsymbol{P}) = \frac{1}{M} \sum_k (1 - F_{kk}), \tag{3.28}$$

where the normalization factor of M ensures that the maximum value of $Q_C(\boldsymbol{P})$ is 1. This expression for cluster quality is higher when listings from cluster k tend to co-occur in search or PDP view sessions with other listings in cluster k, and will be maximized when listings in cluster k *only* co-occur in search or PDP view sessions with other listings in cluster k.

For computational tractability, we choose to construct our proxy matrix using PDP views, as opposed to search impressions. However, we expect that the proxy matrices constructed using these two datasets would be extremely similar.

O Additional Figures

Figure O.1: This figure plots the impact of interference bias on firm profits. The figure shows that profit is maximized when the bias does not exist. Losses increase with interference bias, and this is true for positive and negative bias.

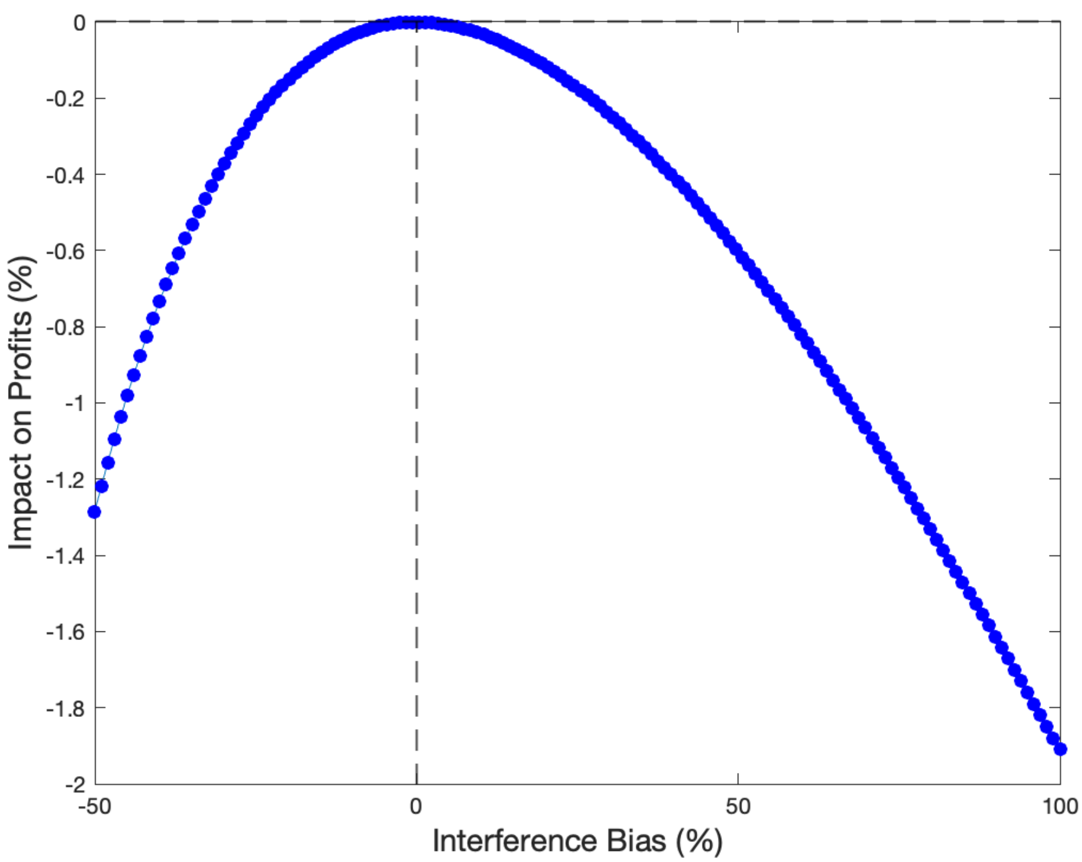


Figure O.2: The geospatial distribution of Airbnb listings in and around Miami. Color corresponds to listing type. This figure was produced with ggmap (Kahle and Wickham, 2013).

