Title
Essays in Public Economics

Author
Giaccobasso Amorena, Matias

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Essays in
Public Economics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Management

by

Matias Giaccobasso Amorena

2023

ABSTRACT OF THE DISSERTATION

Essays in
Public Economics

by

Matias Giaccobasso Amorena
Doctor of Philosophy in Management
University of California, Los Angeles, 2023
Professor Paola Giuliano, Chair

This dissertation is comprised of a series of essays in public economics, development economics, labor economics, and behavioral economics, that provide a comprehensive illustration of the core of my research agenda. In the first paper, I focus on the transfer side of public economics. More specifically, I study the effects of a cash transfer program on individuals' transition to adulthood. In the second and third papers, I focus on the taxation side of public economics. In the second paper, I focus on top-income earners and how they react to a change in the top personal income tax rates. Finally, in the third paper, I focus on how individuals make their tax compliance decisions, in particular, in the role of tax morale mechanisms. Collectively, this series of essays contribute to building our understanding of how individuals interact with tax and transfer policies. Empirical evidence on these interactions are key inputs for the discussion of optimal policy design.

The dissertation of Matias Giaccobasso Amorena is approved.

Melanie Sharon Wasserman

Manisha Shah

Ricardo Perez-Truglia

Paola Giuliano, Committee Chair

University of California, Los Angeles

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Chapter 3 is a version of the submitted article:

Giaccobasso, M., Nathan, B. C., Perez-Truglia, R., & Zentner, A. (2022). Where do my tax dollars go? tax morale effects of perceived government spending (No. w29789). National Bureau of Economic Research. Available at NBER:
https://www.nber.org/papers/w29789 or http://dx.doi.org/10.3386/w29789

All listed co-authors are principal investigators and contributed in equal shares in the elaboration of the article. Full acknowledgments are included in the corresponding chapter.

Chapter 4 is a version of the article in preparation:

Bergolo, M., Burdin, G., De Rosa, M., and Giaccobasso, M., Leites, M., and Rueda, H., (2011). How do Top Earners Respond to Taxation? Evidence from a Tax Reform in Uruguay. (No. 4007698) Social Science Research Network. Available at SSRN:
https://ssrn.com/abstract=4007698 or http://dx.doi.org/10.2139/ssrn.4007698

All listed co-authors are principal investigators and contributed in equal shares in the elaboration of the article. Full acknowledgments are included in the corresponding chapter.

VITA

Education	Ph.D. Candidate in Management UCLA, Anderson School of Management	2017 - 2023
	Bachelors Degree in Economics Universidad de la Republica, Uruguay	2007 - 2013

Articles

"Dissecting Inequality-Averse Preferences" - Bergolo, M., Burdin, G., Burone, S., De Rosa, M., Giaccobasso, M., and Leites, M. (2022). Dissecting Inequality-Averse Preferences. Journal of Economic Behavior & Organization, 200, 782-802

"Tax Audits as Scarecrows. Evidence from a large field experiment" - Bergolo, M.; Ceni, R.; Cruces, G.; Giaccobasso, M. and Perez-Truglia, R. (2021). Tax Audits as Scarecrows. Evidence from a large field experiment. American Economic Journal: Economic Policy, 15(1), 110-153

Digging into the Channels of Bunching: Evidence from the Uruguayan Income Tax. - Bergolo, M., Burdin, G., De Rosa, M., Giaccobasso, M., & Leites, M. (2021). Digging into the Channels of Bunching: Evidence from the Uruguayan Income Tax. The Economic Journal, 131(639), 2726-2762.

"Misperceptions about Tax Audits." - Bergolo, M., Ceni, R., Cruces, G., Giaccobasso, M., and Perez-Truglia, R (2018), Misperceptions about Tax Audits. AEA Papers and Proceedings, 108 : 83-87.

Awards	Dissertation Year Fellowship, UCLA Graduate Division (2022) Global Research Award - Center for Global Management (2020, 22) Research Grant - Universidad de la Republica (2021-2023) Fellowship - Anderson School of Management, UCLA (2017-2022)
Service	Refereeing: Journal of Political Economy: Microeconomics; American Economic Journal: Economic Policy, The Economic Journal, Journal of Public Economics, Public Finance Review, Economic Inquiry, Journal of Economic Inequality, Journal of Behavioral and Experimental Economics, Estudios Economicos
Teaching	Teaching Assistant at UCLA: Business and Economics in Emerging Markets (Prof. Leo Feler), MBA/FEMBA - Spring, 2022; Managerial Economics and Economic Analysis for Managers (2019-2022) Teaching Assistant at UDELAR: Labor Economics, Undergraduate level; Poverty and Inequality, Undergraduate level(2016-2017)
Relevant Employment	Research Assistant for Prof. Natalie Bau UCLA (2020-2022); Prof. Nico Voigtländer, UCLA (2021); Research Fellow California Policy Lab (Fall, 2019); Prof. Melanie Wassermann, UCLA (Spring/Summer, 2019; Summer 2020); Research Assistant Instituto de Economía (IECON), Department of Economics, Universidad de La República (2014-2017)

CHAPTER 1

Introduction

In the last few years, I have developed a research agenda that lies in the intersection of public economics, development economics, labor economics, and behavioral economics. This research agenda consists of a series of projects that aim to understand how different public policies -cash transfers and taxes- and the information environment surrounding them affect individual behavior and preferences, with their implications regarding efficiency and equity. This dissertation is comprised of a series of essays in these areas that illustrate the core of my research agenda.

My research on welfare and anti-poverty programs aims to understand whether current welfare programs are effective in achieving their ultimate medium and long-run goals of improving socio-economic conditions for the next generations and whether they are worth implementing. In Chapter 2, *Growing Up Over the Social Safety Net: The Effects of a Cash Transfer Program on the Transition to Adulthood*, I present novel evidence about the effects of a permanent, large-scale, and government-implemented cash transfer program, the Uruguayan *PANES/AFAM-PE*. I focus on three critical dimensions of individuals' transition to adulthood: education, fertility, and labor market decisions. I use a unique combination of individual-level administrative records that exhaustively describes the year-by-year trajectory of the effects. Using a Regression Discontinuity Design that exploits the use of a poverty score to define eligibility to participate in the program, I show that the program reduces women's teenage pregnancies by 9.4p.p., increases participants' early adulthood labor market participation by 6.4p.p., months worked by 4.4, and earnings by about 12%. The evidence on education outcomes is mixed but suggests a stronger attachment to the secondary education system. Consistent with a postponement of women's first birth being the main driver, changes in labor market outcomes are observed exclusively for women. The evidence suggests that cash transfers may be viable policies to improve children's future life trajectories and contribute to reducing the labor market gender gap.

Taxation is another main area of my research agenda. In this regard, my work focuses on two issues. First, I am interested in understanding how current tax schedules affect individuals' behavior, such as reporting and labor market decisions, and how these responses affect social welfare in general. Second, I am also interested in improving our knowledge about the determinants of tax compliance. In Chapters 3 and 4, I provide some partial answers to these questions.

In Chapter 3, *How do Top Earners Respond to Taxation? Evidence from a Tax Reform in Uruguay*, I focus on top-income earners and how they react to a change in the top personal income tax rates. In particular, we exploit an unprecedented combination of exhaustive administrative records and quasi-random variation in the tax rates for the top 1% labor income earners. This allows us to uncover reliably the different margins of responses, which is particularly important under the presence of fiscal externalities. Exploiting a tax reform implemented in Uruguay in 2012, we estimate an intensive margin elasticity of 0.577, partially explained by a real labor supply adjustment. Responses on the extensive margin are larger (semi-elasticity of 2.479), driven mainly by labor-to-corporate income shifting (semi-elasticity of -1.967). The efficiency costs of the reform represent 31% of the projected tax revenue.

In Chapter 4, *Where Do My Tax Dollars Go? Tax Morale Effects of Perceived Government Spending*, I focus on individuals' determinants of tax compliance. In this case, my co-authors and I aim to answer the following research question: Do perceptions about how the government spends tax dollars affect the willingness to pay taxes? We designed a field experiment to test this hypothesis in a natural, high-stakes context and via revealed preferences. We measured how taxpayers perceive the destination of their tax dollars, such as the percentage of their property taxes that funds public schools. We find that even though accurate information is available, taxpayers still hold substantial misperceptions. We use an information-provision experiment to induce exogenous shocks to these perceptions. Using administrative data on property tax appeals, we measure the causal effect of perceived government spending on the willingness to pay taxes. We find that perceptions about government spending have a significant effect on the probability of filing a tax appeal in a manner that is consistent with reciprocal motivation: individuals are more willing to pay taxes if they believe that the government services funded by those taxes will be of greater personal benefit to them. We discuss implications for the study of tax morale.

Collectively, this series of essays contribute to building our understanding of how individuals interact with tax and transfer policies. Empirical evidence on these interactions is a key input for the discussion of optimal policy design. Furthermore, the three papers included in this dissertation underline the importance of having access to detailed information, either through surveys or administrative records, to understand the nuances of individuals' behavior as a response to changes in the public policy or information environment.

CHAPTER 2

Growing-Up Over The Social Safety Net: The Effects of a Cash Transfer Program on the Transition to Adulthood

Matias Giaccobasso, University of California, Los Angeles [1]

Abstract

Countries spend a large share of their budgets on aid to families with children, with cash transfers being one of the most used policy instruments for this purpose. This paper presents novel evidence about the effects of a permanent, large-scale, and government-implemented cash transfer program, the Uruguayan *PANES/AFAM-PE*. I focus on three critical dimensions of individuals' transition to adulthood:

[1]As described in the acknowledgments page, Chapter 2 is a version of my job market paper. The latest version of this article, as well as the complement online appendix, can be found here. I am very grateful to Ricardo Perez-Truglia, Paola Giuliano, Manisha Shah, Melanie Wasserman, Marcelo Bergolo, and Andrea Vigorito for their continuous support and guidance. I thank Natalie Bau, Sebastian Edwards, Clemence Tricaud, Nico Voigtländer, and Romain Wacziarg for their thoughtful comments and support during my doctoral studies. I also thank Sebastian Calonico, Matias Cattaneo, Raj Chetty, Erzo Luttmer, Dario Tortarolo, and all participants at the NBER Public Economics Meeting for their helpful comments. I thank Elisa Failache, Misha Galashin, Sebastian Ottinger, Zach Sauers, Maria Sauval, and Joan Vila for inspiring discussions. Romina Quagliotti provided superb research assistance. This project was supported by the Center for Global Management at Anderson School of Management.

education, fertility, and labor market decisions. I use a unique combination of individual-level administrative records that exhaustively describes the year-by-year trajectory of the effects. Using a Regression Discontinuity Design that exploits the use of a poverty score to define eligibility to participate in the program, I show that the program reduces women's teenage pregnancies by 9.4p.p., increases participants' early adulthood labor market participation by 6.4p.p., months worked by 4.4, and earnings by about 12%. The evidence on education outcomes is mixed but suggests a stronger attachment to the secondary education system. Consistent with a postponement of women's first birth being the main driver, changes in labor market outcomes are observed exclusively for women. The evidence suggests that cash transfers may be viable policies to improve children's future life trajectories and contribute to reducing the labor market gender gap.

2.1 Introduction

Worldwide, governments spend billions of dollars on social safety net (SSN) policies to reduce poverty, especially for vulnerable households with children.[2] Cash transfers are one of the most used policy instruments for this purpose. Because they represent sizable investments, have the potential to affect multiple generations through a wide range of mechanisms, and trigger ethical debates about who are the deserving beneficiaries, they are a highly controversial topic. For instance,

[2]The world's average SSN expenditure is 1.93% of the GDP (e.g., in Japan) and ranges between 0.01% (Cote d'Ivoire) and 10.1% (South Sudan). In the US, SSN expenditure is 1.34% of the GDP, including programs such as TANF, Child Support programs, WIC, EITC, and Food Stamps. In OECD countries, the average SSN expenditure is 2.6% (e.g., Germany) and ranges between 0.7% (Turkey) and 4.9% (Denmark). In developing countries, expenditure is considerably lower, with an average of 1.7% (Thailand) and a median of 1.23% (China). Both for OECD and developing countries, this is larger, for instance, than the total tax revenue from property taxes.

some argue that cash transfers could be beneficial for children's life trajectories. They might help reduce child poverty, improve economic security, and connect vulnerable individuals to the labor force. Others argue that they could be inefficient or even hurtful for long-run upward mobility, especially when providing unconditional support.[3] The lack of consensus is not exclusive to the policy debate. The academic literature has yet to thoroughly describe whether and how cash transfers affect individuals' life trajectories, especially when the focus is on children. This has substantial policy implications since it could change the direction of the cost-benefit evaluation and the chances of the policy surviving the political cycle in cases where effects do not materialize immediately (Aizer et al., 2022).

This paper fills this gap by studying how a permanent, large-scale, and government-implemented cash transfer program, the Uruguayan *PANES/AFAM-PE*, affects the life trajectories in terms of education, fertility, and labor market decisions of individuals who benefited from the policy during their childhood. These outcomes are strongly correlated with poverty and opportunities for mobility and speak directly to the long-run goal of most cash transfer programs. I focus on individuals' decisions during the period that spans between 15 and 30 years old. This is a period that overlaps with what sociologists and psychologists usually refer to as "transition to adulthood" or "emerging adulthood" (Settersten Jr et al., 2008; Arnett, 2000). Adulthood is a distinct and socially recognized stage of life, usually defined by a series of markers related to the culmination of education cycles, labor market participation, residential independence, marriage, and fertility. Until recently, "adolescence" was the term used to describe the life stage between childhood and

[3]These two types of arguments can be found in recent discussions about the Child Tax Credit expansion in the United States For instance, in the blog post by Scott Winship from the conservative American Enterprise Institute (2021) and in a quote from Rep. Danny Davis (2021) in a press release from the "First Focus Campaign for Children". However, these expressions are representative of the typical discussion surrounding cash transfers across the world.

adulthood. However, this transition has become more nuanced in the last fifty years, and the idea of a uniform adolescence is becoming socially and economically inexact (Settersten Jr et al., 2008). Cash transfers might play a key role in shaping this transition, with long-term consequences in traditional socio-economic outcomes such as income, and human capital accumulation, among others. Understanding the dynamics of these effects is critical to assess if cash transfers are fulfilling their ultimate goal of reducing structural poverty and inequality and increasing opportunities for mobility.

The program that I study, *PANES/AFAM-PE*, was implemented in Uruguay in 2005 and remains in place until today. It consists of a cash transfer that represents, on average, more than 50% of the self-reported pre-program household income. To remain in the program, households were required to satisfy some conditions, such as school attendance and health check-ups, typical of Conditional Cash Transfer programs (CCTs). *PANES/AFAM-PE* has two main goals. First, in the short term, it aims to provide an additional source of income to help beneficiary households to overcome immediate needs related to their disadvantaged socio-economic status. Second, in the long run, the program aims to encourage human capital accumulation of beneficiary children for a more permanent transition out of poverty. *PANES/AFAM-PE* was broadly publicized, even before its implementation, and rapidly became the most generous anti-poverty program in the country's history (Manacorda et al., 2011). It accounts for 0.4% of the Uruguayan GDP and reaches more than 10% of Uruguayan households, comparable to programs such as *PROGRESA* (Mexico) and *Bolsa Familia* (Brazil). Since its inception, the program has only suffered minor changes aimed at increasing its coverage.

The ideal setting to analyze the causal effects of cash transfers on individuals' transition to adulthood would be to randomly assign a group of families to receive government assistance and wait for the children to grow up to compare some key socio-economic outcomes measured at different ages between treated and control families. The research hypothesis is that, compared to children that did not receive government assistance, children that grew up in families that did receive it have better outcomes during their transition to adulthood, such as lower teenage pregnancies or a higher probability of being employed in the formal labor market. Random assignment is atypical if one wants to analyze permanent, large-scale, and government-implemented programs. *PANES/AFAM-PE* is not an exception. Hence, I leverage some features of the program design and rely on a quasi-experimental Regression Discontinuity Design that closely mimics the ideal experiment. More specifically, I leverage the fact that eligibility to participate in *PANES/AFAM-PE* is determined based on a poverty score and exploit the sharp change in the probability of treatment just at the eligibility threshold. Intuitively, I compare individuals who obtained a poverty score just above the eligibility threshold with individuals that obtained a score just below. Under some assumptions (i.e., continuity and monotonicity), this comparison yields an estimate of the (local) average treatment effect of the program.[4]

To conduct the empirical analysis, I have assembled a unique and exhaustive longitudinal dataset that contains individual-level information both on participation and outcomes of interest for the universe of Uruguayan individuals. This rare dataset is built on a series of administrative records provided by different government agencies that can be merged at the individual level using a unique masked

[4]Because the program has been in place uninterruptedly since 2005, households might have applied to it more than once and obtained multiple scores. Following the approach proposed by Jepsen et al. (2016), I use the score of the *first* application to the program as an instrument for treatment status. Section 2.5 discusses more in-depth the challenges of this type of setting and the reasons that lead to the use of the score of the first application as an instrument for the treatment variable.

identification number. In particular, I combine information on *PANES/AFAM-PE* applications and participation, births, education enrollment, and labor market participation from different institutions.

The main findings can be grouped into two. In the first place, *PANES/AFAM-PE* substantially affects individuals' transition to adulthood. In terms of fertility outcomes, measured at 18 years old, for instance, participating in the program reduces women's number of births by 0.108, or 41.95% of the control group average. This effect is statistically significant and economically relevant. First, in percentage terms, this reduction is equivalent to the reduction observed in Uruguay's adolescent fertility rate between 1960 and 2020. Second, the effect sizes are substantially larger compared to other policy changes that reduced teenage pregnancies in Uruguayan women, such as abortion legalization (Cabella and Velázquez, 2022) or a large-scale intervention that granted access to subdermal contraceptive implants (Ceni et al., 2021). This negative effect is also consistent with very recent findings for the EITC in the US (Michelmore and Lopoo, 2021), CCT programs in Latin America (e.g., Araujo and Macours, 2021; Attanasio et al., 2021; Barham et al., 2018), or Africa (e.g., Baird et al., 2011).