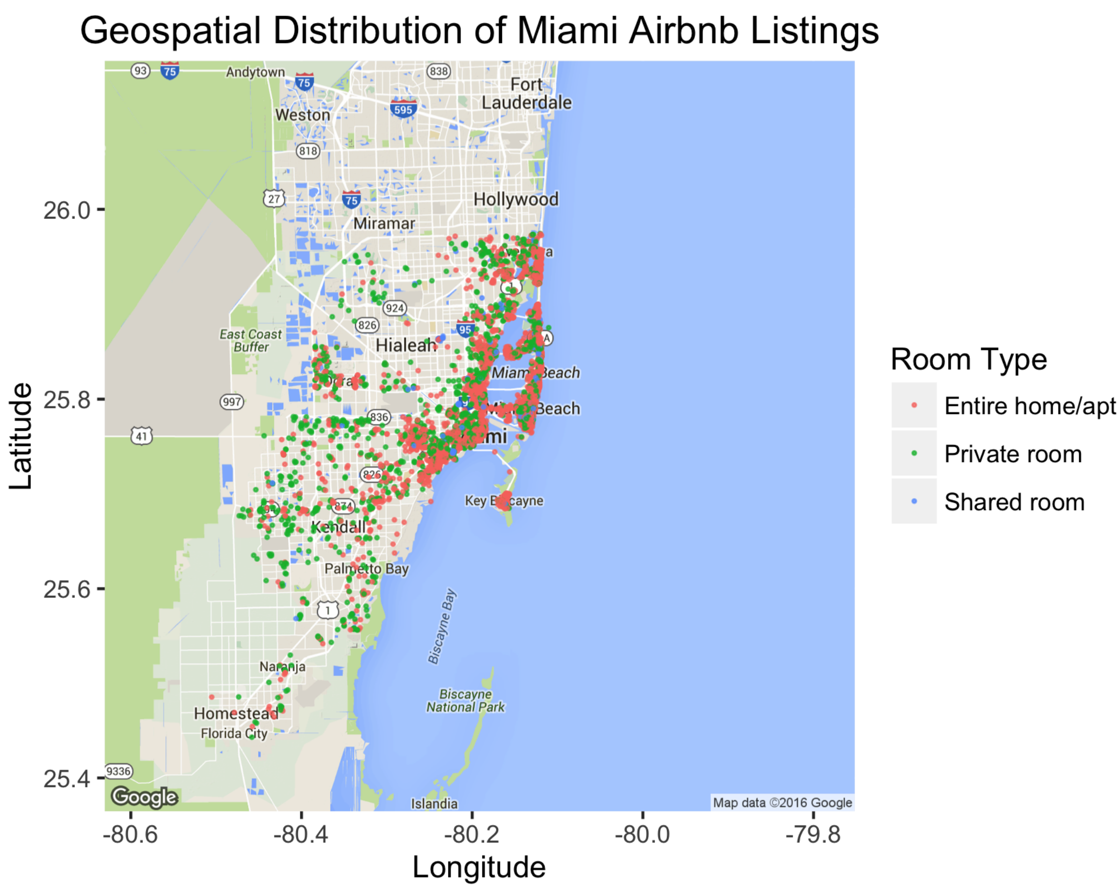
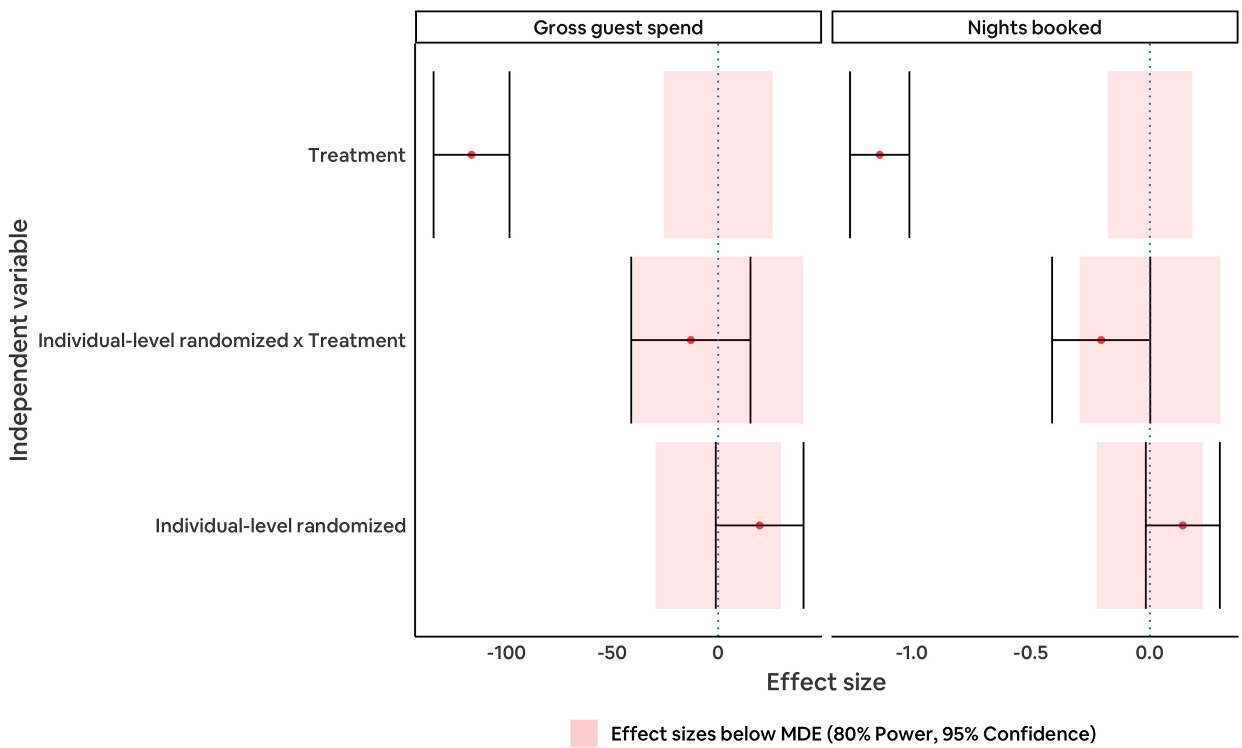


Figure O.3: Comparison of simulated marketplace-wide average outcomes when either 0% or 100% of listings are assigned treatment. The top row shows distributions when the treatment is the price reduction treatment. The bottom row shows distributions when the treatment is the unobserved listing quality change treatment. The left column shows distributions for the listing booked outcome. The right column shows distributions for the listing revenue outcome.

Figure O.4: Coefficient estimates for the joint analysis of the fee meta-experiment (nights booked per listing and gross guest spend per listing). Error bars represent 95% confidence intervals. The dotted blue line corresponds to a treatment effect of 0. The red shaded area corresponds to values that are below the MDE (80% power, 95% confidence).



P Additional Tables

Table P.1: Summary of Airbnb listing covariates for interference simulation

	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Private room	8,855	0.233	0.423	0	0	0	1
Shared room	8,855	0.026	0.158	0	0	0	1
Entire home/apt	8,855	0.742	0.438	0	0	1	1
Reviews	8,855	11.397	22.366	0	0	12	304
Overall satisfaction	6,433	4.588	0.539	1.000	4.500	5.000	5.000
Capacity	6,629	3.060	1.152	1.000	2.000	4.000	8.000
Beds	8,843	1.399	1.028	0.000	1.000	2.000	10.000
Baths	7,922	1.370	0.695	0.000	1.000	2.000	8.000
Price (USD)	8,855	226.016	406.892	15	89	249	10,000
Min Stay	8,418	3.293	9.309	1.000	1.000	3.000	365.000
Lat.	8,855	25.808	0.072	25.443	25.773	25.844	25.974
Lon.	8,855	−80.176	0.070	−80.505	−80.193	−80.129	−80.110

Table P.2: Simulated performance comparison: outcome = bookings

Treatment	Design	Estimator	Bias	RMSE	Coverage
Price Reduction	Individual-level randomization	Difference in means	0.0354	0.0393	6%
Price Reduction	Cluster randomization	Difference in means	0.0248	0.0459	20%
Price Reduction	Cluster randomization	Regression + clustered S.E.	0.0248	0.0459	95%
Unobserved quality	Individual-level randomization	Difference in means	0.0110	0.0125	56%
Unobserved quality	Cluster randomization	Difference in means	0.0039	0.0381	23%
Unobserved quality	Cluster randomization	Regression + clustered S.E.	0.0039	0.0381	99%

Table P.3: Simulated performance comparison: outcome = listing revenue

Treatment	Design	Estimator	Bias	RMSE	Coverage
Price Reduction	Individual-level randomization	Difference in means	6.08	7.26	40%
Price Reduction	Cluster randomization	Difference in means	4.30	9.06	47%
Price Reduction	Cluster randomization	Regression + clustered S.E.	4.30	9.06	97%
Unobserved quality	Individual-level randomization	Difference in means	2.26	3.93	86%
Unobserved quality	Cluster randomization	Difference in means	0.73	7.76	49%
Unobserved quality	Cluster randomization	Regression + clustered S.E.	0.73	7.76	100%

Table P.4: The ratio of demand capture for 1,000 listing threshold clusters and 250 listing threshold clusters, using different demand capture metrics and user subpopulations.

Single views?	Type of viewers	avg. cluster share	avg. HHI	% over 67%	% over 75%	% over 90%
No	All	1.32	1.36	2.36	2.46	2.38
No	Bookers	1.38	1.43	2.48	2.59	2.50
Yes	All	1.16	1.19	1.37	1.33	1.26
Yes	Bookers	1.23	1.27	1.54	1.49	1.37

Table P.5: Results of the fee meta-experiment (nights booked and gross guest spend)

	Dependent variable:	
	Nights booked	Gross guest spend
	(1)	(2)
Treatment	-1.136^{***}	-116.667^{***}
	(0.064)	(9.163)
Individual-level Randomized	0.138^{*}	19.602^{*}
	(0.079)	(10.593)
Individual-level Randomized × Treatment	-0.205^{*}	-12.931
	(0.105)	(14.384)
Pre-treatment bookings	0.297^{***}	24.769^{***}
	(0.003)	(0.376)
Pre-treatment nights booked	0.035^{***}	-4.407^{***}
	(0.001)	(0.129)
Pre-treatment gross guest spend	-0.000^{**}	0.100^{***}
	(0.000)	(0.001)
Pre-treatment nights available	0.011^{***}	1.174^{***}
	(0.001)	(0.082)
Pre-treatment searches/night	0.230^{**}	42.669^{**}
	(0.093)	(17.007)
Stratum F.E.	Yes	Yes
Robust s.e.	Yes	Yes
Semi-clustered s.e.	Yes	Yes
R^2	0.110	0.166
Adjusted R^2	0.110	0.166

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table P.6: Independent results of the fee meta-experiment (simple specification)

	Dependent variable: Bookings	
	Individual-level randomized	Cluster randomized
	(1)	(2)
Treatment	-0.343^{***}	-0.291^{***}
	(0.017)	(0.058)
Constant	2.578^{***}	2.520^{***}
	(0.013)	(0.043)
Clustered s.e.	No	Yes
R^2	0.001	0.000
Adjusted R^2	0.001	0.000

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table P.7: Results of the fee meta-experiment (simple specification)

	Dependent variable:		
	Bookings	Nights booked	Gross guest spend
	(1)	(2)	(3)
Treatment	-0.291^{***}	-1.261^{***}	-123.942^{***}
	(0.058)	(0.195)	(36.306)
Individual-level Randomized	0.058	0.182	28.039
	(0.045)	(0.162)	(28.211)
Individual-level Randomized $\times$ Treatment	-0.052	-0.080	-3.427
	(0.060)	(0.214)	(38.309)
Constant	2.520^{***}	9.517^{***}	$1{,}215.845^{***}$
	(0.043)	(0.148)	(26.830)
Semi-clustered s.e.	Yes	Yes	Yes
R^2	0.001	0.000	0.000
Adjusted R^2	0.001	0.000	0.000

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table P.8: Treatment effect heterogeneity for the fee meta-experiment (interacted)

	Dependent variable:	
	Bookings	
	Supply/demand-constrained	Cluster quality
	(1)	(2)
Treatment	−0.092***	−0.349***
	(0.005)	(0.018)
Individual-level Randomized	0.009	0.032
	(0.005)	(0.021)
Individual-level Randomized × Treatment	−0.021***	−0.076***
	(0.007)	(0.027)
Demand-constrained	−0.073***	
	(0.004)	
High-quality cluster		−0.040**
		(0.018)
Pre-treatment bookings	0.175***	0.175***
	(0.001)	(0.001)
Pre-treatment nights booked	−0.003***	−0.003***
	(0.000)	(0.000)
Pre-treatment gross guest spend	−0.000***	−0.000***
	(0.000)	(0.000)
Pre-treatment nights available	0.000***	0.001***
	(0.000)	(0.000)
Pre-treatment searches/night	0.005**	0.051**
	(0.002)	(0.020)
Individual-level randomized × Demand-constrained	−0.010*	
	(0.006)	
Treatment × Demand-constrained	0.055***	
	(0.005)	
Individual-level Randomized × Treatment × Demand-constrained	0.013	
	(0.008)	
Individual-level randomized × High-quality cluster		−0.024
		(0.028)
Treatment × High-quality cluster		0.142***
		(0.024)
Individual-level Randomized × Treatment × High-quality cluster		0.018
		(0.035)
Stratum F.E.	Yes	Yes
Robust s.e.	Yes	Yes
Clustered s.e.	Yes	Yes
R^2	0.405	0.405
Adjusted R^2	0.405	0.405

*p<0.1; **p<0.05; ***p<0.01

Table P.9: Treatment effect heterogeneity for the fee meta-experiment w.r.t. cluster quality (attribute-based definition)

	Dependent variable:	
	Bookings	
	Low-quality clusters (attributes)	High-quality clusters (attributes)
	(1)	(2)
Treatment	−0.312***	−0.231***
	(0.016)	(0.017)
Individual-level Randomized	0.010	0.033
	(0.019)	(0.020)
Individual-level Randomized × Treatment	−0.052**	−0.092***
	(0.024)	(0.026)
Pre-treatment bookings	0.173***	0.176***
	(0.001)	(0.001)
Pre-treatment nights booked	−0.002***	−0.003***
	(0.000)	(0.000)
Pre-treatment gross guest spend	−0.000***	−0.000***
	(0.000)	(0.000)
Pre-treatment nights available	0.002***	0.002***
	(0.000)	(0.000)
Pre-treatment searches/night	0.199***	0.035*
	(0.022)	(0.018)
Stratum F.E.	Yes	Yes
Robust s.e.	Yes	Yes
Clustered s.e.	Yes	Yes
R^2	0.405	0.406
Adjusted R^2	0.405	0.406

*p<0.1; **p<0.05; ***p<0.01

Bibliography

Abadie, Alberto (2002). "Bootstrap tests for distributional treatment effects in instrumental variable models". In: *Journal of the American statistical Association* 97.457, pp. 284–292.

Airbnb (2019). *Airbnb News: About Us*. URL: https://news.airbnb.com/about-us/ (visited on 11/28/2022).

Alejos, Luis (2018). "Firms' (Mis)-Reporting under a Minimum Tax: Evidence from Guatemalan Corporate Tax Returns". In: *Working Paper*.

Allingham, Michael G. and Agnar Sandmo (Nov. 1972). "Income tax evasion: a theoretical analysis". en. In: *Journal of Public Economics* 1.3, pp. 323–338. ISSN: 0047-2727. DOI: 10.1016/0047-2727(72)90010-2. (Visited on 11/10/2020).

Almunia, Miguel and David Lopez-Rodriguez (Feb. 2018). "Under the Radar: The Effects of Monitoring Firms on Tax Compliance". en. In: *American Economic Journal: Economic Policy* 10.1, pp. 1–38. ISSN: 1945-7731. DOI: 10.1257/pol.20160229. (Visited on 01/08/2020).