The effects on labor market outcomes are also strong. For instance, participating in the program causes an increase of 6.4p.p. (9.69%) in the probability of having worked four consecutive months in the formal sector at or before age 23. In addition, the program increases by 4.4 (or 19.77%) the cumulative number of months worked by age 23 and earnings by around 12%. As a reference, the size of the effect observed for participation is similar in size, but with the opposite sign, to the negative of *PANES/AFAM-PE* on parents' formal labor force participation (Bergolo and Cruces, 2021). These results are also in the same line that recent promising evidence for the US (Barr et al., 2022; Bailey et al., 2020; Bastian and

Michelmore, 2018; Aizer et al., 2016), and Latin American countries (Araujo and Macours, 2021; Attanasio et al., 2021; Parker and Vogl, 2018).

Unlike the effects concerning fertility and employment decisions, the effects on education outcomes are more nuanced. The program weakly increases the number of years enrolled in secondary education at or before age 18 by 0.25 years (or 10%), with null effects on the extensive margin (i.e., ever being enrolled). This effect is consistent with related literature that usually finds increases in years of education that are close to 0.2-0.4 (e.g., Araujo and Macours, 2021; Aizer et al., 2016; Behrman et al., 2011). Exploratory evidence suggests that this less clear pattern of effects masks important heterogeneous effects by education level. In particular, the effects are observed exclusively for higher secondary education, with an increase in the maximum grade of high-school enrollment. This indicates that the relevant margin of response could be associated with academic progress rather than extensive margin responses. Moreover, the weak positive effects observed on education outcomes do not translate into increased enrollment in tertiary education. The mixed evidence found for education outcomes is not exclusive of *PANES/AFAM-PE*. For instance, despite the positive effects reported in Araujo and Macours (2021), Aizer et al. (2016), or Behrman et al. (2011), some other papers find null (Dustan, 2020; Barham et al., 2018), or even negative effects women's education (Bastian et al., 2022).

The second main set of findings suggests that the program's positive effects on employment, months worked, and earnings are explained by changes in the timing of women's fertility decisions. Two pieces of evidence support this interpretation. First, the effects estimated on the overall population, as described in the first group of findings, are driven exclusively for women. For instance, the effect of *PANES/AFAM-PE* on women's labor market participation measured at age 23 is

11.2p.p. (16.98%), while the effect on the number of months worked by this age is 5.92 (26.5%). On the contrary, the estimated effects for men are 4.4p.p. and -0.764 months, respectively, and both are statistically insignificant.

The second piece of supporting evidence is provided by the dynamic analysis. Comparing the age profiles of the effects for each outcome suggests a compelling story. The effects of *PANES/AFAM-PE* start to manifest as early as age 16, with the negative effects on teenage pregnancies. These effects peak (in absolute value) around the ages of 17-18 and continue being negative, although smaller in size, until women reach 27 years old. At this age, they become positive but statistically insignificant. The overall age profile of the effects suggests that *PANES/AFAM-PE* did not change women's overall fertility preferences. Rather, it led to a postponement of births that otherwise would have occurred in the late teens. A similar but oppositely signed pattern is observed when looking at women's labor market outcomes. The positive effects on labor market outcomes start as early as at age 18 and remain positive until the mid-twenties when they start to attenuate and trend toward 0. This attenuation coincides exactly with the attenuation (and even reversal) of the effects observed on fertility outcomes. The fact that labor market effects are exclusively observed on women, combined with such similar but inverse dynamics of fertility and labor market outcomes, suggests that *PANES/AFAM-PE* improves young women's labor market participation through a postponement of births.

The evidence reported in this paper has policy implications for the design, implementation, and evaluation of cash transfer policies. First, it illustrates that cash transfer policies may help reduce labor market gender gaps, even when they are not specifically designed for this purpose. Despite the signs of attenuation observed by the late twenties, changes in the timing of events still have strong consequences from a life-cycle perspective due to the existence of fixed costs and flatter wage profiles for mothers (e.g., Bratti, 2015; Miller, 2011). Given that the motherhood penalty explains a sizable share of the labor market gender gap, policies that promote a postponement of pregnancies that otherwise would have occurred during teenage years might be particularly effective. Second, by having strong effects on critical decisions such as the age of first birth, cash transfers have the potential to spill over to future generations. For instance, later-life pregnancies could lead to higher test scores or improved educational and psychological outcomes for children of the third generation (e.g., as discussed in Sobotka, 2010). These results suggest that cash transfer programs can have long-lasting effects that should be considered when assessing their effectiveness, making them much more attractive. Overall my paper shows that cash transfers can be a viable policy instrument to reduce long-run poverty by improving participants' labor market outcomes.

This paper makes two main contributions. First, it contributes to the literature that analyzes the effect of cash transfers on children's outcomes. In its current status, this literature can be synthesized into two snapshots. On the one hand, there is well-documented evidence about positive early-file effects of cash transfers on children's health and education outcomes in high-, low-, and middle-income countries (Hoynes and Schanzenbach, 2018; Bastagli et al., 2019, 2016; Bosch and Manacorda, 2012; Fiszbein et al., 2009). This literature usually measures children's outcomes at around age 18. On the other hand, some incipient literature focuses on later-life outcomes. This recent literature provides promising evidence for the US (e.g., Barr et al. 2022; Bailey et al. 2020; Price and Song 2018; Bastian and Michelmore 2018; Hoynes et al. 2016), Mexico's *PROGRESA* (Araujo and Macours, 2021), Nicaragua's *Red de Proteccion Social* (Barham et al., 2018), and

Colombia's *Familias en Accion* (Attanasio et al., 2021).[5] In general, this literature focuses on outcomes measured later in life, at around the age of 30. However, to my knowledge, there is still no evidence that describes the full dynamic of the effects of cash transfers on transitions to adulthood. Doing so represents a major empirical challenge, mostly because it is extremely data-demanding.

By collecting administrative records from different data sources that capture both a long period of time and a wide range of outcomes, I am able to overcome this critical empirical challenge. To my knowledge, this is the first paper to exhaustively describe the effects of cash transfers on education, fertility, and labor market decisions during children's transition to adulthood between 18-30 years old with this degree of detail and based on high-quality administrative records. Furthermore, this paper shows that understanding the full dynamics of the effects is key for a correct assessment of the effects of cash transfers on individuals' life trajectories. For instance, if the effects of the program on fertility were measured only at age 18, one would conclude that the program led to a reduction in fertility. On the other hand, focusing on the effects measured at age 30 would lead to the conclusion that the program did not have an effect on fertility. In both cases, one would have completely overlooked the postponement effect.

This paper also presents some additional methodological improvements within this literature strand, which has usually been affected by some data or research design limitations. One example is the use of geographic or temporal variation in the rollout of a program as a source of exogenous variation. Using aggregated units of analysis leads to intention-to-treat estimates rather than average treatment effects. This could be a considerable limitation in contexts where take-up is

[5]Some recent reviews can be found in (Aizer et al., 2022) for the US social safety net, or (Molina Millán et al., 2019) for conditional cash transfers in low- and middle-income countries

imperfect.[6] Having access to individual-level data, jointly with a research design that exploits changes in treatment status at the individual level, allows me to provide estimates of (local) average treatment effects in addition to the intention to treat effects. Local effects may be the parameters of interest from a government's perspective, for instance, when considering an expansion of the program.

Another limitation of the existing literature, particularly relevant in developing countries, is the lack of high-quality administrative records. This implies that researchers usually need to conduct their own follow-up surveys to collect information on the post-intervention period. The high costs associated with this strategy usually result in follow-up surveys comprising very few data points for very specific cohorts. Access to high-quality administrative records allows me to overcome some of the attrition or small sample size issues associated with survey data. Finally, it is also important to note that sometimes the existing evidence on the effects of cash transfers corresponds to very specific settings. For instance, temporary interventions in rural areas, conducted by local or international NGOs. The analysis presented in this paper not only provides the first characterization of the effects of cash transfers on individuals' transition to adulthood but does so in the context of a permanent, large-scale, and government-implemented policy. This is presumably a more general context compared to other small-scale, context-specific interventions.

The second main contribution is to the literature on gender inequality in the labor market (see general surveys in Altonji and Blank, 1999, Blau and Kahn, 2017,

[6]There are some exceptions, such as Aizer et al. (2016) or Price and Song (2018) that do analyze the effects of a cash transfer at the individual level, but they are subject to additional limitations. For instance, Aizer et al. (2016) restrict their analysis to male children, who do not tend to change their surnames and, therefore, can be tracked over time. Price and Song (2018) propose a matching algorithm that only allows measuring effects on families with more than one child.

or Olivetti and Petrongolo, 2016), but with a focus in the relationship between motherhood and labor market outcomes (e.g., Bratti, 2015; Miller, 2011; Waldfogel, 1998 or more recent works such as Kleven et al., 2019,?). Regarding this literature, my paper provides complementary evidence highlighting how fertility decisions, particularly during the critical adolescent ages, might have long-lasting consequences in terms of labor market participation, experience, and earnings. Furthermore, this paper illustrates how cash transfer policies can be useful in reducing labor market gender gaps, even when they are not specifically designed for this purpose.

This paper makes contributions to two additional broader strands of literature. On the one hand, to the literature in Demography that analyzes the causes and consequences of the "postponement transition." The postponement transition describes the increase in the mean age of first birth that has affected rich countries since the 1970s and, more recently, Latin-American countries (Rosero-Bixby et al., 2009). This transition has been explained by several factors, such as the spread of modern contraception or legalization of abortion, but also by changes in socio-economic trends, such as prolonged education, women's emancipation, and the postponement of other adulthood milestones such as finishing education, leaving the parental home, or forming a couple (see Sobotka, 2010; Mills et al., 2011 for exhaustive reviews). Hence, the fertility postponement is strongly connected to the idea of a more diffuse transition to adulthood. This literature has discussed the relationship between fertility postponement and labor market decisions mostly based on macro-level correlations. My paper provides additional evidence using micro-level data that shows a causal relation between improvements in socioeconomic conditions of the households and changes in fertility patterns.

Finally, this paper contributes to a broader literature on the role of household income on children's outcomes. The bulk of the empirical literature has found that early childhood interventions have strong effects on long-term outcomes (see Almond et al., 2018 for a thorough review). However, a growing literature shows that shocks to household income when children are older may also be effective (Bulman et al., 2021; Manoli and Turner, 2018; Cesarini et al., 2016; Akee et al., 2010; Dynarski, 2003). I contribute to this literature by showing the effects of a policy-driven income shock on household income for children that were, on average, 13 years old when they first applied to the program.

The rest of the paper is structured as follows. In Section 3.2, I describe the main features of *PANES/AFAM-PE*. Then, in Section 2.3, I discuss the main mechanisms that could drive the effects of cash transfers on the outcomes of interest, with a specific focus on how these mechanisms might evolve over time. In Section 2.4, I describe the data used in the analysis, while in Section 2.5, I describe the main features of the Regression Discontinuity approach used to estimate the causal effects of the program. In Section 3.5, I report the main results from the empirical analysis. Section 2.7 discusses the main theoretical mechanisms that could explain the results. Finally, Section 2.8 concludes and discusses the main policy implications.

2.2 Institutional Background: *PANES/AFAM-PE*

2.2.1 Context of Implementation

Uruguay is a middle-high-income country in South America with a population of about 3.5 million inhabitants. In 2018, Uruguay had the second largest GDP in the region (USD 23,585), only led by Chile (USD 25,526).[7] In the same year, Uruguay was ranked 55th in the world in terms of Human Development Index and classified within the very high HDI group. Uruguay's lower secondary completion rate in 2018 was 56.8%, which is comparable to Argentina's but lower than in Mexico, Brazil, and Chile; and way behind richer countries such as the United States, Sweden, or even Italy and Spain. Uruguay's adolescent fertility rate (i.e., births per 1,000 women aged 15-19) is 58.24, similar to Brazil and Argentina, but higher than in Chile and Costa Rica, and substantially higher compared to the United States, Norway, Sweden, Spain, and Italy.

Uruguay has a well-established tradition of a strong public sector. In 2018, Uruguay's tax revenue as a percentage of the GDP was 29.2%, the largest in the region, only behind Brazil. Compared to the rest of the world, this share is higher than in the United States and close to the OECD average. In terms of its social protection system, Uruguay has one of the oldest and most developed systems in the region.[8] In 1943, Uruguay implemented family allowances for families with underage children for the first time. However, until the end of the 90s, these benefits were restricted to registered employees.

The program I focus on, *PANES/AFAM-PE*, was implemented in 2005. It was conceived as a temporary social relief program in response to the economic downturn that affected most Latin American countries in the early 2000s, and it remained in place until December, 2007.[9] In the next section, I describe in detail

[7] See Online Appendix for further details.

[8] For instance, old age pensions were established for the first time in 1919; maternity leave was implemented in 1937; sickness and disability insurance in 1950; and unemployment benefits in 1958.

[9] After the economic crisis of the early 2000s, unemployment and poverty sky-rocketed. By the end of 2004, the poverty rate for urban areas reached 40%, and the unemployment rate was close to 15%.

the key elements of its design.

2.2.2 Design of *PANES/AFAM-PE*

The implementation of *PANES/AFAM-PE* can be divided into two phases. The first started in 2005 under the name of *PANES* and remained in place until 2007. The second, *AFAM-PE*, started immediately after. The program was widely publicized and rapidly became the largest anti-poverty program in the country's history (Manacorda et al., 2011). *PANES/AFAM-PE* is comparable both in its design and in its relative size to programs such as *PROGRESA-Oportunidades* (Mexico) and *Bolsa Familia* (Brazil). Its total cost has been consistently around 0.4% of the Uruguayan GDP.

The main component of *PANES* was a cash transfer targeted at the poorest 150,000 households in the country. The program had two primary goals. First, in the short run, it aimed to alleviate the high poverty levels caused by the economic crisis.[10] Second, in the medium- and long-run, its goal was to encourage human capital accumulation in poor households to help them move out of structural poverty. The base cash transfer was USD 133, expressed in January 2008 PPP terms.[11] In addition, the program provided a supplementary transfer between USD 29 and USD 78 to households with underage children (70% of the participant households). Overall, the cash transfer represented more than 50% of the average self-reported household income in the application forms.[12]

---

[10] In 2005, the country's poverty rate was close to 21%. However, the child poverty rate was even higher: 36.6% for all children in urban areas and 60% for children between 0-5 years old.

[11] In local currency, this corresponded to UYU 1,360. In what follows, all income variables are converted to 2008 PPP using CPI and PPP conversion factors.

[12] See Online Appendix for a more detailed description of the characteristics of the universe of application forms. It is important to note that the income used as a reference to calculate this share is self-reported. However, since the program also had an income threshold rule to

Between 2005 and 2007, more than 180,000 different households (about 17.6% of all households in the country) applied to *PANES/AFAM-PE*. Eligibility was determined based on two criteria. First, applicant households must have a per-capita household income below USD 131 (or between 27.9% and 41.7% of the April 2005 poverty line). Second, households must have a poverty score below an arbitrarily defined threshold that varies by region. Households were visited by program officials who conducted a thorough interview to evaluate their socio-economic situation. This information was used to compute the poverty score, which consists of the predicted probability of being below a critical per capita income level.[13] Households with a poverty score above a certain threshold are eligible to participate, while households with a score below the threshold are deemed ineligible. After being accepted, participant households were supposed to satisfy school attendance, regular health check-ups, and monthly per-capita income requirements, but the program did not rigorously enforce these conditions until April 2013.

On January 1st, 2008 *PANES* was expanded and re-branded into *AFAM-PE*. While formally, *AFAM-PE* was a new program that substituted the original *PANES*, in practice, it was implemented as an expansion with very slight differences. The program's main components - i.e., eligibility criteria and type of benefits and conditionalities - remained the same. There were only three differences between *PANES* and *AFAM-PE*. The first one is that *AFAM-PE* established

define eligibility, households may have misreported income to become eligible. Therefore this share must be interpreted as an upper bound. As an alternative reference, in April 2005, the household *per capita* poverty line was USD 314.19 for rural areas and USD 469.95 for urban areas in 2008 PPP terms.

[13]The variables used to calculate the score included the overall quality of the building, the number of people living in the household, the number of rooms, the presence of underage children, average years of education, and type of employment, among others. More details about how the poverty score was computed can be found in Online Appendix and in Manacorda et al., 2011; Amarante et al., 2016.

the presence of underage children in the household as a requirement for eligibility. The second is a more lenient poverty score eligibility threshold. This change aimed to increase the coverage of the program. Finally, the program changed the formula used to define the amount to be transferred. The new structure established a baseline payment of USD 57 per child from 0-17 but was subject to an equivalence scale of 0.6. In addition to the base payment, each household would receive an additional USD 24 per child enrolled in the secondary education system, also subject to an equivalence scale of 0.6. Finally, *AFAM-PE* beneficiaries were also supposed to fulfill education and health check-up conditionalities. However, these started to be enforced only beginning in April 2013. In subsequent years the enforcement quality depended on the will of the Ministry of Social Development and other high-ranked officials.

The transition between the two phases was straightforward. Provided that families had underage children, *PANES* participants were automatically enrolled in *AFAM-PE*. Furthermore, households rejected during the first phase were automatically enrolled in the second phase if they satisfied the new eligibility requirements. Figure 2.1 presents a summary of the main components of *PANES/AFAM-PE*.

2.3 Conceptual Framework: Cash Transfers and Decisions Within the Household

Cash transfers may cause behavioral responses in several margins across all household members. In its simplest form (i.e., unconditional), cash transfers induce a pure income effect that leads households to increase their demand for normal goods (e.g., consumption goods or leisure). However, cash transfers contingent on certain behaviors (i.e., conditional) are associated with a more complex set of potential behavioral responses. These may trigger reactions associated with a substitution effect due to changes in the relative opportunity costs of the alternatives included in individuals' choice sets. The analysis becomes even more convoluted when decisions are allowed to interact inter-temporarily or when market imperfections such as information frictions or collective household models are considered. In this section, I motivate the research hypotheses by broadly discussing how *PANES/AFAM-PE*, or any other similar CCT program, may affect education, labor market, and fertility decisions of individuals who benefited from the program in their childhood.[14] The list of mechanisms discussed in this section is not intended to be exhaustive. The goal is to provide an overview of what the literature has proposed and discussed when analyzing the effects of CCTs on education, fertility, and labor market decisions.

Income and substitution effects: Consider a simple unitary model where households decide over leisure, school, fertility, and labor market activities. In this setting, CCTs could imply both income and substitution effects. The income effect, associated with additional household resources, would increase the quantity demanded of normal goods (e.g., leisure) to the detriment of labor market activities for all household members. Furthermore, if households obtain direct utility from children's current human capital or schooling (e.g., Todd and Wolpin 2006, 2008; Keane and Wolpin 2010), the income effect could also lead to an increased demand

[14] Developing a theoretical model that contemplates all these potential interactions is beyond the scope of this paper. However, it is important to mention that Keane and Wolpin (2010), for instance, formalize a similar decision process, focusing exclusively on women's decisions. More specifically, they estimate a structural model in which women's choice set is comprised of work, marriage, schooling, fertility, and welfare participation. A very simple but illustrative example of the complexity of this setting is that women make between 18 and 36 mutually exclusive choices in each period, depending on their fecundity stage.

for children's education.[15] Income effects are crucial for poor households in the presence of credit constraints. In such settings, families might decide not to send their children to school because they cannot afford it. The cash transfer would work as a mechanism that relaxes those constraints, allowing beneficiary households to increase their expenditure on school-related goods and services. This would enable children to enroll and remain at school (e.g., books, clothing, transportation costs, etc.). A similar mechanism could also explain changes in fertility decisions if, for instance, there is a direct dis-utility associated with early life childbearing, and household members cannot buy contraceptives in the absence of cash transfers.[16]

CCTs can also affect household decisions through a substitution effect since they make participation contingent on specific behaviors, typically school enrollment, attendance, and health check-ups. Education requirements reduce the opportunity cost of schooling, and make it more attractive compared to any other non-education-related activity such as labor market participation or becoming a parent (e.g., Parker and Todd, 2017). A substitution effect could also affect children's education enrollment through parents' time allocation if children's engagement with the education system depends, at least partly, on the time they spend together (e.g., in the spirit of Martinelli and Parker, 2008). In this case, the reduction in parents' time allocated to labor market activities through the substitution effect would free time that could be re-directed toward time spent with children. A reduction in parents' time allocated to labor market activities also increases

[15]It is beyond the scope of this paper to discuss the non-pecuniary benefits of schooling or if it should be considered a (normal) consumption good. Oreopoulos (2011) and MacLeod and Urquiola (2019) provide in-depth reviews about the status of this discussion in the literature.

[16]One alternative way in which the cash transfer can affect fertility rates of young women through an income effect is when their labor market activities are associated with transactional sex activities (see Baird and Özler 2016; LoPiccalo et al. 2016 for a review of the relation between income and transactional sex)

children's share of supervised time. This reduces the possibility of engaging in risky behaviors that could lead to early-life pregnancies. In sum, both income and substitution effects are expected to reduce children's labor market participation, increase children's education enrollment and reduce young women's fertility when they receive the cash transfer.

Dynamic effects: The effects discussed so far correspond to a static model. When individuals make decisions that have consequences for multiple periods, the set of potential behavioral responses becomes broader and even more complex. One example is what Black et al. (2008) refers to as the "future human capital effect". Consider the income and substitution effects discussed in the previous paragraph as effects in the "current" or today's time. The reduction in the marginal cost of schooling increases current investment in education. However, additional education today also increases the opportunity cost of education tomorrow. The more schooling children accumulate, the higher the wage offers they receive. If there are diminishing marginal returns to schooling, there would be a point where the marginal cost of an additional year of schooling will be larger than the marginal benefit. This would lead some individuals to choose labor market participation instead of more schooling (Behrman et al., 2011).

A similar reasoning can be applied to fertility decisions. There is a strong link between expected future labor market income and fertility decisions. Models that aim to characterize early fertility decisions propose that young women compare the lifetime expected utility of having vs. not having a teen birth (e.g., Duncan and Hoffman, 1990; Wolfe et al., 2001). Because they reduce the marginal cost of schooling, CCTs increase the expected utility of delaying fertility through higher expected adult wages. This leads to more women deciding to delay fertility. While delaying fertility might seem relatively costless in the short run, it is also reasonable to expect these costs to increase in the long run, for instance, due to a reduced probability of having a successful healthy pregnancy or due to an increase in biological or psychological costs associated with later-life pregnancies (Schmidt et al., 2012; Gustafsson, 2001). Hence, at some point, even when the opportunity cost of having a child is large due to high wages, the marginal cost associated with keep delaying childbearing might be sufficiently high to more than offset the potential gains in earnings. Under such circumstances, it is reasonable to expect that initial negative effects on fertility might start to fade or even reverse in the long run. In this case, one should be cautious in how the early negative effects are interpreted. These could be more associated with delays rather than actual changes in fertility preferences. However, potential effects of CCTs on overall preferences for fertility cannot be ruled out ex-ante.

Another example of how current decisions might have strong implications on future choices stems from the models of skill formation in the presence of dynamic complementarities (Cunha and Heckman, 2007). In these models, today's education investments increase education returns in subsequent stages, which promote a more extended stay in the education system to the detriment of other activities, such as labor market participation and childbearing.

It is important to note that besides the direct effects of CCTs on education, fertility, and labor market decisions, these decisions might also have direct effects on each other. For instance, education could affect fertility decisions if there is a trade-off between quality and quantity of children (Becker and Lewis, 1973); if it improves current women's ability to predict better labor market outcomes associated with delaying childbearing (referred to as "current human capital effect" in Black et al., 2008); if it improves access to contraceptives and family planning and health care services which are critical determinants of fertility decisions (e.g., as in Kearney and Levine, 2009; Bailey, 2006; Lundberg and Plotnick, 1995); or by changing women empowerment, attitudes, and values toward maternity, just to name a few.[17,18] Fertility could also affect education decisions, for instance, through the effect of child care time on the marginal cost of school time (Klepinger et al., 1999). Similarly, education can affect labor market decisions by affecting children's perceptions about how the process of earning better wages works, the current sacrifices required for better future wages, by improving expectations about achievable goals, or by providing different role models, etc.

Other mechanisms: While in a friction-less model, conditionalities associated with cash transfers would cause efficiency losses, they usually aim to correct potential sub-optimal decisions due to market failures, such as information frictions, differences in discount rates, or intra-household bargaining problems (Parker and Todd, 2017; Baird et al., 2014). Hence, under more realistic circumstances, CCTs may also affect households' decisions through mechanisms other than the standard income and substitution effects. For instance, CCTs are usually entitled to the mother of the eligible children. Moving from a unitary to a collective household decision model (e.g., Chiappori, 1988, 1992; Browning and Chiappori, 1998) opens the door for CCTs to change household members' bargaining power (Martinelli and Parker, 2003, 2008; Attanasio et al., 2012), which could re-direct part of the house-

[17] Related literature (e.g., Black et al., 2008) also defines an "incarceration effect" of education on fertility, i.e., more time spent at school reduces the time available to engage in risky behavior. While this mechanism is plausible, in this discussion, it is captured by the idea that education and fertility are mutually exclusive or highly substitutes

[18] Alternatively, attending school might also increase the social interactions of young girls with other potential sex partners that they meet at school or in related environments. However, for this to have an effect, the new interactions should more than offset the existing interactions outside the school that are lost due to the increased time at education institutions.

hold expenditure toward goods and services that are more favorable to children (e.g. Thomas, 1990; Duflo, 2003, or more specifically about *PANES/AFAM-PE* Bergolo and Galván, 2018).

The information environment and expectations about returns to education are also key determinants of current education decisions (Jensen, 2010). For instance, by participating in a CCT, parents are more exposed to highly educated professionals, which could change their expectations about the opportunities for their children and the investment required to reach them. Parents' improved expectations can also be transmitted to their children. This would lead to higher enrollment and permanence in the education system (Attanasio and Kaufmann, 2014; Chiapa et al., 2012). On the contrary, children's expected returns to education can be negatively affected by the CCT if parents substantially increase their time allocated to leisure activities because of the income effect. Perceptions about expected future outcomes are also highly relevant for fertility decisions. Kearney and Levine (2014) propose a model where fertility decisions are determined by the perceived probability of achieving a high utility state, which is only feasible if women delay childbearing. Perceptions of the likelihood of success are a function of current socio-economic status and inequality. Hence, CCTs may also affect fertility decisions by changing the current socio-economic situation of poor women or, more generally, by reducing inequality in their society.

Finally, alternative mechanisms such as a reduction in household economic stress that could create a better environment for child development (Gershoff et al., 2007; Yeung et al., 2002; Conger et al., 1993); or improved children's health outcomes due to better parental socioeconomic conditions (e.g., Currie, 2009); or social interactions and peer effects (e.g. Bobonis and Finan, 2009; Lalive and Cattaneo, 2009) might also affect education, fertility, labor market participation decisions.

In sum, the related theoretical literature provides mostly unambiguous predictions about the short-run effects of CCTs on education, fertility, and labor market decisions for individuals that benefited from a CCT program when they were young. More specifically, CCTs are expected to reduce teenage pregnancies, increase education enrollment, and reduce children's labor market participation. However, in a dynamic setting, the expected effects are ambiguous and depend on individual preferences and institutional characteristics. The fact that these effects can interact in complex and theoretically ambiguous ways illustrates the need for a dynamic analysis to understand how CCTs affect the current and future lives of the beneficiaries and the mechanisms involved. In the end, the effects of CCTs on the trajectories that mark children's transition to adulthood and early adulthood outcomes is mostly an empirical matter.

2.4 Data Sources, Measurement, and Sample of Interest

The analysis of the effects of CCTs on the transition to adulthood is highly data demanding for two reasons. First, it requires information on a large number of individual characteristics. Because adulthood is defined not just by one but by a series of markers in different life spheres - including education, fertility, and labor market markers among the most important ones - the transition to it also needs to be characterized in terms of such dimensions. Second, because transitions are a dynamic phenomenon by nature, its analysis requires longitudinal information that allows for a complete description of the individual trajectories. Having both is extremely difficult and costly.

The data used in this paper satisfy these two requirements. First, the empirical analysis is based on an exhaustive compilation of administrative records from different sources for the universe of applicants to *PANES/AFAM-PE*. These can be linked at the individual level and contain information about fertility, education, and labor market outcomes. Second, because the data is based on administrative records, all of these variables are observed for a long span of years and for the universe of interest. In the next section, I explain in detail the main features of the dataset assembled for the analysis, as well as the key outcome variables.

2.4.1 Data Sources and Measurement

***PANES/AFAM-PE* records: Application and participation variables.** These records are used to measure all the application- and participation-related variables, which are mostly used as treatment or control variables. They were provided by the Ministry of Social Development, which is in charge of implementing the program, and contain information about the universe of successful and unsuccessful applications to *PANES/AFAM-PE* between April 2005 and December 2017 at the form, household, and individual level. The information at the form level includes city, date of application, poverty score, resolution, and in case of acceptance, the participation history. Information at the household level includes the house's building materials, structure, appliances, and access to public services, among other information used to compute the poverty score. Individual level information contains the baseline information about education, employment status, income, date of birth, and gender, for each household member reported in the application form. The total number of application forms included in the raw participation data is 747,204, corresponding to 1,476,696 unique individuals.[19]

[19] Online Appendix contains a more detailed description of the participation data.

Birth Records: Fertility outcomes I use birth records to measure the fertility outcomes reported throughout the paper. These were provided by the Ministry of Health and consist of an individual-level dataset that includes the universe of births in Uruguay between 2005 and 2019. Birth records contain information such as birth date, type of institution where the child was born (public, private, or others), the mother's age, birth weight, and gestation weeks. In addition, they also include identification information of the mother, which allows me to link this information with *PANES/AFAM-PE* participation records at the individual level.

Concerning fertility outcomes, it is important to note that these variables are defined exclusively for women due to the typical limitations in the information reported on birth certificates about newborns' fathers. As for every outcome variable described hereon, I define two types of variables. A binary variable that indicates whether a woman has given birth *at or before* a certain age and a continuous variable that reports the number of births by a given age. The binary variable is associated with extensive margin responses, i.e., it will capture the effect of *PANES/AFAM-PE* on giving birth versus not giving birth. The continuous variable will also capture responses in the intensive margin. In addition, I define different variables for each age between 15 and 30 years old. This allows me to provide a full description of whether and how the effects of *PANES/AFAM-PE* materialize throughout the transition to adulthood. All the outcome variables are defined exclusively based on the post-application period. In the specific case of fertility outcomes, I define the post-treatment period as starting seven months after the application date. As a robustness test, I will also report estimates based on a binary variable defined *at* a given age, as opposed to *at or before*.[20]

[20] Online Appendix provides summary statistics for each of the outcome variables based on

Secondary and tertiary education administrative records: Education outcomes. I use the secondary and tertiary education administrative records to measure the by-age effects on education enrollment. These records come from three different public institutions: 1) National Council of Secondary Education, 2) National Council of Technical and Professional Education, and 3) *Universidad de la Republica*, which is the largest public university in the country. Information from the National Council of Education corresponds to the traditional public secondary education system, which is analogous to grades 1-6 of middle and high school education in the United States. These records contain yearly information for the universe of students enrolled in secondary public schools in 2006-2012, 2014, 2017, and 2018. National Council of Technical and Professional Education records contain information on vocational and technical public school enrollment for the same period. Careers offered by technical and vocational schools can be classified into middle school, high school, and tertiary careers, based on enrollment requirements.[21] Finally, the information provided by *Universidad de la Republica* consists of individual-level information that identifies the year of enrollment of every student enrolled at the University between 2005 and 2020.[22]

To analyze the effects of *PANES/AFAM-PE* on education decisions, I focus

---

the samples that are defined in Section 2.4.2

[21]For instance, a middle-school-analogous vocational education program is a program that requires individuals to have completed primary school. A high-school-analogous vocational educational program is a program that requires to have completed middle school and so on.

[22]While I do not have access to education enrollment information in private institutions for any of the education systems or levels, it is important to note that the (free) public education system is probably the relevant choice set of schools for the population of interest, given that private institutions usually offer a limited number of grants and have relatively expensive tuition. For instance Ramírez Leira (2021) shows that the probability of enrolling in a public institution for individuals in the first quintile of the income distribution is larger than 95% in 2017

specifically on enrollment outcomes. These variables are the most complete and reliable among the ones included in the administrative records. They are also easily comparable across levels and institutions.[23] I define education enrollment outcomes using binary and continuous variables. For secondary enrollment, The binary variable indicates whether an individual was enrolled at any secondary education institution, either traditional or vocational, at or before a given age. The continuous variable measures the total years of enrollment. In both cases, different variables are computed for ages 12-23. Enrollment in tertiary education works differently than enrollment at secondary education institutions. In particular, once registered for the first time, students are not required to re-enroll periodically to take classes. For this reason, the information provided by *Universidad de la Republica* only allows me to define a binary variable that indicates if the individual has ever been enrolled at the university or any other tertiary level course at the vocational/technical institutions, by a certain age. Tertiary education outcomes are computed between 18-23 [24]

Social Security Agency (SSA) labor histories: Labor market outcomes. I use SSA labor histories to construct the labor market outcomes. They contain monthly individual-level information about wages, hours worked, activity type, and employers' industry sector for each position for the universe of registered employees. The main limitation of SSA labor histories is that they provide infor-

[23]It is important to note that students who promoted the current grade are automatically enrolled for the next academic year. Hence, enrollment variables do not necessarily represent an explicit decision to sign up for the current academic year. Moreover, to some extent, individual enrollment for a given grade could be interpreted as a signal of academic progress.

[24]95% of the individuals in the sample who were ever enrolled in tertiary education did it before the age of 23. Because the share of individuals ever enrolled in tertiary education is already extremely low, first-time enrollment between 23-30 only corresponds to a handful of cases. Hence, for simplicity, I report estimates for tertiary enrollment only until age 23.

mation only on the formal labor market. However, it is important to note that the informal labor sector in Uruguay is currently relatively small compared to other non-high-income countries, and it only represents 17% of total employment. In any case, to interpret the results reported in this paper, one should keep in mind that someone that does not show up in SSA labor histories can be either unemployed or employed in the informal sector. Therefore, results must be interpreted exclusively as concerning the formal employment sector.

To analyze the effect of *PANES/AFAM-PE* on labor market outcomes, the baseline binary labor market outcome indicates whether someone has had a registered employment spell that lasted at least four consecutive months at or before a given age. Using spells of four consecutive months rules out potential temporary employment, such as Summer jobs, and reflects more stable links to the labor market. In addition, I define two complementary continuous variables. First, the total number of months an individual has worked in the formal sector by a certain age. Second, a similar variable that measures cumulative earnings, i.e., the sum of all labor income earned by a given age. These three variables are calculated for ages between 14 and 30 years old. As in the other dimensions, I report robustness tests using the same variables defined *at* a given age, instead of *at or before.*

2.4.2 Sample of Interest: Definition

Since this paper focuses on the effects of *PANES/AFAM-PE* on the transition to adulthood for children that benefited from the program during their childhood, the empirical analysis is restricted to individuals of households that applied to the program when they were eighteen years old or younger and had at least fifteen years old by April 2018. The latter restriction ensures that every individual included in the analysis has had the chance to enroll in high secondary education for at least one year. In addition, the analysis is always restricted to individuals of households who applied for the first time between 2005-2012, which represent 95% of the sample.

A dynamic analysis of the age-by-age effects of the program on transitions to adulthood presents some challenges associated with the definition of the sample of interest, especially for a permanent and relatively young program such as PANES/AFAM-PE. For instance, it is impossible to calculate the effect of the program at age 30 for someone that has not turned that age by the time the outcomes are last observed.[25] The simplest alternative would be to keep the sample composition constant and only use individuals who have already turned 30 on a given date. However, this would substantially reduce the sample size and compromise estimates' statistical precision. To balance the trade-off between keeping the sample composition constant and maximizing the use of information available, I present two main sets of results. First, a group of estimates based on a *main sample* comprised of 23 years old or older individuals in December 2019. Second, for the dynamic analysis and estimates at older ages, I use a series *dynamic samples* that aim to include as many observations as possible. These *dynamic samples* vary their composition depending on the age at which an outcome is measured but use the maximum amount of information available.[26] However, for the sake of transparency, I will also report the most conservative estimates based on an extreme-balanced sample of individuals who were 30 years old by December 2019. Estimates based on this fully-balanced sample allow us to completely rule out

[25] Online Appendix reports the distribution of ages on December 31, 2019. This corresponds to the last day available in birth records.

[26] For instance, estimates of the effect of the program at age 27 will be based on individuals who had already turned 27 by the time an outcome is last measured, which depends on the outcome.

that the results are driven by changes in the composition, although at the cost of statistical precision.