Angrist, Joshua, Peter Hull, and Christopher R Walters (2022). "Methods for Measuring School Effectiveness". In.

Aral, Sinan and Christos Nicolaides (2017). "Exercise contagion in a global social network". In: *Nature communications* 8.1, pp. 1–8.

Ariel, Barak, Alex Sutherland, and Lawrence W Sherman (2019). "Preventing treatment spillover contamination in criminological field experiments: the case of body-worn police cameras". In: *Journal of Experimental Criminology* 15.4, pp. 569–591.

Aronow, Peter M (2012). "A general method for detecting interference between units in randomized experiments". In: *Sociological Methods & Research* 41.1, pp. 3–16.

Aronow, Peter M and Cyrus Samii (2017). "Estimating average causal effects under general interference, with application to a social network experiment". In: *The Annals of Applied Statistics* 11.4, pp. 1912–1947.

Athey, Susan, Dean Eckles, and Guido W Imbens (2018). "Exact p-values for network interference". In: *Journal of the American Statistical Association* 113.521, pp. 230–240.

Auerbach, Alan (Oct. 2005). *Who Bears the Corporate Tax? A review of What We Know*. en. Tech. rep. w11686. Cambridge, MA: National Bureau of Economic Research, w11686. DOI: 10.3386/w11686. (Visited on 06/29/2020).

Autor, David, Claudia Goldin, and Lawrence F Katz (2020). "Extending the race between education and technology". In: *AEA Papers and Proceedings*. Vol. 110. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 347–351.

Azevedo, Eduardo M et al. (2020). "A/b testing with fat tails". In: *Journal of Political Economy* 128.12, pp. 4614–000.

Bachas, Pierre and Mauricio Soto (2021). "Corporate Taxation under Weak Enforcement". en. In: *American Economic Journal: Economic Policy* Forthcoming. ISSN: 1945-7731. DOI: `10.1257/pol.20180564`. (Visited on 03/09/2021).

Bailey, M. et al. (2020). "Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program". In: *NBER Working Paper* 26942. URL: `https://dx.doi.org/10.3386/w26942`.

Baird, Sarah et al. (2018). "Optimal design of experiments in the presence of interference". In: *Review of Economics and Statistics* 100.5, pp. 844–860.

Basri, M. Chatib et al. (Aug. 2019). *Tax Administration vs. Tax Rates: Evidence from Corporate Taxation in Indonesia*. Working Paper 26150. Series: Working Paper Series. National Bureau of Economic Research. DOI: `10.3386/w26150`. (Visited on 08/26/2020).

Basse, Guillaume and Avi Feller (2018). "Analyzing two-stage experiments in the presence of interference". In: *Journal of the American Statistical Association* 113.521, pp. 41–55.

Bastani, Spencer and Håkan Selin (Jan. 2014). "Bunching and non-bunching at kink points of the Swedish tax schedule". en. In: *Journal of Public Economics* 109, pp. 36–49. ISSN: 00472727. DOI: `10.1016/j.jpubeco.2013.09.010`. (Visited on 07/28/2020).

Bastani, Spencer and Daniel Waldenström (July 2020). "How Should Capital be Taxed?" en. In: *Journal of Economic Surveys*, joes.12380. ISSN: 0950-0804, 1467-6419. DOI: `10.1111/joes.12380`. (Visited on 07/10/2020).

Baumgartner, Erick, Raphael Corbi, and Renata Narita (2022). *Payroll Tax, Employment and Labor Market Concentration*. Tech. rep. University of Sao Paulo (FEA-USP).

Belot, Michele, Philipp Kircher, and Paul Muller (2019). "Providing advice to jobseekers at low cost: An experimental study on online advice". In: *The review of economic studies* 86.4, pp. 1411–1447.

Benmelech, Efraim, Nittai K Bergman, and Hyunseob Kim (2022). "Strong employers and weak employees how does employer concentration affect wages?" In: *Journal of Human Resources* 57.S, S200–S250.

Berger, David, Kyle Herkenhoff, and Simon Mongey (2022). "Labor market power". In: *American Economic Review* 112.4, pp. 1147–93.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile prices in market equilibrium". In: *Econometrica: Journal of the Econometric Society*, pp. 841–890.

Bertanha, Marinho, Andrew H. McCallum, and Nathan Seegert (2018). "Better Bunching, Nicer Notching". en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: `10.2139/ssrn.3144539`. (Visited on 07/10/2020).

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How much should we trust differences-in-differences estimates?" In: *The Quarterly journal of economics* 119.1, pp. 249–275.

Best, Michael Carlos, Anne Brockmeyer, et al. (Dec. 2015). "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan". In: *Journal of Political Economy* 123.6, pp. 1311–1355. ISSN: 0022-3808. DOI: `10.1086/683849`. (Visited on 01/07/2020).

Best, Michael Carlos, James S. Cloyne, et al. (Mar. 2020). "Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches". en. In: *The Review of Economic Studies* 87.2. Publisher: Oxford Academic, pp. 656–690. ISSN: 0034-6527. DOI: `10.1093/restud/rdz025`. (Visited on 09/23/2020).

Biro, Aniko et al. (2022). "Firm heterogeneity and the impact of payroll taxes". In.

Blake, Thomas and Dominic Coey (2014). "Why marketplace experimentation is harder than it seems: The role of test-control interference". In: *Proceedings of the fifteenth ACM conference on Economics and computation*. ACM, pp. 567–582.

Blomquist, Soren and Whitney K. Newey (Nov. 2017). *The Bunching Estimator Cannot Identify the Taxable Income Elasticity*. en. SSRN Scholarly Paper ID 3100040. Rochester, NY: Social Science Research Network. (Visited on 08/25/2020).

Blondel, Vincent D et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10, P10008.

Bojinov, Iavor, David Simchi-Levi, and Jinglong Zhao (2022). "Design and analysis of switchback experiments". In: *Management Science*.

Bond, Robert M et al. (2012). "A 61-million-person experiment in social influence and political mobilization". In: *Nature* 489.7415, pp. 295–298.

Bowers, Jake, Mark M Fredrickson, and Costas Panagopoulos (2013). "Reasoning about interference between units: A general framework". In: *Political Analysis* 21.1, pp. 97–124.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani (2018). "The morale effects of pay inequality". In: *The Quarterly Journal of Economics* 133.2, pp. 611–663.

Bright, Ido, Arthur Delarue, and Ilan Lobel (2022). "Reducing Marketplace Interference Bias Via Shadow Prices". In: *arXiv preprint arXiv:2205.02274*.

Bronzini, Raffaello and Eleonora Iachini (2014). "Are incentives for R&D effective? Evidence from a regression discontinuity approach". In: *American Economic Journal: Economic Policy* 6.4, pp. 100–134.

Brown, Dorothy A (2022). *The whiteness of wealth: How the tax system impoverishes Black Americans–and how we can fix it*. Crown.