2.4.3 Sample of Interest: Description

Table 2.1 describes the main characteristics for the *main* (columns 1 and 2) and the *dynamic* (columns 3 through 6) samples.[27] Columns 3 and 4 describe the sample of analysis used for fertility estimates at age 30, while columns 4 and 5 do the same for the sample used for labor market outcomes. These are two extreme examples of the *dynamic samples* used in the longer run analysis. Odd columns include all individuals in each of the samples. Even columns are restricted to individuals with an application score within the optimal RDD bandwidth chosen for the baseline estimates. The procedure used to select the optimal bandwidth is explained in detail in Section 2.5.[28]

Panel a. focuses on individual characteristics. There are 224,413 individuals in the *main sample* who are equally split between men and women, were on average 26.9 years old by December 31, 2019, and belonged to about 1.8 households. Individuals are typically included in 2.6 application forms. The average age at first application is 13.4 years old. About 84.2% of these individuals were accepted to *PANES/AFAM-PE* at least once before age 18.[29] 78.4% of the individuals show up in at least one application form to the *PANES* phase. At the same time, 96.1% are included in at least one *AFAM-PE* form. Panel b. describes the characteris-

[27]Online Appendix report more detailed descriptive statistics at the form level, including information about the 747,204 application forms filled between 2005 and 2017

[28]For exposition purposes, to describe the samples used in the analysis, I selected the largest optimal bandwidth among the estimates that use each specific sample.

[29]Online Appendix provides the full distribution of age at first application and age at first acceptance for both *main* and *dynamic* samples.

tics of the application forms for these individuals. Because they may be included in multiple application forms, this table reports the characteristics of the earliest application form filled.[30,31] The average centered poverty score is 0.18. This means that the average application corresponds to an eligible form. Consistently, the share of individuals whose first application form was accepted is 71.8%. The first application form for most individuals (78.4%) was filled during the *PANES* phase, and 31% of the applications corresponded to individuals of households in the capital city (Montevideo). This means that individuals from the capital city are under-represented in this sample since about half of Uruguay's population lives in Montevideo. Finally, panel c. describes the household characteristics. Individuals in the sample belong to households that are comprised, on average, of 4.9 individuals, of whom 2.9 are underage children. Slightly less than half of the households correspond to single-parent households, and the average age of household members (including children) is 23.1. Household heads have, on average low education levels, i.e., about seven years, slightly more than the equivalent of completed primary school. 63.4% of them are employed, and their average income is USD 144.33, which is comparable in size to the cash transfer value, as described in Section 3.2.

The sub-group of individuals that belong to the *main sample* and have a poverty score within the optimal bandwidth has very similar characteristics to the full *main sample* in terms of variables that are not related to the poverty score. These are different by construction. Besides these variables, the exception is on the share of individuals from the capital city, which is smaller than in the

[30]A more detailed discussion of the reasons for selecting this application form is provided in Section 2.5, since this is also a critical decision for the empirical design

[31]Online Appendix compares the characteristics of *all* application forms versus *first* application forms.

full *main sample*. The *dynamic samples* also closely resemble the *main sample* except for age-related variables, which are mechanically different. Individuals in the *dynamic samples* are, on average, between 3 and 4 years older than in the main sample (31.1 and 29.97 compared to 26.91), and were also older when they applied to the program for the first time (16.8 and 15.98 years old compared to 13.4). Because of the age restriction used to define these *dynamic samples*, most of the first application forms for these individuals corresponded to *PANES* applications. This also implies that they had less potential time of exposure before turning 18 years old and translates into a lower share of individuals ever accepted to *PANES/AFAM-PE* (71.0% and 80.9% compared to 84.2%). Except for the mechanical differences in variables related to the definition of each sample, the subgroup of individuals that comprise the *dynamic samples* is very similar to the full *main sample*.

2.5 Empirical Strategy

As described in Section 3.2, eligibility to participate in *PANES/AFAM-PE* is based on a poverty score. More specifically, let z be the poverty score centered around the eligibility threshold and D an indicator variable such that positive values of z indicate eligibility (i.e., $D = 1$) and negative values indicate ineligibility (i.e., $D = 0$). The use of an arbitrary threshold to define whether a household is eligible to participate in *PANES/AFAM-PE* provides a quasi-random source of exogenous variation to identify the causal effects of the program using a Regression Discontinuity Design (RDD) (Thistlethwaite and Campbell, 1960). Intuitively, under perfect compliance and a continuity assumption, (local) average treatment effects of the program can be obtained by comparing the regression functions of the outcome of interest at both sides of the threshold (Hahn et al., 2001).[32]

To illustrate how the *PANES/AFAM-PE* eligibility rule works, Figure 2.2 describes the relation between a variable that indicates if an application was successful (*y-axis*) and the centered poverty score (*x-axis*). Panel a. depicts this relation for the full support of the running variable. Each bin in the figure represents the percent of accepted forms within the bin.[33] In the background, vertical bars represent the distribution of the poverty score. Panel b. zooms into observations close to the threshold, i.e., within five percentage points distance, grouped into 10 bins of half percentage point width. From the program administrator's perspective, the eligibility rule was applied correctly, although not perfectly. Figure 2.2 shows a pronounced change in the acceptance rate just at 0, i.e., at the eligibility threshold. The size of the change is 60.0p.p., and it is statistically significant at usual levels ($p - value < 0.001$).[34] Different reasons can explain the fuzziness observed on both sides of the threshold. For instance, to the left of 0, it could be due to applications below the eligibility threshold that were rejected when filed but were automatically enrolled after the threshold became more lenient.[35] On the right-hand side, it could be due to rejections based on reasons other than the

[32]Formally, let Y be any of the outcomes of interest. Under perfect compliance, the key identification assumption in RDD is that Y is continuous at $z = 0$ if the regression functions for the outcome variable - $\mathbb{E}[Y(1)|Z = z]$ and $\mathbb{E}[Y(0)|Z = z]$ - are continuous functions at $z = 0$, then: $\mathbb{E}[Y(1) - Y(0)|Z = z] = \lim_{z \downarrow 0} \mathbb{E}[Y|Z = z] - \lim_{z \uparrow 0} \mathbb{E}[Y|Z = z]$.

[33]To the left of 0, observations are binned in ten quantile-spaced bins. To the right of 0, observations are binned in fifty quantile-spaced bins. The relation in the number of bins used at each side of 0 is based on the relative number of observations between the two sides.

[34]The procedure used to calculate the change in the probability of acceptance is based on the data-driven approach proposed by Calonico et al. (2019). Hence, the optimal bandwidth is selected such that it optimizes the Mean Squared Error (MSERD). This will be explained in detail before the end of this section.

[35]Unfortunately, the administrative records do not identify these cases, and date of application corresponds to the day on which the application was submitted.

poverty score, such as income or no qualifying underage children.

Despite this sharp change, the use of an RDD in the *PANES/AFAM-PE* setting presents one additional challenge. As discussed in Section 3.2, *PANES/AFAM-PE* has been in place uninterruptedly since 2005. This means that households might have applied to the program multiple times, introducing two concerns about key elements of the research design. The first one is how to define the running variable when households have multiple application scores. The second is how to address the possibility of endogenous sorting around the eligibility threshold induced by re-applications. This could happen if re-applicant households that are close to the threshold and were originally rejected are different from non-reapplicant households, and these differences are correlated to the outcomes of interest. In this case, the RDD estimates will be biased.[36] To address these concerns, I follow the approach proposed by Jepsen et al. (2016) who suggest implementing a fuzzy RDD where eligibility based on the first application score is used as an instrument for treatment in contexts where there are re-applications. The intuition is that the first score is presumably the score that is less subject to manipulation.[37] Hence, the RDD will be based on the following variables:

Exogenous variable: eligibility based on the score of the first application form (D^{1st}): I define the first application form (or reference form) as the earliest

[36] Endogenous sorting in settings with multiple applications is also an issue in different contexts such as close elections (e.g. Cellini et al., 2010), analysis of returns to education using test scores (e.g. Clark and Martorell, 2014); or evaluation of the effects of remedial education(e.g., Martorell and McFarlin, 2011).

[37] Jepsen et al. (2016) analyze the effects of GED scores on employment and earnings. In this setting, the discontinuity exploited is the passing grade of the exam, and concerns about endogenous sorting arise because students can take the exam multiple times. The issue for identification is that re-takers can be different from non-re-takers in ways that are also correlated with the outcome of interest. If this is the case, using the final score obtained in the GED exam will not provide an adequate source of identification for the effects of the GED.

application form by any of the households that an individual has ever belonged to, as long as the individual had not left the household by the time of application. By going as far back as possible when defining the value of the running variable, I am taking a conservative approach to minimize any possible concern about endogenous sorting.[38] Hence, eligibility based on the first application is a binary variable that takes the value of 1 if the score obtained in the first application corresponds to an eligible form and 0 otherwise.

Endogenous variables: participation in *PANES/AFAM-PE* (T): The baseline treatment variable (T) is a binary variable that indicates whether an individual was ever accepted to *PANES/AFAM-PE* before turning eighteen years old. In addition, I define analogous variables for ages twelve through seventeen that will be used in estimates where the outcome is measured before 18 years old. As a robustness test, I present estimates based on two complementary continuous treatment variables: 1) the number of months treated and 2) the net present value of the cash transfer collected by the household. Hence, the analysis of the causal effects of *PANES/AFAM-PE* on the different outcomes of interest is based on the following specification:

$$Y_i = \mu + \tau T_i + \beta_1 Z_i^{1st} + \beta_2 Z_i^{1st} T_i + u_i \tag{2.1}$$

where Y_i is the outcome of interest for individual i, (Z_i^{1st}) is the score obtained in the first application, and T_i corresponds to i's treatment status. Because T_i and Y_i are endogenous, T_i is instrumented using D_i^{1st} based on the following first-stage

[38] For instance, household h_1 applied to *PANES/AFAM-PE* with forms $f_{h_1,A}$ and $f_{h_1,B}$. Individual i was born in h_1 after $f_{h_1,A}$ was filed, but before $f_{h_1,B}$ was filed. In this case, $f_{h_1,A}$ is still the reference form for individual i, even when she was not included in $f_{h_1,A}$.

equation:

$$T_i = \alpha + \delta D_i^{1st} + \gamma_1 Z_i^{1st} + \gamma_2 Z_i^{1st} D_i^{1st} + \epsilon_i \tag{2.2}$$

Following Imbens and Lemieux (2008) and Calonico et al. (2014), I estimate this model using local linear regressions fitted separately to each side of the threshold with observations that are sufficiently close to it. The estimation procedure follows Calonico et al. (2014), who provide robust standard errors and confidence intervals. The threshold is defined optimally following the data-driven approach by Calonico et al. (2019) and the default options: selection of bandwidth by optimization of Mean Squared Error (MSERD) and a triangular kernel function that puts more weight on observations that are close to the threshold. To assess the robustness of the results to these arbitrary choices, I present specification curves based on all possible combinations of options for each baseline outcome. In all cases, standard errors are clustered at the household level.

The described strategy provides an estimate that should be interpreted as a local average treatment effect. In addition, I will also report the intention to treat effects. The difference between the ITT and LATE effects is that the latter scales up the ITT effect by the size of the first stage, i.e., by the actual change in the probability of participation at the eligibility threshold. ITT estimates are based on the following specification:

$$Y_i = \tilde{\mu} + \tilde{\tau}\mathbb{1}(Z_i^{1st} > 0) + \tilde{\beta}_1 Z_i^{1st} + \tilde{\beta}_2 Z_i^{1st}\mathbb{1}(Z_i^{1st} > 0) + u_i \tag{2.3}$$

Equation 2.3 is the reduced form specification for equation 2.1, but using an indicator variable for eligibility ($\mathbb{1}(Z_i^{1st} > 0)$) instead of the treatment binary variable (T). The coefficient of interest is $\tilde{\tau}$, the ITT effect, which measures the difference in the intercepts of the two local linear regressions fitted separately to each side of the eligibility threshold within the optimal bandwidth. ITT estimates are important to provide an idea of the magnitude and standard errors associated with the visual representations of the RDD.

Finally, it is important to note that, compared to sharp RDDs, fuzzy RDDs require an additional identifying assumption of monotonicity or "no defiers" (Imbens and Lemieux, 2008; Cattaneo et al., 2019). In this paper, monotonicity implies that an application form with a score z that is rejected when the threshold is set at 0 would also be dismissed for any alternative threshold greater than 0. Conversely, any application form with a score z that is accepted when the cutoff is 0, would also be accepted if the cutoff is $\tilde{z} < 0$.

2.6 Results

In this section, I present the main empirical analysis. First, I illustrate the validity of the RDD by reporting first-stage results, manipulation, and balance tests used typically in these settings. Second, I report the reduced form estimates (ITT), including visual evidence, and the LATE effects for a group of baseline results measured at ages 18, 23, and 30. For fertility and labor market outcomes, I report estimates for each of these three ages. For education outcomes, I focus exclusively on outcomes measured up to the age of 23. This is because after 23 years old, there are almost no changes in education enrollment variables. Education variables at age 18 correspond to secondary education, while at 23 correspond to tertiary education. Traditional and vocational/technical school enrollment variables are pooled in both cases. Third, I report analogous estimates focusing on heterogeneity by sex. Finally, I report the full dynamic analysis measuring the effects of *PANES/AFAM-PE* at all possible ages.

It is important to note that the choice of the specific age cutoffs used for the baseline results is arbitrary and mainly for illustration purposes. Estimates on outcomes measured at 18 and 23, based on the *main sample*, are chosen to coincide with the age at which someone on track would complete secondary and tertiary education, respectively. Estimates on outcomes measured at age 30, based on the *dynamic sample*, are the last ones to provide a reasonable sample size to implement an RDD. This way of presenting the results is also illustrative of the two typical snapshots of early- and later-life results (18, and 30 years old, respectively) that can be found in the literature. In any case, the full set of results is discussed in the dynamic analysis, which provides the full description of the trajectory of the effects.

2.6.1 Validity of the RDD Design

In this section, I report evidence that supports the use of an RDD to analyze the causal effects of *PANES/AFAMPE*. First, Figure 2.3 depicts the relation between the score obtained by an individual in her first application form (Z^{1st}) - measured in the *x-axis* - and *PANES/AFAM-PE* participation before eighteen years old (T) - measured in the *y-axis*. Panel a. reports the relation for the full support of Z^{1st}, while panel b. zooms into a narrower bandwidth of 5p.p.. In all cases, as in Figure 2.2, each circle represents the average value of the treatment variable within a 0.5p.p. width bin.

Overall, Figure 2.3 shows an abrupt discontinuity in the probability of ever being accepted into the program before turning eighteen years old, just at the eligibility threshold. This probability changes by 50% (29.3p.p.), and the change is statistically significant at traditional levels ($p-value \leq 0.001$). Table 2.2 presents the analogous regression estimates. Column (1) reports the baseline estimates using a linear polynomial function and a triangular kernel function, while columns (2) through (4) show that the baseline specification is robust to changes in the polynomial degree and kernel function.[39]

In addition, Figure 2.4 and Table 2.3 report some of the typical tests performed when using RDD to validate the identification assumption of continuity. Panel a. in Figure 2.4 illustrates that the distribution of the poverty score is smooth around the threshold. Panel b. provides a formal test of continuity of the running variable based on Cattaneo et al. (2018) and McCrary (2008). This test provides no evidence to reject the null hypothesis of continuity ($p-$value=0.715). Table 2.3 reports the RDD analogous to a balance table comprised of a series of falsification tests that replicate the baseline RDD strategy on pre-treatment covariates. As expected, Table 2.3 shows that baseline variables are continuous at the threshold. When p-values are adjusted by the expected false discovery rate (Anderson, 2008), all estimates are statistically insignificant.[40] Furthermore, when the falsification test is conducted on a variable that predicts the eligibility status based on all the other baseline covariates, there are no signs of discontinuity at the threshold (p-value = 0.635). This indicates that observable characteristics do not change abruptly at the threshold.[41] Overall, the tests reported in this section provide robust support for the validity of the identification strategy. However, out of

[39] Online Appendix reports several additional robustness tests, such as using alternative endogenous variables, estimates on the dynamic samples, and falsification tests.

[40] When taken individually, in some cases there are statistically significant differences, but in all cases, these are economically irrelevant. For instance, the average age of household members for eligible individuals in the *main sample* is 0.31 years higher compared to ineligible individuals. Eligible individuals also live in households with, on average, 0.07 fewer members and are also 0.13 years younger by December 2019.

[41] Online Appendix reports visual evidence about the continuity of the predicted eligibility status and similar estimates for the *dynamic samples*. Both pieces of evidence also support the baseline variables' continuity at the threshold.

caution, the empirical analysis will be complemented with several tests to further prove their robustness.

2.6.2 Baseline Estimates

Figures 2.5 and 2.6 depict the corresponding binary or continuous outcome variable as a function of the score obtained in the first application. For an easier comparison across outcomes, visual evidence is reported for a bandwidth of $\pm$ 5p.p.. Observations at each side of the threshold are grouped into bins of 0.5p.p. Table 2.4 reports the analogous regression estimates for a bandwidth chosen by optimizing MSERD (Section 2.5 described this in detail). For consistency, and to provide the most transparent representation possible, Table 2.4 and Figures 2.5 and 2.6 are based on specifications without additional covariates. Table 2.5 reports the baseline LATE results.

Fertility outcomes

Panel a1 in Figures 2.5 and 2.6 already show signs of a discontinuity at 0 both for the probability of having a birth before the age of 18, and the number of births by the same age. The fact that the visual evidence already shows clear signs of a discontinuity, even when a large share of ineligible individuals - based on the first score - has participated in the program, is suggestive of the strength of the effects. Table 2.4 reports formal estimates of the size of these changes. When measured at age 18, eligibility to participate in *PANES/AFAM-PE* has an ITT effect on the probability of giving birth of -3.1p.p. This effect represents 13.5% of the number of women in the ineligible group within the optimal bandwidth who gave birth by age 18, and is statistically significant at traditional levels (p-value = 0.005). A similar effect is observed in panel b. for the number of births, which decreased by

0.038 (or 14.9% with a p-value = 0.003). LATE estimates reported in Table 2.5, which re-scale the ITT effects by the size of the first stage, show effects of -9.4p.p. (-41.2%, p-value = 0.015), and -0.108 (-41.9%, p-value=0.015), respectively.

Both the visual and the econometric evidence suggest a decay in the effects when measured at age 23. For instance, the estimated LATE effect in the probability of having a birth before the age of 23 is -0.047 (-8.52%) and statistically insignificant. The effect on the number of births remains statistically significant at a 10% level, but is less than half of the effect estimated for age 18 (-41.9% vs. -16.86%). A more drastic attenuation, or even reversal, is observed when the outcome is measured at age 30. In this case, both the binary and continuous variables seem to have been unaffected by the program. Furthermore, in the case of the probability of having at least one birth before the age of 30, the effect has even changed its sign, although it is statistically insignificant (p-value = 0.288). Overall, the results suggest that *PANES/AFAM-PE* negatively affects fertility at early ages, but these effects attenuate at age 23 and even reverse by age 30. I will go back to this discussion in Section 2.6.4 when I report the age-by-age results for the whole period covered by the data.

The results reported on the baseline set of fertility outcomes indicate a strong and negative effect of *PANES/AFAM-PE* on the probability of having a teenage pregnancy. This effect is statistically significant and economically relevant. For instance, a 41.9% reduction in the number of births by age 18 is equivalent in percentage terms to the reduction observed in Uruguay's adolescent fertility rate between 1960 and 2020, which changed from 5% to 3% in the period. Compared to other policy interventions carried out in Uruguay, the effects of *PANES/AFAM-PE* are substantially larger than, for instance, legalization of abortions (Cabella and Velázquez, 2022), or a large-scale intervention that granted access to subdermal contraceptive implants (Ceni et al., 2021). The effect is also consistent with very recent empirical evidence from other programs in high-, middle-, and low-income countries. For instance, in the US, Michelmore and Lopoo (2021) find that additional exposure to the EITC during childhood leads to a 2%–3% decline in a woman's likelihood of having a first birth by her early 20s. Perhaps in a more similar context, Attanasio et al. (2021) find remarkably similar estimates of the effects of an expansion of *Familias en Accion* on teenage pregnancies measured at age 18 of -9.3p.p., while Barham et al. (2018) find that a CCT in Nicaragua reduced the number of women's births at ages 18-21. A qualitatively similar result is observed for a temporary cash transfer implemented in rural Malawi, although, in this case, the effects were observed for an unconditional type of transfer (Baird et al., 2011).

Education outcomes

The effects on education outcomes are more nuanced. In the case of secondary education, measured at age 18, the visual evidence reported in panels b1. of Figures 2.5 and 2.6 shows mixed evidence. First, there is no sign of a discontinuity when using the binary variable as the outcome variable. This is confirmed by the econometric evidence reported in Tables 2.4 and 2.5 that show ITT and LATE effects of 0.008 (1.06%) and 0.017 (2.21%), respectively, both statistically insignificant (p-value=0.204 and p-value=0.469). On the contrary, when looking at the number of years enrolled in secondary education, both the visual and the econometric evidence suggest a positive effect, with an ITT effect of 0.086 years (3.41%) and a LATE effect of 0.253 (10.05%), both statistically significant (p-value=0.029 and p-value=0.027).

The program does not seem to affect individuals' enrollment in tertiary education, measured at age 23. The ITT estimates show an increase of 0.5p.p. (5.83%) on the probability of enrollment, but imprecisely estimated and statistically insignificant (p-value=0.445). A similar null effect is observed for LATE estimates (p-value=0.526). Unfortunately, the data available does not allow me to analyze any measure of academic progress in tertiary education.

The more nuanced evidence on the effects of cash transfers on secondary education enrollment outcomes is consistent with findings in related literature. On the one hand, the moderate increase of about a quarter of a year in secondary education enrollment is similar to previous findings that show that conditional cash transfers programs improve years of education between 0.2-0.4 years (Araujo and Macours, 2021; Behrman et al., 2011; Barham et al., 2018). In a different setting, but also focusing on the role of a cash transfer program, Aizer et al. (2016) find that the Mothers' Pension program in the US increased children's years of education by 0.3-0.4 years. Additional literature provides evidence of both stronger and weaker effects. For instance, Attanasio et al. (2021) report that *Familias en Accion* strongly reduced dropouts by 18p.p.; Molina Millán et al. (2020) find that a CCT program in Honduras led to a large increase in secondary completion rates; and Cahyadi et al. (2020) find an increase of 29% in high school completion for an Indonesian cash transfer program. On the contrary, some additional evidence suggests null (Dustan, 2020), or even negative (Bastian et al., 2022) effects of cash transfers or expanded access to welfare.[42] Focusing on tertiary education outcomes, the null effects of *PANES/AFAM-PE* are in contrast with Molina Millán et al. (2020) who find a strong increase in the probability of reaching university, or

[42]For instance, Dustan (2020) finds null effects of a CCT implemented in Mexico City that paid students to be enrolled in a public high school. Bastian et al. (2022) find evidence that an expansion of the safety-net reform in the US might have reduced educational attainment for women and had small positive effects on men.

with Attanasio et al. (2021) who find an increase in tertiary education enrollment for men. Similarly, some evidence about the effects of social safety net policies in the US point in the same direction and find positive effects of increased income on college enrollment (Bastian and Michelmore, 2018; Manoli and Turner, 2018).

Together, the results reported in this section indicate that the mixed evidence for education outcomes is explained by the margin of response considered. In particular, behavioral responses seem to be associated with the intensive margin, i.e., an increase in the number of years enrolled, rather than by changes in the extensive margin, i.e., the probability of being enrolled. Howecver, the comparison in the previous paragraph must be taken with a grain of salt since the estimates reported so far only inform about enrollment. Additional evidence reported in Section 2.6.4 provides preliminary evidence of the effects on *PANES/AFAM-PE* academic progress.

Labor market outcomes

Effects on labor market outcomes measured at age 18 are null. First, the visual evidence reported in panels c1 of Figures 2.5 and 2.6 does not suggest any evidence of discontinuity at the threshold neither for the binary or the continuous variable. The regression estimates for ITT and LATE both indicate similar patterns, with the effects being small and statistically insignificant.

When looking at age 23, the picture is extremely different. The visual evidence reported in panel c2 depicts a sizable jump in both outcome variables. The regression estimates show an ITT effect on the probability of having had at least one spell of four consecutive months in the formal labor market of 2.0p.p. (3.85%) and of 0.816 (3.65%) on the number of months worked. Both effects are statistically significant (p-value=0.022 and p-value = 0.045). Similarly, the estimated

LATE effects are 6.4p.p. (9.69%) for the binary variable (p-value=0.062) and 4.4 months (19.77%) for the continuous variable (p-value=0.005). Effects of this size are not rare for this program. For instance, Bergolo and Cruces (2021) find that *PANES/AFAM-PE* had a similarly-sized effect but with the opposite sign on parents' adult formal labor market participation. It is important to note that this effect is explained by the income threshold used to define eligibility for the program, which mostly affects adults' labor market decisions.

As in the case of women's fertility, by age 30, both the ITT and LATE effects on labor market outcomes seem to have attenuated. For instance, the estimated LATE effect on the extensive margin of participation is 1.6p.p. (1.98%) and statistically insignificant ($p - value = 0.570$), and 1.259 months (2.31%) for the number of months worked, also statistically insignificant (p-value=0.324).

Positive effects on labor market outcomes have been found in recent literature for other cash transfer programs. For instance, Barham et al. (2018) find that a CCT program in Nicaragua increases employment and earnings at the ages 19-22. However, the attenuation of the effects by the age of 30 contrasts with Araujo and Macours (2021) who find that, by this same age, PROGRESA improved earnings by 16%. For the US, most of the evidence also suggests a positive effect of increased exposure to social safety net policies (Barr et al., 2022; Bailey et al., 2020; Bastian and Michelmore, 2018; Aizer et al., 2016), although the only purely experimental piece of evidence suggests that SIME/DIME had null effects on children's labor market outcomes (Price and Song, 2018).

The set of results reported so far illustrates the overall effects of the program on fertility, education, and labor market decisions in the way that snapshots would do. Taken individually, the effects are consistent with most of the literature but fail to explain how *PANES/AFAM-PE* has affected the full transition to adulthood.

The fact that the effects are substantially different depending on the outcome and the age at which they are estimated highlights the need to analyze the trajectories more in detail.

Robustness Tests

Online Appendix report additional robustness and sensitivity tests to validate the ITT and LATE results reported in the previous sections:

Randomization Inference: First, in the spirit of randomization inference, I replicate the baseline ITT estimates using different placebo cutoffs. More specifically, I iterate the baseline ITT specification using every possible cutoff in the range [-0.08,0.50] in steps of 0.0025, excluding values close to the actual threshold, i.e., between -0.01 and 0.01. These tests show that estimates that are statistically significant using the true cutoff fall in the extremes of the distribution of the placebo estimates. A similar pattern is obtained when looking at the sorted p-values. Furthermore, as expected, all the distributions of the placebo estimates are centered around 0 and have averages that are very close to 0.

Specification curves: Second, to rule out that the effects are driven by specific choices of the RDD parameters, I report specification curves for each estimate included in Tables 4 and 5. More specifically, I plot the point estimate and 90% confidence intervals for all possible combinations of choices of 1) criteria used to define optimal bandwidth, 2) kernel functions, 3) polynomial degree, and 4) use of covariates, sorted by point estimate. Overall, the specification curves illustrate that the size and direction of the effect are not driven by a specific choice of one of these parameters. Furthermore, the baseline estimates are usually close to the median estimates and, if anything, err toward null effects.

Inclusion/exclusion of covariates Third, I test whether ITT estimates are robust to the inclusion/exclusion of additional baseline variables as control variables. Both for ITT and LATE estimates, the inclusion/exclusion of covariates provides estimates that closely resemble the baseline specifications in magnitude, size, and statistical significance.

Balanced sample: Fourth, I report the baseline estimates but using a fully balanced sample instead of the *main sample.* The fully balanced sample is comprised exclusively of individuals who were 30 or older in December 2019. Hence, the sample composition is held constant for every estimate reported in these tables. ITT estimates are very similar in direction and slightly stronger in size. Furthermore, the balanced sample shows a weak positive ITT effect on secondary enrollment by age 18, even when using the binary outcome variable. However, in most cases, the effects are more imprecisely estimated because of the substantial reduction in the sample size. LATE estimates based on the balanced sample are also very similar to the baseline estimates based on the *main sample*, although with some differences in the magnitude and, in some cases, in the statistical significance of the effects due to the reduced sample size. The more pronounced difference is observed in the years of secondary education enrollment, which is substantially smaller and statistically insignificant, contributing to the nuanced pattern of effects in this dimension

Alternative endogenous variables: Finally, I replicate the baseline LATE analysis but use alternative definitions of the endogenous treatment variable. First, I substitute the binary treatment variable for a continuous variable that indicates the number of years in the program before turning 18. Estimates based on this alternative definition are almost identical in direction, statistical significance, and size when scaled up by the average value of the treatment variable. The same is true for estimates based on a continuous variable that measures the net present value of the total cash transfer amount collected by the household before the individual turns eighteen years old.

2.6.3 Heterogeneous Responses by Sex

In this section, I replicate the baseline estimates but split the sample by gender. Tables 2.6 and 2.7 report the LATE estimates for men and women respectively, while Figure 2.7 summarizes these results and reports the p-value of a test of difference of coefficients between these two groups. Panel a. reports estimates for the binary outcome variables, while panel b. reports the estimates associated with the continuous variables. To make the comparison easier, the figure reports standardized effects.[43]

First, the estimated effects on the probability of ever being enrolled in secondary education are small (4.55% and 0.65%) and statistically insignificant for both groups (p-value = 0.341 and p-value = 0.807). The effects on the number of years enrolled are also very similar (10.89% and 8.60%) but only statistically significant for men. In both cases, the differences between men and women are statistically insignificant (p-value = 0.549 and p-value = 0.767, respectively). The same is observed for tertiary education enrollment, where differences between men and women are not significant either (p-value = 0.392). The existing evidence on the effects of cash transfers on education outcomes does not provide a clear

[43]Because fertility variables are only measured for women, the estimates reported in Figure 2.7 are the same as in Table 2.4, except for the standardization.

pattern of heterogeneous effects by gender either. For instance, while Araujo and Macours (2021) find that educational gains from PROGRESA are slightly larger for women, Parker and Vogl (2018) find stronger effects on college enrollment for men, consistent with the evidence reported for Colombia in Attanasio et al. (2021). On the other hand, for a CCT in Honduras, Molina Millán et al. (2020) find similar effects for men and women.

Unlike education outcomes, the effects of *PANES/AFAM-PE* on labor market outcomes present a clear and strong differential pattern between men and women. Measured at age 18, the LATE effect on having worked four consecutive months is 2.1p.p. (9.7%) for women, but -0.5p.p. for men (-2.1%). However, even in spite of the different signs of the effect, both coefficients taken individually are statistically insignificant, and so is the difference between them (p-value = 0.710). Stronger differences are observed when comparing the effects on the number of months worked (-0.253 vs. 0.455). However, it still cannot be ruled out that both coefficients are the same ($p - value = 0.344$). When measured at age 23, the differential effects become larger. When looking at the binary outcomes, the effect of *PANES/AFAM-PE* on women is 11.2p.p. (17.0%, p-value = 0.051) versus 3.4p.p. (6.7%, p-value = 0.268) for men. As in previous cases, I still cannot rule out that both effects are the same ($p - value = 0.369$). Regarding the number of months worked, for men, the estimated effect on the number of months worked is -0.764 (-3.40%) and statistically insignificant ($p - value = 0.975$). For women, the estimated effect is 5.92 months (26.5%), and statistically significant ($p - value = 0.009$). The differences between the two coefficients now become statistically significant at a 5% level ($p - value = 0.046$). Finally, a similar comparison can be made for estimates measured at age 30. However, the effects are more imprecisely estimated due to the reduced sample size, and neither men nor women show statistically significant effects on their labor market outcomes by this age. Despite being less precisely estimated, the effects are still larger for women, and we can rule out that both effects are the same at a 10% (p-value = 0.099). A more in-depth discussion of this differential pattern is presented in the next section when the full dynamic effects are described.

2.6.4 Dynamic Effects

In this section, I provide a more thorough analysis of the effects of *PANES/AFAM-PE* measured age-by-age. In particular, I report results for outcomes measured as early as age 12 in the case of education outcomes and at age 15 for fertility and labor market outcomes. To maximize the use of information, the effects reported in this section are based on the *dynamic samples*. However, to rule out that the estimated trajectories of the effects are driven by changes in the sample's composition, Online Appendix reports estimates based on the fully balanced sample. The main caveat with using the balanced sample is that estimates are more imprecisely estimated due to the reduced sample size. Given the strong heterogeneous effects reported in the previous section, especially for labor market outcomes, the dynamic analysis is presented for men and women separately.

Fertility outcomes

Figures 2.8 and 2.9 report the main findings for the binary and continuous variables. To make comparisons easier, estimates for continuous variables are expressed in standard deviations.[44] In panel a. I report estimates of the effect of *PANES/AFAM-PE* on fertility outcomes measured at different ages. For instance,

[44] Online Appendix include Tables with point estimates, standard errors, robust p-values, and p-value of the equality of coefficients tests.

the estimated effect on the binary outcome reported at age 25 corresponds to the effect on the probability that a woman has given birth at or before age 25. The trajectory of the effects on fertility outcomes is similar for the binary and continuous variables. In both cases, the program's effects are strong, negative, and statistically significant when measured at ages around 17 and 18 years old. Estimates are also negative between 20 and 25 years old but slightly smaller in magnitude and statistically insignificant.[45] The effects start to trend toward the positive side starting at age 25. Overall, this pattern suggests that the effect of the program, at least until the age of 30, is associated with a postponement of the age of women's first birth rather than changes in overall preferences for the number of children.

The postponement effect is consistent with the scarce existing related literature that also finds stronger effects of cash transfers on fertility at early ages. For instance, using cross-section data, Araujo and Macours (2021) find that PROGRESA increased the age at which women had their first child by 0.5 years. In a very different context, Michelmore and Lopoo (2021) shows that exposure to EITC benefits in the US has stronger effects on early-life pregnancies around the age of 20 compared to the effects estimated around the mid-twenties. However, their analysis only covers the 16-25 period, so it is not clear what the trajectory of the effects on later-life outcomes is going to be.

[45]The only exceptions are the coefficients on the number of births measured at ages 23 and 24 (-0.14 and -0.17 births, respectively), which are negative and statistically significant ($p-value = 0.067$ and $p-value = 0.031$). While the magnitude of these coefficients measured in percentage points is larger compared to the effects measured at 17 or 18 years old, the size of the effect relative to the control mean is much smaller (16.7% and 18.6%).

Education outcomes

Panel b. in Figures 2.8 and 2.9 report estimates for secondary education outcomes. The effects on the extensive margin of secondary enrollment are mostly null for men and women. A slightly different story is observed when the effects are measured using the continuous variable. First, between ages 18 and 22 there is a consistently positive and statistically significant effect for men. During these four years, point estimates are between 0.15 and 0.2 standard deviations (i.e., a third of a year, or about 14% of the control mean). In all cases, the estimates are statistically significant ($p-values$ range between 0.019 and 0.058). On the contrary, estimates for women are smaller and statistically insignificant. However, the equality of coefficients cannot be rejected.[46] In sum, the age-by-age estimates on education outcomes again yield mixed evidence about the program's effects when estimated separately on men and women.

One alternative way of looking at effects on education outcomes in a dynamic setting is to analyze the effects of *PANES/AFAM-PE* on secondary enrollment for each grade separately. Online Appendix reports these results. First, when considering men and women together, the results indicate that *PANES/AFAM-PE* has null effects on enrollment in grades 1-3 (middle school), while it has strong and positive effects on enrolment in grades 4-6 (high-school). More specifically, *PANES/AFAM-PE* increases enrollment in 4th grade by 9.3p.p. (32.07%), in 5th grade by 5.9p.p. (28.83%), and in 6th grade by 4.7p.p. (32.36%). In each of these three cases, the effects are statistically significant (p-values of 0.003, 0.024, and 0.030, respectively). Combined with the baseline estimates reported above, the fact that there is an increase in the number of years enrolled in secondary

[46]For instance, $p-values$ are between 0.107 and 0.130 for estimates measured between ages 18 and 20 and between 0.220 and 0.525 for ages 21 and 22.

education, which is driven mostly by changes in high-school enrollment, suggests that academic progress might be playing a role.