Burdett, Kenneth and Dale T Mortensen (1998). "Wage differentials, employer size, and unemployment". In: *International Economic Review*, pp. 257–273.

Bustos, Sebastián et al. (June 2022). *The Race Between Tax Enforcement and Tax Planning: Evidence From a Natural Experiment in Chile*. en. Tech. rep. w30114. Cambridge, MA: National Bureau of Economic Research, w30114. DOI: `10.3386/w30114`. URL: `http://www.nber.org/papers/w30114.pdf` (visited on 11/03/2022).

Caballero, Ricardo J et al. (1995). “Plant-level adjustment and aggregate investment dynamics”. In: *Brookings papers on economic activity*, pp. 1–54.

Cameron, A Colin and Douglas L Miller (2015). “A practitioner’s guide to cluster-robust inference”. In: *Journal of human resources* 50.2, pp. 317–372.

Campos, Christopher and Caitlin Kearns (2022). “The Impact of Neighborhood School Choice: Evidence from Los Angeles’ Zones of Choice”. In: *Available at SSRN 3830628*.

Carbonnier, Clement et al. (2022). “Who benefits from tax incentives? The heterogeneous wage incidence of a tax credit”. In: *Journal of Public Economics* 206, p. 104577.

Card, David et al. (2018). “Firms and labor market inequality: Evidence and some theory”. In: *Journal of Labor Economics* 36.S1, S13–S70.

Carrillo, Paul, Dave Donaldson, et al. (July 2022). *Ghosting the Tax Authority: Fake Firms and Tax Fraud*. Working Paper. DOI: `10.3386/w30242`. URL: `https://www.nber.org/papers/w30242` (visited on 12/11/2022).

Carrillo, Paul, Dina Pomeranz, and Monica Singhal (Apr. 2017). “Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement”. en. In: *American Economic Journal: Applied Economics* 9.2, pp. 144–164. ISSN: 1945-7782. DOI: `10.1257/app.20140495`. (Visited on 03/18/2020).

Chetty, Raj (Aug. 2009). “Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance”. en. In: *American Economic Journal: Economic Policy* 1.2, pp. 31–52. ISSN: 1945-7731. DOI: `10.1257/pol.1.2.31`. (Visited on 06/09/2020).

Chetty, Raj, John N Friedman, and Emmanuel Saez (2013). “Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings”. In: *American Economic Review* 103.7, pp. 2683–2721.

Chetty, Raj, John N. Friedman, et al. (May 2011). “Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records”. en. In: *The Quarterly Journal of Economics* 126.2, pp. 749–804. ISSN: 0033-5533. DOI: `10.1093/qje/qjr013`. (Visited on 01/07/2020).

Chetty, Raj, Adam Looney, and Kory Kroft (2009). “Salience and taxation: Theory and evidence”. In: *American economic review* 99.4, pp. 1145–1177.

Chin, Alex (2018). “Central limit theorems via Stein’s method for randomized experiments under interference”. In: *arXiv preprint arXiv:1804.03105*.

Chirinko, Robert S, Steven M Fazzari, and Andrew P Meyer (2011). “A new approach to estimating production function parameters: the elusive capital–labor substitution elasticity”. In: *Journal of Business & Economic Statistics* 29.4, pp. 587–594.

Choi, Hana and Carl F Mela (2019). “Monetizing online marketplaces”. In: *Marketing Science* 38.6, pp. 948–972.

Cochran, William G and Donald B Rubin (1973). “Controlling bias in observational studies: A review”. In: *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 417–446.

Cox, David Roxbee (1958). “Planning of experiments.” In.

Criscuolo, Chiara et al. (2019). “Some causal effects of an industrial policy”. In: *American Economic Review* 109.1, pp. 48–85.

Cruces, Guillermo, Sebastian Galiani, and Susana Kidyba (2010). "Payroll taxes, wages and employment: Identification through policy changes". In: *Labour economics* 17.4, pp. 743–749.

Currie, Janet et al. (2001). "Explaining recent declines in food stamp program participation [with comments]". In: *Brookings-Wharton papers on urban affairs*, pp. 203–244.

Curtis, E Mark et al. (2021). *Capital investment and labor demand*. Tech. rep. National Bureau of Economic Research.

Dal Bo, Ernesto, Frederico Finan, and Martin A Rossi (2013). "Strengthening state capabilities: The role of financial incentives in the call to public service". In: *The Quarterly Journal of Economics* 128.3, pp. 1169–1218.

Dallava, Caroline Caparroz (2014). "Impactos da desoneracaoo da folha de pagamantos sobre o nivel de emprego no mercado de trabalho brasileiro: Um estudo a partir dos dados da RAIS". PhD thesis.

Devereux, Michael P., Li Liu, and Simon Loretz (May 2014). "The Elasticity of Corporate Taxable Income: New Evidence from UK Tax Records". en. In: *American Economic Journal: Economic Policy* 6.2, pp. 19–53. ISSN: 1945-7731. DOI: 10.1257/pol.6.2.19. (Visited on 02/15/2020).

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard (2019). "Fairness and frictions: The impact of unequal raises on quit behavior". In: *American Economic Review* 109.2, pp. 620–63.

Dube, Arindrajit, Jeff Jacobs, et al. (2020). "Monopsony in online labor markets". In: *American Economic Review: Insights* 2.1, pp. 33–46.

Dubé, Jean-Pierre and Sanjog Misra (2017). *Scalable price targeting*. Tech. rep. National Bureau of Economic Research.

Eckles, Dean, Brian Karrer, and Johan Ugander (2017). "Design and analysis of experiments in networks: Reducing bias from interference". In: *Journal of Causal Inference* 5.1.

Elzayn, Hadi et al. (2023). *Measuring and mitigating racial disparities in tax audits*. Stanford Institute for Economic Policy Research (SIEPR).

Fack, Gabrielle and Camille Landais (2016). "The effect of tax enforcement on tax elasticities: Evidence from charitable contributions in France". In: *Journal of Public Economics* 133.C. Publisher: Elsevier, pp. 23–40. ISSN: 0047-2727. (Visited on 09/21/2020).

Feit, Elea McDonnell and Ron Berman (2019). "Test & roll: Profit-maximizing A/B tests". In: *Marketing Science* 38.6, pp. 1038–1058.

Felix, Mayara (2021). "Trade, Labor Market Concentration, and Wages". In: *Job Market Paper*.

Filippas, Apostolos, Srikanth Jagabathula, and Arun Sundararajan (2019). "Managing Market Mechanism Transitions: A Randomized Trial of Decentralized Pricing Versus Platform Control". In: *Proceedings of the 2019 ACM Conference on Economics and Computation*. ACM.