Ideally, one would want to test this hypothesis directly and report estimates of the effects on actual years completed. However, this information is not available in the current data. For this reason, I conduct an alternative approach which could also be informative about potential effects on academic progress. First, I report estimates on the number of different grades in which an individual was enrolled to, separated by secondary education level. This variable takes a value between 0-3, where 0 corresponds, for instance, to someone that was not enrolled in middle school, while 3 corresponds to someone that was enrolled in grades 1st to 3rd. Consistent with previous findings, the effects of *PANES/AFAM-PE* are null on the number of middle school grades while positive and statistically significant for the number of high school grades. In this case, the effect is 0.187 (30.39%) with a p-value=0.002. In addition, I also define a variable that contains the maximum grade in which someone was enrolled from 1st to 6th. A positive effect on this variable would provide some suggestive evidence of academic progress. Consistent with this hypothesis, the estimated effects of *PANES/AFAM-PE* on the maximum grade enrolled is 0.35 (9.49%) and statistically significant (p-value=0.018).

However, when looking at the effects on education outcomes by grade and gender, there are some differential patterns between men and women. First, for men, only the effect on the probability of being enrolled in 4th grade is statistically significant. The size of this effect is large, 9.2p.p. (42.35%) and statistically significant (p-value = 0.001). In addition, there is an effect of *PANES/AFAM-PE* on men's number of grades of high school - 0.139 (31.68%) with a p-value of 0.014 - but there is no effect on the maximum grade in which an individual was enrolled. In this case, the effect is still sizable - 0.324 (9.63%) - but statistically insignificant (p-value=0.157). For women, the effects are larger and more widespread. For instance, there is a positive effect of the program both on 4th and 5th grade enrollment. These effects are 8.2p.p. (21.56%) and 10.1p.p. (36.78%), respectively, and statistically significant (p-value=0.084 and p-value=0.012). The effect on 6th grade enrollment is still sizable - 6.4p.p. (32.29%) - but statistically insignificant (p-value=0.119). The effects on women's number of grades enrolled in middle school are null, while for high school they are positive - 0.226 (27.75%) - and statistically significant (p-value=27.75%). Furthermore, the effect on the maximum grade enrolled is 0.366 (9.06%), which is slightly larger than for men but statistically significant (p-value = 0.047).

The main lessons that can be learned from the estimates on education outcomes are the following. First, *PANES/AFAM-PE* does not seem to affect enrollment in the secondary education system in the extensive margin. On the contrary, effects seem to be associated with changes in the number of years of secondary enrollment, and in particular, to an increase in the number of years enrolled in high school rather than middle school. In terms of differential patterns by gender, changes in the number of years of secondary education enrollment are driven by men. However, even though women do not show changes in the number of years enrolled in secondary, they do show an increase in the number of grades enrolled in high school and, most importantly, in the maximum grade in which they were enrolled. Together, these results suggest that effects on men's enrollment could be associated with a more passive enrollment, where they just stay in the system for longer but do not make significant academic progress, while women do. However, the evidence is not conclusive, and more work is required to confirm this interpretation.

Finally, panel c. in Figure 2.8 depicts the analogous estimates for tertiary education enrollment. The trajectory of the effects is consistent with the results discussed in Section 2.6.2. *PANES/AFAM-PE* does not seem to affect the probability of ever being enrolled in tertiary education for any of the ages considered.

Labor Market Outcomes

Binary labor market outcomes are reported in panel d. of Figure 2.8. Estimates for the continuous variables are reported in panels c. and d. of Figure 2.9. Panel d. of Figure 2.9 presents additional complementary evidence of effects on cumulative earnings.

The differences in the effects of the program between men's and women's labor market outcomes are striking. For men, *PANES/AFAM-PE* does not seem to have affected either labor market participation, months worked, or total cumulative earnings. Only 3 out of 45 p-values estimated for the 15 ages, and the 3 outcomes are smaller than 0.100, and those who are, do not follow any clear pattern. On the contrary, estimates for women provide substantial, robust evidence of positive effects on each of the three variables. These start as early as around 17-18 years old and continue relatively stable until their mid- or late-twenties. The differences between men and women are large and, in some cases, statistically significant. For instance, the effect of *PANES/AFAM-PE* on women's probability of having at least one four-month spell in the formal labor market measured at or before age 18 is 13p.p. (56.38%), while for men is -3p.p. and statistically insignificant ($p-value = 0.568$). The test of equality of coefficients for this outcome is rejected at a 5% level ($p-value = 0.027$). The differences are even larger for months worked and earnings. For instance, at age 24, the effect of *PANES/AFAM-PE* on women's months worked is 9.62 (40.79%), while for men, it is -3.22 (-9.97%) and statistically insignificant (p-value=0.497). The differential in the effects of the program on months worked between men and women measured at age 24 is 12.83, and it is statistically significant (p-value=0.004). More generally, this differential remains statistically significant between ages 17 and 27. The heterogeneous responses between men and women are not as large when comparing effects on cumulative earnings.

Finally, for each of the three labor market outcomes utilized, the effects attenuate by the late twenties and become null when reaching the 30s. Measured at 28, 29, or 30 years old, the program's effects are null both for men and women. Furthermore, by this age, there are no signs of the strong positive differential in favor of women's outcomes observed in the early twenties.

Stronger effects on women's labor market outcomes are also observed in related literature. For instance, in the US, Bastian et al. (2022); Hoynes et al. (2016); Bitler and Figinski (2019) find that the effect of different social safety net policies on children's adult outcomes is stronger for women. For PROGRESA, both Araujo and Macours (2021) and Parker and Vogl (2018) find similar heterogeneous responses where effects are more pronounced on women, although in some cases, the differences are not statistically significant. One recent piece of evidence that, at first glance, goes in the opposite direction is Barr et al. (2022), who show that the effects of additional exposure to EITC during childhood on early adulthood labor market outcomes are mostly driven by men. However, when making a more detailed comparison, the contrasting pattern is not so strong, and it could be explained mostly due to differences in the periods covered by the analysis. More specifically, Barr et al. (2022) do not report effects measured in the early 20s, which is the period where I find stronger effects on women's outcomes. Furthermore, unlike the estimates I provide in this paper, their analysis extends to the mid-thirties. At these ages effects on women's labor market outcomes become stronger. If one extrapolates this finding to the setting of *PANES/AFAM-PE*, this could suggest that the improved labor market outcomes might show up again after a period of high fertility. However, this is just speculative and can only be tested in the future when women are older.

Summary of results

The results discussed in this section illustrate the trajectory of the effects measured for ages 15 to 30 for each one of the dimensions included in the analysis. Comparing the age profiles of the effects for each outcome suggests a compelling story. Figure 2.10 illustrates this story by putting together the two main results, i.e., the age trajectories of the effects on women's fertility and labor market outcomes. This exercise aims to provide visual evidence of regular trends in the trajectories of the effects on different outcomes that could suggest the potential mechanisms behind these responses. When the two series are put together, a clear pattern emerges. Figure 2.10 illustrates that effects on fertility and labor market outcomes evolve inversely. The effects of *PANES/AFAM-PE* start to manifest as early as age 16, with negative effects on teenage pregnancies. These effects peak (in absolute value) around the ages of 17-18 and continue being negative, although smaller in size, until women reach 27 years old. At this age, they become positive but statistically insignificant. The overall age profile of the effects suggests that *PANES/AFAM-PE* did not change women's overall fertility preferences. Rather, it led to a postponement of births that otherwise would have occurred in the late teens. A similar but oppositely signed pattern is observed when looking at women's labor market outcomes. The positive effects on labor market outcomes start as early as age 18 and remain positive until the mid-twenties when they start to attenuate and trend toward 0. This attenuation coincides exactly with the attenuation (and even reversal) of the effects observed on fertility outcomes. The fact that labor market effects are exclusively observed on women, combined with such similar but inverse dynamics of fertility and labor market outcomes, suggests that *PANES/AFAM-PE* improves young women's labor market participation through a postponement of births.

In addition, while weaker and perhaps not as conclusive, the effects reported on women's academic progress in secondary education are also consistent with the trajectories observed for fertility and labor market outcomes. In particular, it is important to note that the effects on labor market outcomes become statistically significant only after age 18. Altogether, these results could indicate that, first, women make additional progress in secondary education, and then, they anticipate their entry into the formal labor market. However, as explained in the previous section, results on education outcomes must be interpreted cautiously because of the mixed and weaker patterns.

In the following section, I discuss in detail some of the main theoretical mechanisms that could explain these results.

2.7 Discussion

The results presented in previous sections indicate that *PANES/AFAM-PE* had strong effects on women's transition to adulthood, mostly on labor market outcomes, and coinciding with a postponement of their first birth. In addition, there is mixed evidence about the effect of the program on education decisions. While effects on tertiary education enrollment are consistently null, effects on secondary education seem to be strong for women, particularly in terms of high school enrollment. Furthermore, suggestive evidence shows that the effect on women's education might be explained by academic progress.

As discussed in Section 2.3, there are several potential mechanisms that could explain the effects reported in the previous section. For instance, the negative effects of *PANES/AFAM-PE* on teenage pregnancies can be explained by increased household income. On the one hand, higher income could lead to increased access to contraceptives, which in turn could reduce teenage pregnancies (e.g., as in Kearney and Levine, 2009; Bailey, 2006; Lundberg and Plotnick, 1995). Similarly, higher household income might reduce households' economic stress, making them a more attractive place to stay and reducing the incentives for young children to leave their parents' household to form a new one.

The suggestive evidence on education outcomes might also help explain reductions in teenage fertility. First, additional years in the education system might lead to a reduction in activities associated with risky behaviors that could lead to early-life pregnancies (Black et al., 2008; Berthelon and Kruger, 2011). Second, in a human capital framework, increased education improves expectations about future labor market outcomes, which is one of the key components of the opportunity costs of motherhood. In particular, this mechanism is supported by Araujo and Macours (2021) who show that PROGRESA improves children's earnings expectations, and also with literature on career choice that shows that the expected starting wage and the steepness of the earnings profile are strongly associated with fertility postponement (Van Bavel, 2010). This mechanism is also consistent with the literature in Demography that explains fertility postponement, partly by an increase in women's education (see Sobotka, 2010 for a thorough review).

The dynamic analysis discussed in Section 2.6.4 shows that fertility postponement was strongly associated with earlier participation in the labor market. This finding is also consistent with a whole strand of literature in Demography that discusses the relationship between fertility and labor market outcomes. For instance, since 1995, countries with higher delays in fertility have been associated with an increase in labor market participation and better wages for women (See Bratti, 2015 for a brief review). The micro-level evidence that uses biological fertility shocks to analyze the causal effects of fertility postponement on labor market outcomes also points in the same direction (Miller, 2011; Bratti and Cavalli, 2014).

The findings in this paper, at least for women's labor market outcomes, are also consistent with a broader literature that discusses the role of income during childhood on adulthood labor market outcomes (Akee et al., 2010; Bulman et al., 2021; Cesarini et al., 2016). However, if income *per se* was the main mechanism, one would also expect to observe a positive effect on men's outcomes. Overall, the fact that labor market effects are exclusively driven by women, jointly with such similar but oppositely signed trajectories for the effects on the women's labor market and fertility outcomes, are consistent with the interpretation that a postponement of women's age of first birth is one of the main channels for an earlier entry to the formal labor market and for the increased months and earnings.

2.8 Conclusion

Worldwide, governments spend billions of dollars on social safety net (SSN) policies to reduce poverty and inequality. With different designs depending on the context, cash transfers are one of the simplest and most used policy instruments for this purpose. Cash transfers can affect the lives of beneficiary household members in several ways. For instance, they can change parents' time allocation between labor, leisure, and housework, or children's school enrollment and healthcare decisions. All these changes affect the current lives of individuals but can also have long-lasting consequences, especially for children who benefited from the program in early life. This paper presents evidence of how a large-scale and government-implemented cash transfer program, the Uruguayan *PANES/AFAM-PE*, affects the transition to adulthood of individuals that benefited from the program when they were young.

Using a Regression Discontinuity Design that exploits the use of a poverty score to define eligibility to participate in the program, I show that the program reduces women's teenage pregnancies by 9.4p.p., increases participants' early adulthood labor market participation by 6.4p.p., months worked by 4.4, and earnings by about 12%. The evidence on education outcomes is mixed but suggests a stronger attachment to the secondary education system. For women, this stronger attachment is possibly explained by academic progress. Consistent with a postponement of women's first birth being the main driver, changes in labor market outcomes are observed exclusively for women.

The evidence reported in this paper has implications for the design, implementation, and evaluation of cash transfer policies. In particular, it suggests that cash transfers may play a key role in reducing labor market gender gaps, even when they are not specifically designed for this purpose. A back-of-the-envelope calculation illustrates that the differential effects of *PANES/AFAM-PE* on women's earnings represent a large share of the earnings gender gap. For instance, at age 25, the differential effect observed for women represents 57% of the earnings gap of a "pure" control group that is close to the threshold but never participated in the program.[47] At later ages, the positive effects on women's labor market outcomes seem to attenuate. Hence, one could be worried that the reduction in the

[47]One could be worried that never-treated are a group of reference that is strongly selected. Using the group of ineligible individuals based on the score obtained in the first application yields similar results. The differential effects represent about 39.7% of the gender gap, conditional on having earned income

labor market gender gap is only short-lived. However, changes in the timing of events still have strong consequences from a life-cycle perspective due to the existence of fixed costs and flatter wage profiles for mothers (e.g., Bratti, 2015; Miller, 2011). Cash transfers might help mitigate the motherhood penalty by delaying the time of a woman's first birth, even if they do not change the overall number of children. . This is particularly important in the context of an anti-poverty program in Uruguay, where the motherhood penalty is larger for low-income mothers, although it has reduced over time (Querejeta and Bucheli, 2022). One relevant question, that exceeds the case of Uruguay, is what is the role of public policy in reducing the motherhood penalty and, in general, the labor market gender gap. My paper illustrates that public policy has the potential to play a key role. Given that the motherhood penalty explains a sizable share of the labor market gender gap, policies that promote a postponement of pregnancies that otherwise would have occurred during teenage years might be particularly effective in reducing long-term labor market gender gaps.

In addition, the evidence reported in this paper suggests that cash transfers might also induce strong intergenerational effects, for instance, by affecting critical decisions such as the age of first birth. In this regard, the literature has shown that later-life pregnancies are associated to higher test scores or improved educational and psychological outcomes for children of the third generation (e.g., as discussed in Sobotka, 2010).

Overall, the study of the effects of *PANES/AFAM-PE* on individuals' transition to adulthood suggests that cash transfer programs can have long-lasting effects that should be considered when assessing their effectiveness, making them much more attractive. In particular, the positive effect on women's labor market outcomes combined with potential improvement in subsequent generations suggests that cash transfers can be a viable policy instrument to reduce long-run poverty and inequality in the long run.

Despite the thorough analysis reported in this paper, there are at least two key questions that remain unanswered since they require waiting for a longer time to observe these same individuals at older ages. The first one corresponds to the intergenerational effects of welfare participation. In particular, it is important to understand whether children that benefited from parents participating in welfare programs will also increase their own participation as adults. The empirical literature provides mixed evidence in this regard (Dahl and Gielen, 2021; Dahl et al., 2014; Hartley et al., 2022; Deshpande, 2016; Price and Song, 2018). Overall, improved labor market outcomes could suggest that they will not require to participate in welfare programs as adults. However, the attenuation observed by the late twenties weakens this interpretation. Unfortunately, the participation records used in the analysis only contain information until 2017. Hence, they do not allow me to provide precise estimates on children's adult participation in *PANES/AFAM-PE*, yet. However, this is a key topic for future research.

The second important question that remains unanswered is how the program would end up affecting overall fertility and, most importantly, what are the welfare implications. While the postponement of fertility has improved women's labor market outcomes, these improvements might come with a cost. In particular, birth at later ages, especially after mid-thirties, might lead to higher risks such as a higher probability of infertility, increased risk of miscarriage, and higher risk of pregnancy complications, among others, which also entail higher expected pecuniary and psychological costs of pregnancies (Schmidt et al., 2012; Gustafsson, 2001). In addition, one must consider that by delaying the age of first birth, the whole fertility cycle becomes shorter, and some women might be prevented from achieving their desired fertility plans. In this regard, demographers have suggested that the postponement transition is one of the reasons that explain a reduction in the total fertility rate observed in some societies for more than three decades (Kohler et al., 2002; Sobotka, 2004). The effects found in this paper correspond mostly to a postponement of birth that otherwise would have happened in the teenage years. Hence, the increased costs of postponing the age of firth birth are probably not as significant as they would be if the delay corresponded to pregnancies in the mid-, late-twenties. Furthermore, such early-life pregnancies also have some additional non-pecuniary costs that must be considered in an overall welfare assessment. In any case, a correct evaluation of welfare effects must weigh the positive effects on labor market outcomes and reduction in labor market gender gaps against potential changes in pecuniary and non-pecuniary costs of changes in the timing of the pregnancies. Furthermore, this has implications for policy design. For instance, if the costs of postponement are too large, it would be better from a social welfare perspective for governments to find a way to balance fertility and labor market decisions without inducing large delays. One way to go could be to increase investment in family-friendly policies such as early care education centers or parental leave policies.

Finally, the results in this paper illustrate the importance of dynamic analysis to assess the effects of public policy correctly. The findings on fertility outcomes provide a clear example. If the effects of the program on fertility are measured at around age 18, one could conclude that the program led to a reduction in fertility. On the other hand, focusing on the effects measured at age 30, one could conclude that the program did not have an effect on fertility. In both cases, one would have completely overlooked the postponement effect. Such a finding resembles the literature that discusses non-linear trajectories in the dynamic effects of early childhood policy interventions Almond et al. (2018); Chetty et al. (2011). This paper provides evidence based on a different type of intervention, i.e., cash transfers, at ages not necessarily restricted to early childhood. Furthermore, Almond et al. (2018) argue that the potential existence of non-linear patterns is one of the main reasons that justify studying the “missing middle”. In Almond et al. (2018), this missing middle refers to the lack of knowledge about the years between early childhood and adulthood in terms of developmental trajectories. The idea of an understudied missing middle that prevents us from understanding individuals’ life trajectories also relates to the primary motivation for this paper. My paper contributes to this strand of literature by providing novel evidence about a different “missing middle”.

Figure 2.1: Description of the Program

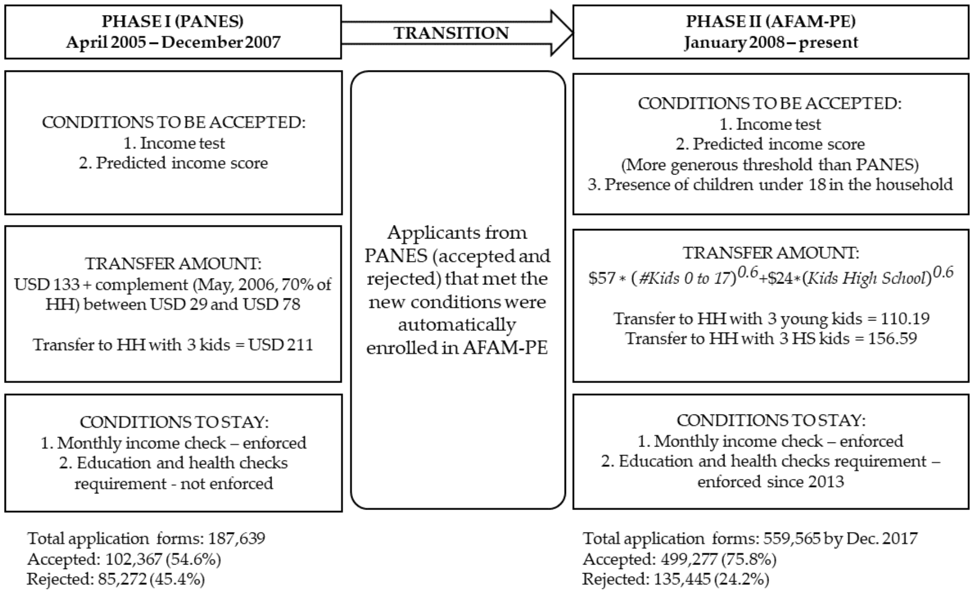
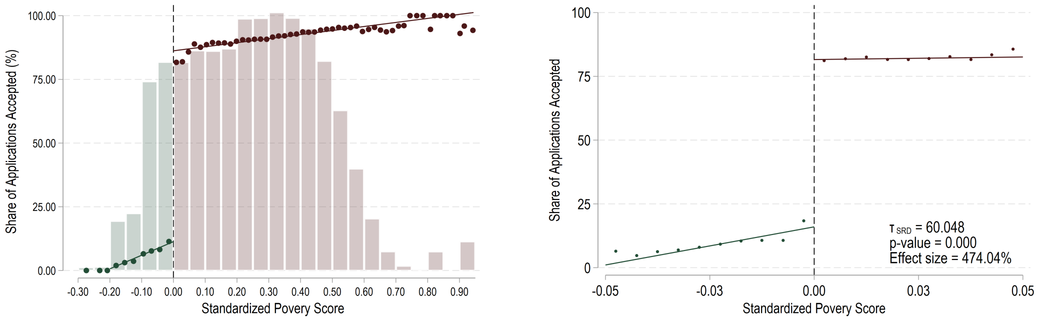


Figure 2.2: Relation Between Application Form Eligibility and Resoultion - Main Sample

a. Full Support

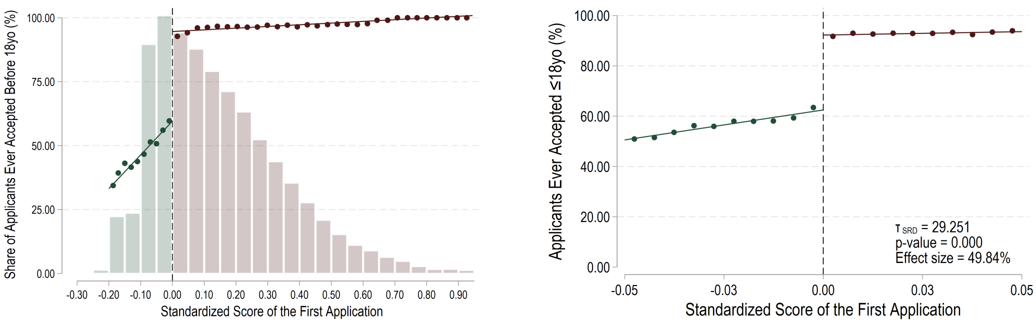
b. Optimal Bandwidth



Notes:This figure reports the share of application forms that were accepted as a function of the standardized poverty score (z) for the forms corresponding to individuals in the *main sample* as defined in Section 4.2.. Each observation used to construct this figure corresponds to an application form. Panel a. reports this relation for the full support of z. Negative values of z (depicted in green) indicate that the application does not meet the eligibility requirements, while positive values (depicted in red) correspond to eligible applications. Bars in the background depict the distribution of z. Each dot in the figure represents the average share of application forms accepted within a bin. The number of bins was selected manually such that the number of bins for negative values of z relative to the number of bins for the positive values of z represents the distribution of z. Panel b. focuses on application forms that are located within a bandwidth of 5p.p. of the eligibility threshold. In this case, observations are grouped into bins of 0.5p.p. width. In addition, the figure reports the point estimate of the local difference in the share of application forms accepted just at the threshold (τ_{SRD}), the p-value corresponding of a test of continuity, and the effect size expressed as a percent of the share of applications accepted for the ineligible group within the bandwidth depicted. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), the bandwidth used is selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights and robust standard errors clustered at the household level. For transparency and consistency between the point estimates reported and the graphical representation, estimates are based on a specification that does not include any additional covariates.

Figure 2.3: Participation Rule Using First Application Form - Main Sample Ever Treated Before 18yo

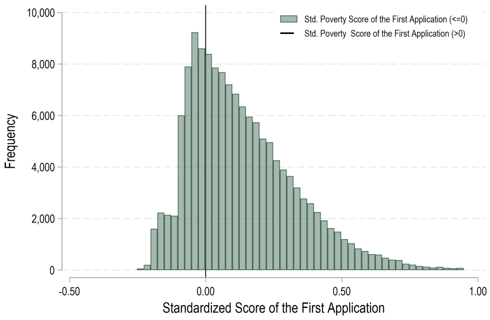
a. Full Support b. Optimal Bandwith



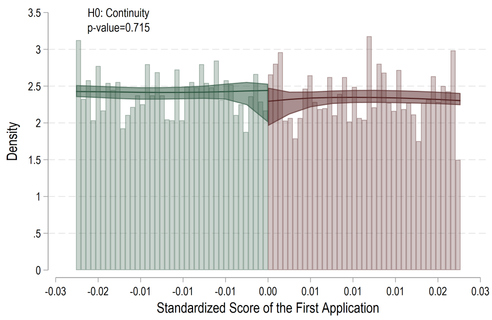
Notes: This figure reports the share individuals that were ever accepted to *PANES/AFAM-PE* before turning eighteen years old as a function of the standardized poverty score obtained in the first application (Z_i^{1st}) for the *main sample* as defined in Section 4.2.. Each observation used to construct this figure corresponds to an individual of the main sample as defined in Section XX. Panel a. reports this relation for the full support of z. Negative values of z (depicted in green) indicate that the score obtained in the first application form does not meet the eligibility requirements, while positive values (depicted in red) indicate that it does. Bars in the background depict the distribution of z. Each dot in the figure represents the average share of application form accepted within a bin. The number of bins was selected manually such that the number of bins for negative values of z relative to the number of bins for the positive values of z represents the distribution of z. Panel b. focuses on application forms that are located within a bandwidth of 5p.p. of the eligibility threshold. In this case, observations are grouped into 0.5p.p. width bins. In addition, the figure reports the point estimate of the local difference in the share of individuals ever treated just at the threshold (τ_{SRD}), the p-value corresponding of a test of continuity, and the effect size expressed as a percent of the share of ever treated individuals in the ineligible group within the bandwidth depicted. Estimates are based on the specification reported in equation 3. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), the bandwidth used is selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights and robust standard errors clustered at the household level. For transparency and consistency between the point estimates reported and the graphical representation, estimates are based on a specification that does not include any additional covariates. Additional details on the estimation procedure are reported in Table 2.

Figure 2.4: Continuity of the Poverty Score in 1st. Application Form - Main Sample

a. Distribution of the Poverty Score



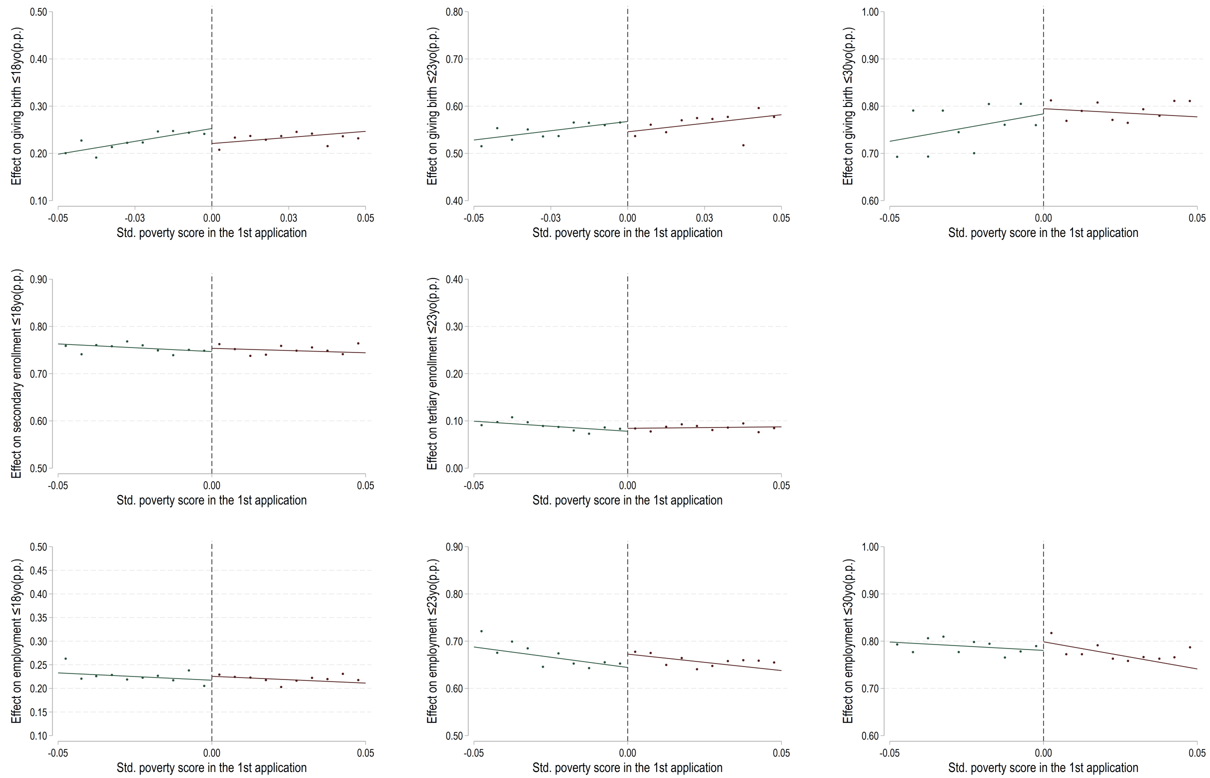
b. Cattaneo, Jansson, and Ma Test



Notes: This figure illustrates the distribution of the standardized poverty score obtained in the first application form (Z_i^{1st}) for the *main sample* as defined in Section 4.2.. Panel a. reports the distribution of Z_i^{1st} for its full support. Panel b. provides an illustration of a continuity test of Z_i^{1st} at the eligibility threshold as proposed by Cattaneo, Janson, and Ma (2020) and using the default options in the *rddensity* Stata command.

Figure 2.5: Graphic Evidence: Intention to Treat Effects, by Age - Binary Variables

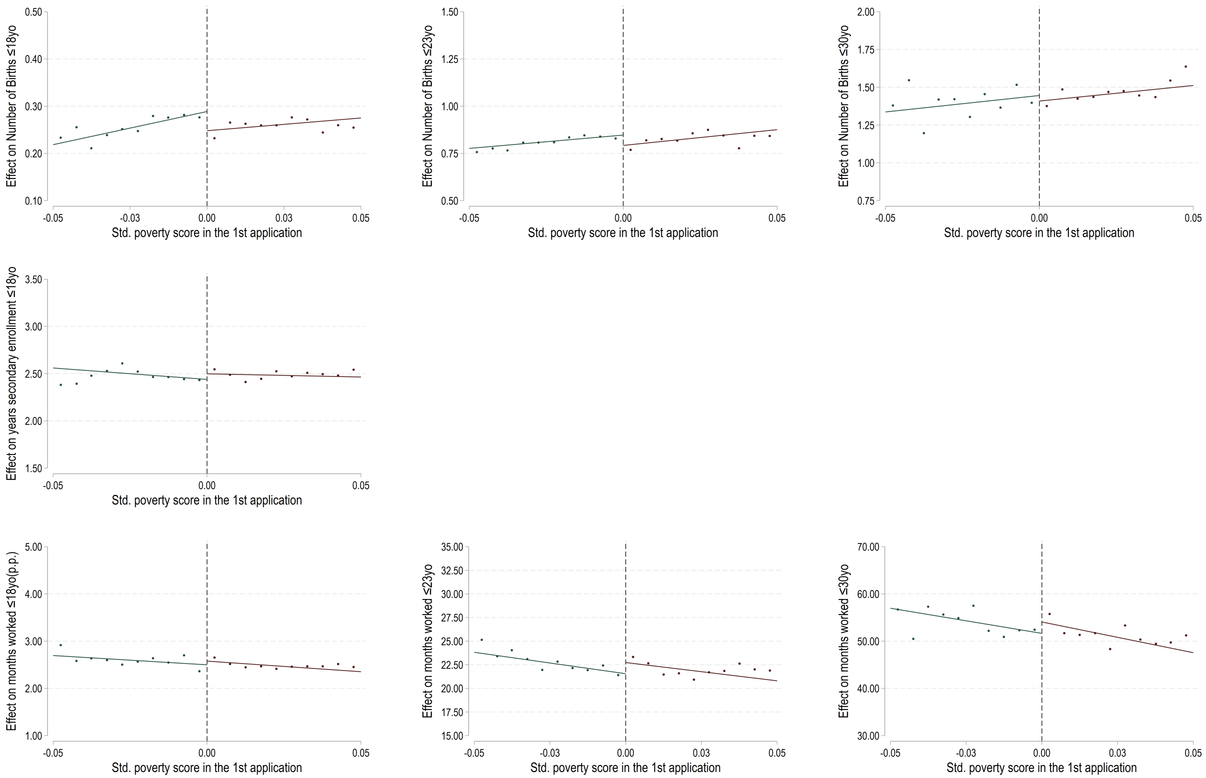
a. 18 years old b. 23 years old c. 30 years old



Notes: This figure illustrates the local intention to treat effects of the program on the different outcomes of interest using the score obtained in the first application (Z_i^{1st}) as the running variable. Panel a. depicts these effects for outcomes measured at the age of 18, panel b. does the same for outcomes measured at the age of 23, while panel c. uses outcomes measured at the age of 30. As explained in Section 4.2., panels a. and b. are based on the *main sample*, while estimates reported in panel c. are based on the *dynamic sample* of individuals with at least 30 years old by December, 2019. Row 1 focuses on women's fertility outcomes. In this case, the outcome variable takes the value of 1 if a woman has given birth by the corresponding age and 0 otherwise. Row 2 is focused on men and women's education outcomes. The outcome variable reported in panel a. takes the value of 1 if an individual has ever been enrolled into secondary education by the age of 18 and 0 otherwise. In panel b., the outcome variable is defined similarly but for tertiary education enrollment. As explained in Section 5, education outcomes are not meaningful at older ages. Therefore, I do not include estimates for tertiary enrollment by the age of 30. Finally, row 3 illustrates the effects of the program for men a women's labor market outcomes. In this case, the outcome variable is defined as 1 if the individual has had a four-consecutive-months employment spell in the labor market by the corresponding age and 0 otherwise. For comparison purposes all panels in this figure focus on application forms that are located within a bandwidth of 5p.p. of the eligibility threshold. In this case, observations are grouped into bins of 0.5p.p. width. Point estimates, standard errors, and additional details about the estimation procedure are reported in Table 4.

Figure 2.6: Graphic Evidence: Intention to Treat Effects, by Age - Continuous Variables

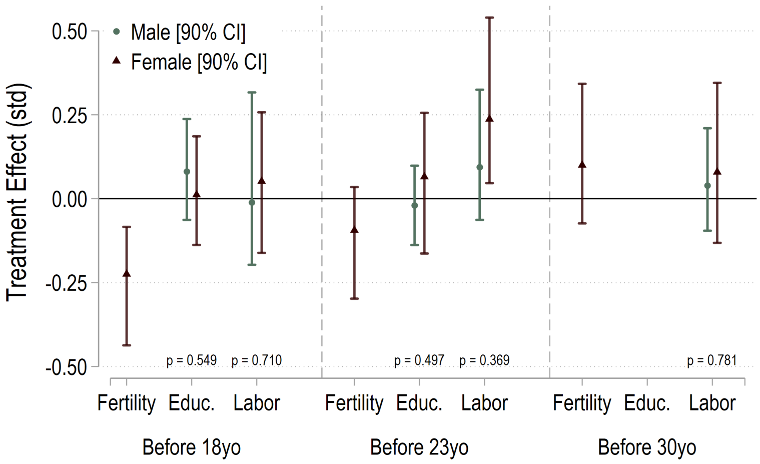
a. 18 years old b. 23 years old c. 30 years old



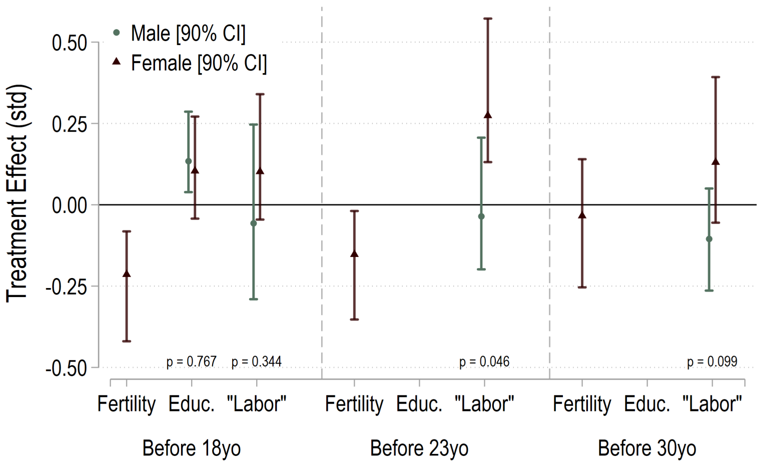
Notes: This figure illustrates the local intention to treat effects of the program on the different outcomes of interest using the score obtained in the first application (Z_i^{1st}) as the running variable. Panel a. depicts these effects for outcomes measured at the age of 18, panel b. does the same for outcomes measured at the age of 23, while panel c. uses outcomes measured at the age of 30. As explained in Section 4.2., panels a. and b. are based on the *main sample*, while estimates reported in panel c. are based on the *dynamic sample* of individuals with at least 30 years old by December, 2019. Row 1 focuses on women's fertility outcomes. In this case, the outcome variable measures the number of births that a woman has had by the corresponding age and 0 otherwise. Row 2 is focused on men and women's education outcomes. The outcome variable reported in panel a. measures the number of years that an individual has been enrolled into secondary education by the age of 18. As explained in Section 5, the information available for tertiary education only allows to measure enrollment. Therefore, tO do not estimate effects of the program on the number of years enrolled in tertiary education. Finally, row 3 illustrates the effects of the program for men a women's labor market outcomes. In this case, the outcome variable is defined as the number of months that an individual has worked in the formal labor market by the corresponding age. For comparison purposes all panels in this figure focus on application forms that are located within a bandwidth of 5p.p. of the eligibility threshold. In this case, observations are grouped into bins of 0.5p.p. width. Point estimates, standard errors, and additional details about the estimation procedure are reported in Table 4.