Finkelstein, Amy (2009). "E-ztax: Tax salience and tax rates". In: *The Quarterly Journal of Economics* 124.3, pp. 969–1010.

Fradkin, Andrey (2015). "Search frictions and the design of online marketplaces". In: *Work. Pap., Mass. Inst. Technol.*

Fuest, Clemens, Andreas Peichl, and Sebastian Siegloch (Feb. 2018a). "Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany". en. In: *American Economic Review* 108.2, pp. 393–418. ISSN: 0002-8282. DOI: 10.1257/aer.20130570. (Visited on 08/26/2020).

— (2018b). "Do higher corporate taxes reduce wages? Micro evidence from Germany". In: *American Economic Review* 108.2, pp. 393–418.

Garin, Andrew and Filipe Silvério (2019). *How responsive are wages to demand within the firm? evidence from idiosyncratic export demand shocks*. Tech. rep.

Gechert, Sebastian et al. (2022). "Measuring capital-labor substitution: The importance of method choices and publication bias". In: *Review of Economic Dynamics* 45, pp. 55–82.

Gelber, Alexander M., Damon Jones, and Daniel W. Sacks (Jan. 2020). "Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test". en. In: *American Economic Journal: Applied Economics* 12.1, pp. 1–31. ISSN: 1945-7782. DOI: 10.1257/app.20170717. (Visited on 01/08/2020).

Gerber, Alan S and Donald P Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Gordon, Roger and Wei Li (2009). "Tax structures in developing countries: Many puzzles and a possible explanation". In: *Journal of Public Economics* 93.7-8. Publisher: Elsevier, pp. 855–866. ISSN: 0047-2727. (Visited on 03/18/2020).

Grbovic, Mihajlo and Haibin Cheng (2018). "Real-time personalization using embeddings for search ranking at Airbnb". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, pp. 311–320.

Gruber, Jonathan (1997). "The incidence of payroll taxation: evidence from Chile". In: *Journal of labor economics* 15.S3, S72–S101.

Guo, Audrey (2023). "Payroll Tax Incidence: Evidence from Unemployment Insurance". In: *arXiv preprint arXiv:2304.05605*.

Haanwinckel, Daniel (2023). *Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution*. Tech. rep. National Bureau of Economic Research.

Haanwinckel, Daniel and Rodrigo R Soares (2021). "Workforce Composition, Productivity, and Labour Regulations in a Compensating Differentials Theory of Informality". In: *The Review of Economic Studies* 88.6, pp. 2970–3010.

Hamermesh, Daniel S (1996). *Labor demand*. princeton University press.

Harasztosi, Péter and Attila Lindner (2019). "Who Pays for the minimum Wage?" In: *American Economic Review* 109.8, pp. 2693–2727.

Harberger, Arnold C. (1962). "The Incidence of the Corporation Income Tax". In: *Journal of Political Economy* 70.3. Publisher: University of Chicago Press, pp. 215–240. ISSN: 0022-3808. (Visited on 07/28/2020).

Heckman, James J and Jeffrey A Smith (2004). "The determinants of participation in a social program: Evidence from a prototypical job training program". In: *Journal of Labor Economics* 22.2, pp. 243–298.

Hendren, Nathaniel (2016). “The policy elasticity”. In: *Tax Policy and the Economy* 30.1, pp. 51–89.

Hendren, Nathaniel and Ben Sprung-Keyser (2020). “A unified welfare analysis of government policies”. In: *The Quarterly Journal of Economics* 135.3, pp. 1209–1318.

Hines Jr., James R (Dec. 2010). “Treasure Islands”. en. In: *Journal of Economic Perspectives* 24.4, pp. 103–126. ISSN: 0895-3309. DOI: 10.1257/jep.24.4.103. (Visited on 05/08/2022).

Ho, Daniel E et al. (2007). “Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference”. In: *Political analysis* 15.3, pp. 199–236.

Holland, David and Richard J. Vann (June 1998). “Income Tax Incentives for Investment”. In: *Tax Law Design and Drafting, Volume 2*. International Monetary Fund. ISBN: 978-1-55775-633-6. DOI: 10.5089/9781557756336.071. (Visited on 07/13/2020).

Holtz, David et al. (2020). “Interdependence and the cost of uncoordinated responses to COVID-19”. In: *Proceedings of the National Academy of Sciences* 117.33, pp. 19837–19843.

Holtz, David Michael (2018). “Limiting bias from test-control interference in online marketplace experiments”. MA thesis. Massachusetts Institute of Technology.

Holtzblatt, Janet et al. (2023). “Racial Disparities in the Income Tax Treatment of Marriage”. In.

Hoopes, Jeffrey L, Daniel H Reck, and Joel Slemrod (2015). “Taxpayer search for information: Implications for rational attention”. In: *American Economic Journal: Economic Policy* 7.3, pp. 177–208.

Howell, Sabrina T (2017). “Financing innovation: Evidence from R&D grants”. In: *American Economic Review* 107.4, pp. 1136–64.

Hudgens, Michael G and M Elizabeth Halloran (2008). “Toward causal inference with interference”. In: *Journal of the American Statistical Association* 103.482, pp. 832–842.

Ifrach, Bar et al. (May 2016). *Demand Prediction for Time-Expiring Inventory*. US Patent App. 14/952,576.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto (2013). “Experimental designs for identifying causal mechanisms”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176.1, pp. 5–51.

International Monetary Fund (2018). “Honduras: 2018 Article IV Consultation”. en. In: *IMF*. ISSN: 9781484366066/1934-7685. (Visited on 02/05/2020).

— (2019a). *Corporate Taxation in the Global Economy*. English. Washington, D.C.: International Monetary Fund. (Visited on 02/14/2020).

— (2019b). *World Revenue Longitudinal Data*. URL: https://data.world/imf/world-revenue-longitudinal-dat.

Jacobson, Louis S, Robert J LaLonde, and Daniel G Sullivan (1993). “Earnings losses of displaced workers”. In: *The American economic review*, pp. 685–709.

Jäger, Simon and Jörg Heining (2022). *How substitutable are workers? evidence from worker deaths*. Tech. rep. National Bureau of Economic Research.

Janet, Currie, Card David, Quigley John, et al. (2006). “The Take Up of Social Benefits”. In: *Public Policy and the Income Distribution*, pp. 80–148.

Johari, Ramesh et al. (2022). "Experimental design in two-sided platforms: An analysis of bias". In: *Management Science*.

Johnson, George E (1997). "Changes in earnings inequality: the role of demand shifts". In: *Journal of economic perspectives* 11.2, pp. 41–54.

Kahle, David and Hadley Wickham (2013). "ggmap: Spatial Visualization with ggplot2". In: *The R Journal* 5.1, pp. 144–161. URL: `http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf`.

Kang, Ji Hoon, Chan Hee Park, and Seoung Bum Kim (2016). "Recursive partitioning clustering tree algorithm". In: *Pattern Analysis and Applications* 19.2, pp. 355–367.

Karabarbounis, Loukas and Brent Neiman (2014). "The global decline of the labor share". In: *The Quarterly journal of economics* 129.1, pp. 61–103.

Katz, Lawrence F and Kevin M Murphy (1992). "Changes in relative wages, 1963–1987: supply and demand factors". In: *The quarterly journal of economics* 107.1, pp. 35–78.

Kleven, Henrik J (2018). "Calculating Reduced-Form Elasticities Using Notches". en. In: *Technical note*.

Kleven, Henrik J and Mazhar Waseem (2013a). "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan". In: *The Quarterly Journal of Economics* 128.2, pp. 669–723.

— (May 2013b). "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan". en. In: *The Quarterly Journal of Economics* 128.2, pp. 669–723. ISSN: 0033-5533. DOI: `10.1093/qje/qjt004`. (Visited on 01/07/2020).

Kleven, Henrik Jacobsen (2016). "Bunching". In: *Annual Review of Economics* 8.1, pp. 435–464. DOI: `10.1146/annurev-economics-080315-015234`. (Visited on 01/07/2020).

Kleven, Henrik Jacobsen, Martin B. Knudsen, et al. (2011). "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark". en. In: *Econometrica* 79.3, pp. 651–692. ISSN: 1468-0262. DOI: `10.3982/ECTA9113`. (Visited on 07/22/2020).

Kleven, Henrik Jacobsen and Claus Thustrup Kreiner (2006). "The marginal cost of public funds: Hours of work versus labor force participation". In: *Journal of Public Economics* 90.10-11, pp. 1955–1973.

Kline, Patrick et al. (2019). "Who profits from patents? rent-sharing at innovative firms". In: *The quarterly journal of economics* 134.3, pp. 1343–1404.

Kroft, Kory et al. (2020). *Imperfect competition and rents in labor and product markets: The case of the construction industry*. Tech. rep. National Bureau of Economic Research.

Krusell, Per et al. (2000). "Capital-skill complementarity and inequality: A macroeconomic analysis". In: *Econometrica* 68.5, pp. 1029–1053.

Kugler, Adriana and Maurice Kugler (2009). "Labor market effects of payroll taxes in developing countries: Evidence from Colombia". In: *Economic development and cultural change* 57.2, pp. 335–358.

Kugler, Adriana, Maurice Kugler, and Luis Omar Herrera Prada (2017). *Do payroll tax breaks stimulate formality? Evidence from Colombia's reform*. Tech. rep. National Bureau of Economic Research.

Lachowska, Marta, Alexandre Mas, and Stephen A Woodbury (2020). "Sources of displaced workers' long-term earnings losses". In: *American Economic Review* 110.10, pp. 3231–66.

Lagos, Lorenzo (2019). "Labor Market Institutions and the Composition of Firm Compensation: Evidence from Brazilian Collective Bargaining". In: *Job Market Paper*.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler (2022). "Imperfect competition, compensating differentials, and rent sharing in the US labor market". In: *American Economic Review* 112.1, pp. 169–212.

Li, Hannah et al. (2022). "Interference, bias, and variance in two-sided marketplace experimentation: Guidance for platforms". In: *Proceedings of the ACM Web Conference 2022*, pp. 182–192.

Liu, Lan and Michael G Hudgens (2014). "Large sample randomization inference of causal effects in the presence of interference". In: *Journal of the american statistical association* 109.505, pp. 288–301.

Liu, Min, Jialiang Mao, and Kang Kang (2021). "Trustworthy and Powerful Online Marketplace Experimentation with Budget-split Design". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3319–3329.

Londoño-Vélez, Juliana and Javier Ávila-Mahecha (2019). "Can Wealth Taxation Work in Developing Countries? Quasi-Experimental Evidence from Colombia". In: Job Market paper.

Mahalanobis, Prasanta Chandra (1936). "On the generalized distance in statistics". In: *Proceedings of the National Institute of Sciences (Calcutta)* 2, pp. 49–55.

Maloney, William F (2004). "Informality revisited". In: *World development* 32.7, pp. 1159–1178.

Manning (2011). "Imperfect competition in the labor market". In: *Handbook of labor economics*. Vol. 4. Elsevier, pp. 973–1041.

— (2021). "Monopsony in labor markets: A review". In: *ILR Review* 74.1, pp. 3–26.

Manski, Charles F (2000). "Economic analysis of social interactions". In: *Journal of economic perspectives* 14.3, pp. 115–136.

Mayshar, Joram (1990). "On measures of excess burden and their application". In: *Journal of Public Economics* 43.3, pp. 263–289.

McFadden, Daniel et al. (1973). "Conditional logit analysis of qualitative choice behavior". In.

Mikolov, Tomas, Kai Chen, et al. (2013). "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781*.

Mikolov, Tomas, Ilya Sutskever, et al. (2013). "Distributed representations of words and phrases and their compositionality". In: *Advances in neural information processing systems*, pp. 3111–3119.

Miratrix, Luke W, Jasjeet S Sekhon, and Bin Yu (2013). "Adjusting treatment effect estimates by post-stratification in randomized experiments". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.2, pp. 369–396.

Mittal, Shekhar, Ofir Reich, and Aprajit Mahajan (June 2018). "Who is Bogus?: Using One-Sided Labels to Identify Fraudulent Firms from Tax Returns". en. In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. Menlo Park and San Jose CA USA: ACM, pp. 1–11. ISBN: 978-1-4503-5816-3. DOI: 10.1145/3209811.3209824. URL: https://dl.acm.org/doi/10.1145/3209811.3209824 (visited on 11/03/2022).

Moffitt, Robert A et al. (2001). "Policy interventions, low-level equilibria, and social interactions". In: *Social dynamics* 4.45-82, pp. 6–17.

Moffitt, Robert A (2007). *Means-tested transfer programs in the United States*. University of Chicago Press.

Moore, Ryan T (2012). "Multivariate continuous blocking to improve political science experiments". In: *Political Analysis* 20.4, pp. 460–479.

Mosberger, Pálma (2016). *Accounting versus real production responses among firms to tax incentives: bunching evidence from Hungary*. en. Tech. rep. 2016/3. Magyar Nemzeti Bank (Central Bank of Hungary). (Visited on 02/15/2020).

Naritomi, Joana (2019). "Consumers as tax auditors". In: *American Economic Review* 109.9, pp. 3031–3072.

Nevo, Aviv (2000). "A practitioner's guide to estimation of random-coefficients logit models of demand". In: *Journal of economics & management strategy* 9.4, pp. 513–548.

Oberfield, Ezra and Devesh Raval (2021). "Micro data and macro technology". In: *Econometrica* 89.2, pp. 703–732.

OECD (2019). *Taxing Wages 2019*, p. 640. DOI: https://doi.org/https://doi.org/10.1787/tax_wages-2019-en. URL: https://www.oecd-ilibrary.org/content/publication/tax_wages-2019-en.

Ohrn, Eric (2023). "Corporate tax breaks and executive compensation". In: *American Economic Journal: Economic Policy* 15.3, pp. 215–255.

Perry, Guillermo (2007). *Informality: Exit and exclusion*. World Bank Publications.

Piketty, Thomas and Gabriel Zucman (2014). "Capital is back: Wealth-income ratios in rich countries 1700–2010". In: *The Quarterly journal of economics* 129.3, pp. 1255–1310.

Pouget-Abadie, Jean et al. (2018). "Optimizing cluster-based randomized experiments under monotonicity". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2090–2099.

Raval, Devesh R (2019). "The micro elasticity of substitution and non-neutral technology". In: *The RAND Journal of Economics* 50.1, pp. 147–167.

Risch, Max (2024). "Does taxing business owners affect employees? Evidence from a change in the top marginal tax rate". In: *The Quarterly Journal of Economics* 139.1, pp. 637–692.

Rolnick, David et al. (2019). "Randomized experimental design via geographic clustering". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2745–2753.

Rosenbaum, Paul R (2007). "Interference between units in randomized experiments". In: *Journal of the american statistical association* 102.477, pp. 191–200.

Rosenbaum, Paul R and Donald B Rubin (1984). "Reducing bias in observational studies using subclassification on the propensity score". In: *Journal of the American statistical Association* 79.387, pp. 516–524.

Rubin, Donald B (1974). "Estimating causal effects of treatments in randomized and non-randomized studies." In: *Journal of educational Psychology* 66.5, p. 688.

Saez, Emmanuel (Aug. 2010). "Do Taxpayers Bunch at Kink Points?" en. In: *American Economic Journal: Economic Policy* 2.3, pp. 180–212. ISSN: 1945-7731. DOI: 10.1257/pol.2.3.180. (Visited on 01/07/2020).

Saez, Emmanuel, Manos Matsaganis, and Panos Tsakloglou (2012). "Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece". In: *The Quarterly Journal of Economics* 127.1, pp. 493–533.

Saez, Emmanuel, Benjamin Schoefer, and David Seim (2019). "Payroll taxes, firm behavior, and rent sharing: Evidence from a young workers' tax cut in Sweden". In: *American Economic Review* 109.5, pp. 1717–63.

Saez, Emmanuel and Gabriel Zucman (2019). *The triumph of injustice: How the rich dodge taxes and how to make them pay*. WW Norton & Company.

Saveski, Martin et al. (2017). "Detecting network effects: Randomizing over randomized experiments". In: *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, pp. 1027–1035.

Sävje, Fredrik, Peter Aronow, and Michael Hudgens (2021). "Average treatment effects in the presence of unknown interference". In: *Annals of statistics* 49.2, p. 673.

Scherer, Clovis (2015). "Payroll tax reduction in Brazil: Effects on employment and wages". In: *ISS Working Paper Series/General Series* 602.602, pp. 1–64.

Sinclair, Betsy, Margaret McConnell, and Donald P Green (2012). "Detecting spillover effects: Design and analysis of multilevel experiments". In: *American Journal of Political Science* 56.4, pp. 1055–1069.

Slee, Tom (2015). *Airbnb Data Collection: Methodology and Accuracy*. URL: http://tomslee.net/airbnb-data-collection-methodology-and-accuracy.

Slemrod, J. and S. Yitzhaki (2001). "Integrating Expenditure and Tax Decisions: The Marginal Cost of Funds and the Marginal Benefit of Projects". In: *National Tax Journal* 54.2, pp. 189–201. URL: https://dx.doi.org/10.17310/ntj.2001.2.01.

Slemrod, Joel and Wojciech Kopczuk (2002). "The optimal elasticity of taxable income". en. In: *Journal of Public Economics* 84.1. Publisher: Elsevier, pp. 91–112. (Visited on 09/21/2020).

Srinivasan, Sharan (2018). *Learning Market Dynamics for Optimal Pricing*. URL: https://medium.com/airbnb-engineering/learning-market-dynamics-for-optimal-pricing-97cffbcc53e3 (visited on 05/06/2019).

Suarez Serrato, Juan Carlos and Owen Zidar (2023). "Who Benefits from State Corporate Tax Cuts? A Local Labor Market Approach with Heterogeneous Firms: Reply". In: *American Economic Review*.

Suárez Serrato, Juan Carlos and Owen Zidar (Sept. 2016a). "Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms". en.

In: *American Economic Review* 106.9, pp. 2582–2624. ISSN: 0002-8282. DOI: 10.1257/aer.20141702. (Visited on 06/29/2020).

Suárez Serrato, Juan Carlos and Owen Zidar (2016b). "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms". In: *American Economic Review* 106.9, pp. 2582–2624.

Tchetgen, Eric J Tchetgen and Tyler J VanderWeele (2012). "On causal inference in the presence of interference". In: *Statistical methods in medical research* 21.1, pp. 55–75.

Ugander, Johan et al. (2013). "Graph cluster randomization: Network exposure to multiple universes". In: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 329–337.

Ulyssea, Gabriel (2018a). "Firms, Informality, and Development: Theory and Evidence from Brazil". In: *American Economic Review* forthcoming.

— (2018b). "Firms, informality, and development: Theory and evidence from Brazil". In: *American Economic Review* 108.8, pp. 2015–47.

Waseem, Mazhar (Apr. 2020). *Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax*. en. SSRN Scholarly Paper ID 3594223. Rochester, NY: Social Science Research Network. (Visited on 02/16/2022).

Yagan, Danny (2015). "Capital tax reform and the real economy: The effects of the 2003 dividend tax cut". In: *American Economic Review* 105.12, pp. 3531–3563.

Ye, Peng et al. (2018). "Customized Regression Model for Airbnb Dynamic Pricing". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, pp. 932–940.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein (2022). "Monopsony in the US labor market". In: *American Economic Review* 112.7, pp. 2099–2138.

Zucman, Gabriel (Nov. 2014). "Taxing across Borders: Tracking Personal Wealth and Corporate Profits". en. In: *Journal of Economic Perspectives* 28.4, pp. 121–148. ISSN: 0895-3309. DOI: 10.1257/jep.28.4.121. (Visited on 02/14/2020).

Zwick, Eric (2021). "The costs of corporate tax complexity". In: *American Economic Journal: Economic Policy* 13.2, pp. 467–500.

Zwick, Eric and James Mahon (2017). "Tax policy and heterogeneous investment behavior". In: *American Economic Review* 107.1, pp. 217–48.