Figure 2.7: Heterogeneity by Gender

a. Binary outcomes



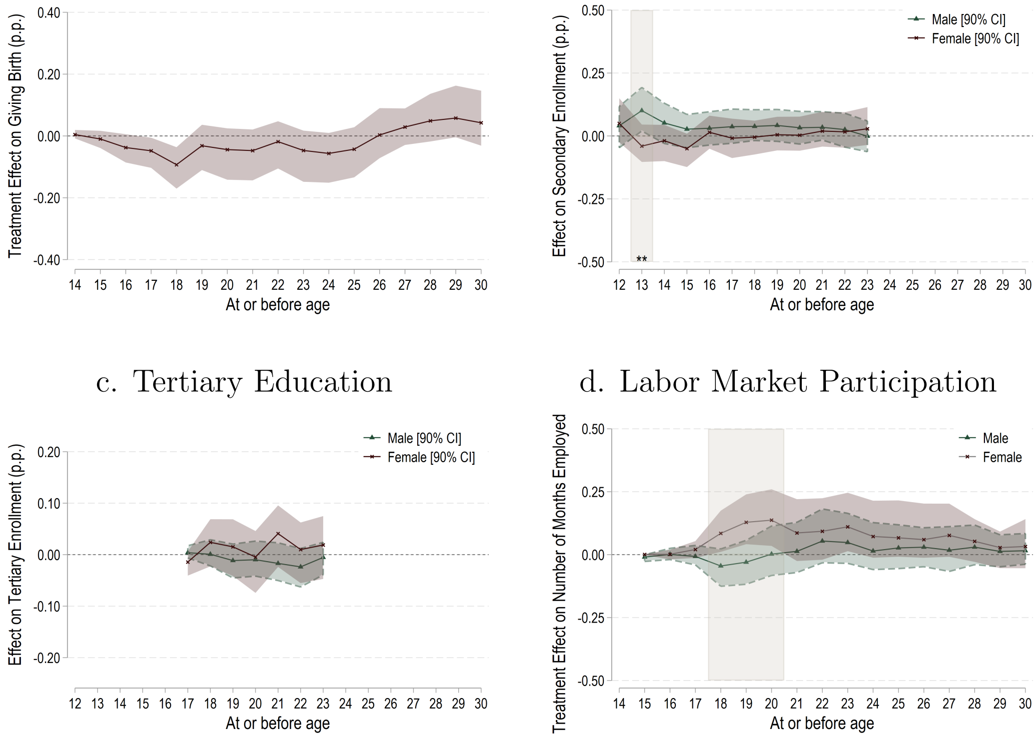
b. Number of events



Notes: This figure illustrates the local average treatment effects of the program on the different outcomes of interest, measured at different ages, for men (green) and women (red) separately. As explained in Section 4.2., estimates measured at 18 and 23 years old are based on the *main sample*, while estimates measured at the age of 30 are based on the *dynamic sample* of individuals with at least 30 years old by December, 2019. Panel a. reports estimates that correspond to the binary outcome variables, as defined in Figure 5. Panel b. replicates the analysis for the continuous outcome variables, as defined in Figure 6. In both cases, effects on fertility outcomes are reported exclusively for women. In all cases, for a simpler comparison across groups and outcomes, estimates are expressed in standard deviations. In addition to the point estimates and the 90% robust confidence intervals based on Cattaneo, Janson, and Ma (2020), the figure also reports the p-value for a test of equality of coefficients between men and women. See notes in Tables 6 and 7 for additional details about the estimation procedure.

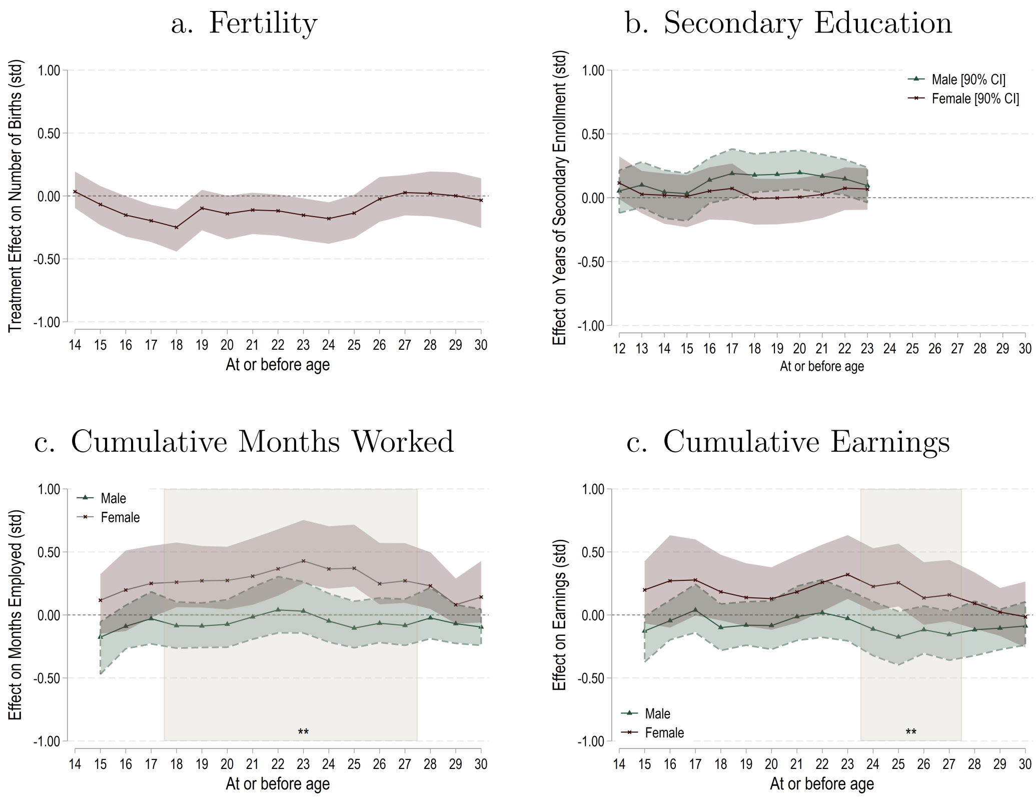
Figure 2.8: Dynamic Effects, by outcome and gender - Binary Variable

a. Fertility b. Secondary Education



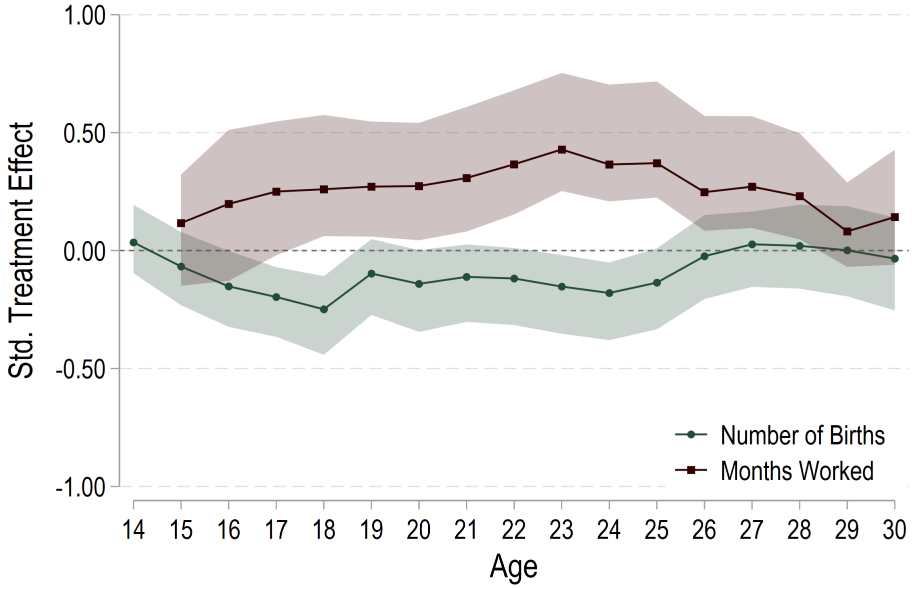
Notes: This figure reports the age-by-age estimates of the local average treatment effects on a set of outcomes of interest for men (green) and women (red) separately. Estimates are based on the *dynamic samples* as defined in Section 4.2.. Each dot in a figure represents a different RDD estimate. These estimates are based on the specification described by equation 1 estimated using 2SLS where eligibility based on the score obtained in the first application is used as an instrument for the endogenous variable of ever treatment. Panel a. reports the estimates for the binary fertility outcome. This variable takes the value of 1 if a woman gave birth by a certain age, measured in the x-axis. For instance, when the x-axis takes the value of 25, the outcome variable is defined as 1 if a woman has given birth at or before the age of 25 and 0 otherwise. Panel b. reports the estimates corresponding to the binary education outcomes, i.e. a variable that takes the value of 1 if an individual was enrolled in secondary education at or before a certain age and 0 otherwise. Panel c. focuses on a similarly defined variable but for tertiary education enrollment. Finally, panel d. reports estimates corresponding to a binary variable that takes the value of 1 if an individual has ever had an employment spell that lasted at least for consecutive months at or before a given age and 0 otherwise. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), each estimate corresponds to an optimal bandwidth selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights. All estimates correspond to a model that includes the pre-treatment variables described in Table 3 (except predicted eligibility) as control variables. 90% confidence intervals correspond to robust standard errors clustered at the household level. Full estimates are reported in Online Appendix as well as additional details on the estimation procedure.

Figure 2.9: Dynamic Effects, by outcome and gender - Continuous Variable



Notes: This figure reports the age-by-age estimates of the local average treatment effects on a set of outcomes of interest for men (green) and women (red) separately. Estimates are based on the *dynamic samples* as defined in Section 4.2.. Each dot in a figure represents a different RDD estimate. These estimates are based on the specification described by equation 1 estimated using 2SLS where eligibility based on the score obtained in the first application is used as an instrument for the endogenous variable of ever treatment. In all cases, estimates are standardized for an easier comparison across groups and outcomes. Panel a. reports the estimates for the continuous fertility outcome. This variable measures the number of births a woman has had by a certain age, measured in the x-axis. For instance, when the x-axis takes the value of 25, the outcome variable is defined as the number of births of a woman by the age of 25. Panel b. reports the estimates corresponding to the continuous education outcomes, i.e. a variable that counts the number of years an individual was enrolled in secondary education by a certain age. Panel c. focuses on a continuous measure of labor market participation. This variable counts the number of months that an individual has worked in the formal sector by a given age. Finally, panel d. reports an analogous measure but for the cumulative earnings by a given age. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), each estimate corresponds to an optimal bandwidth selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights. All estimates correspond to a model that includes the pre-treatment variables described in Table 3 (except predicted eligibility) as control variables. 90% confidence intervals correspond to robust standard errors clustered at the household level. Full estimates are reported in Online Appendix as well as additional details on the estimation procedure.

Figure 2.10: Dynamic Effects, Combined



Notes: This figure reports the age-by-age estimates of the local average treatment effects of the program on women's fertility and labor market outcomes. These series of effects correspond to the ones reported in Figure 9, panels a. and c. See notes in Figure 9 for details on the estimation procedure.

Table 2.1: Descriptive Statistics: Individual Characteristics - By Sample

	Main Sample: ≥ 23 years old At Dec, 2019		Dynamic Sample: ≥ 30 years old Fertility		Dynamic Sample: ≥ 30 years old Labor Market	
	All	Opt. Bandwidth	All	Opt. Bandwidth	All	Opt. Bandwidth
	(1)	(2)	(3)	(4)	(5)	(6)
a. Individual Characteristics						
Female (%)	50.32	50.70	49.88	51.02	52.13	53.60
	(50.00)	(50.00)	(50.00)	(49.99)	(49.96)	(49.87)
Number of HH.	1.77	1.63	1.93	1.67	1.99	1.80
	(1.10)	(0.97)	(1.20)	(1.03)	(1.20)	(1.05)
Age at 31 Dec. 2019	26.91	27.11	31.11	31.09	29.97	30.01
	(2.56)	(2.61)	(0.68)	(0.66)	(1.20)	(1.19)
Age of 1st application	13.42	13.40	16.84	16.87	15.98	15.98
	(2.59)	(2.61)	(0.69)	(0.67)	(1.22)	(1.19)
Accepted before 18yo (%)	84.16	73.62	71.04	53.30	80.94	65.63
	(36.52)	(44.07)	(45.36)	(49.89)	(39.28)	(47.49)
Number of app. forms	2.56	2.47	2.44	2.23	2.71	2.62
	(1.44)	(1.35)	(1.31)	(1.22)	(1.32)	(1.23)
In *PANES* form (%)	78.39	86.89	100.00	100.00	91.27	94.94
	(41.16)	(33.75)	(0.00)	(0.00)	(28.23)	(21.93)
In *AFAM-PE* form (%)	96.08	93.21	91.40	85.94	99.85	99.73
	(19.42)	(25.16)	(28.04)	(34.77)	(3.92)	(5.20)
b. Reference Form						
Std. Score	0.18	-0.00	0.16	-0.00	0.18	-0.00
	(0.25)	(0.05)	(0.27)	(0.03)	(0.26)	(0.05)
App. Accepted (%)	71.82	49.08	69.58	50.52	73.27	50.69
	(44.99)	(49.99)	(46.01)	(50.00)	(44.25)	(50.00)
PANES (%)	78.39	86.89	100.00	100.00	91.27	94.94
	(41.16)	(33.75)	(0.00)	(0.00)	(28.23)	(21.93)
Capital City (%)	31.25	18.18	29.64	15.64	31.27	17.90
	(46.35)	(38.56)	(45.67)	(36.33)	(46.36)	(38.34)
c. Household characteristics (ref. form)						
Single Parent (%)	46.73	48.96	48.30	54.65	47.14	49.56
	(49.89)	(49.99)	(49.97)	(49.79)	(49.92)	(50.00)
Number of members	4.92	4.38	4.98	4.23	5.06	4.45
	(1.98)	(1.82)	(2.12)	(1.91)	(2.07)	(1.87)
Number of children	2.95	2.41	2.91	2.27	3.02	2.43
	(1.69)	(1.45)	(1.78)	(1.46)	(1.76)	(1.48)
Avg. age	23.13	25.13	24.35	26.65	23.59	25.79
	(7.60)	(8.00)	(7.85)	(8.16)	(7.62)	(8.02)
Household Head: Ed. years	6.83	7.20	6.72	6.94	6.71	7.05
	(3.41)	(3.55)	(3.47)	(3.51)	(3.42)	(3.55)
Household head: Employed (%)	63.43	64.55	65.31	63.48	64.36	64.29
	(48.16)	(47.84)	(47.60)	(48.15)	(47.89)	(47.92)
Household head: income	143.33	159.37	128.79	133.69	130.82	147.87
	(172.54)	(171.86)	(140.96)	(134.38)	(145.62)	(152.77)
Observations	224,413	76,593	34,754	7,971	59,667	21,779

Notes: Table 1 reports a series of descriptive statistics for some of the samples used in the analysis. Columns (1) and (2) are focused on the *main sample* as described in Section 4.2.. Columns (3) and (4) are based on the *dynamic sample* used for estimates on fertility outcomes. This dynamic sample is defined as individuals who were younger than 18 years old at the time of their first application, and at least 30 years old by December, 2019, the latest date included in birth records. Columns (5) and (6) are based on the *dynamic sample* used for estimates on labor market outcomes that is defined analogously but for individuals with at least 30 years old by October, 2021, the latest date included in labor histories. Odd columns report statistics that describe individuals across the full support of the running variable, while even columns report statistics corresponding to individuals that are within the optimal bandwidths used in the RDD estimates of the effects of the program. Panel a. reports information on a series of characteristics at the individual level. Panel b. focuses on the characteristics of the reference form, i.e., the application form corresponding to the first application. Finally, panel c. reports information about the characteristics of the household defined in the first application form.

Table 2.2: First Stage - Main Sample

	First Stage Estimates - Alternative Specifications			
	(1)	(2)	(3)	(4)
a. Dep. Var: Ever Treated Before 18 Years Old				
Eligibility	29.251***	29.089***	29.677***	29.620***
	(0.892)	(1.013)	(0.921)	(1.067)
Robust p-value	0.000	0.000	0.000	0.000
Effect Size (%)	49.84%	52.37%	50.08%	51.72%
Bwd.	[0.033;0.033]	[0.054;0.054]	[0.026;0.026]	[0.043;0.043]
Observations	31,413	52,538	24,551	40,813
Parameter Selection:				
Pol. Degree	1	2	1	2
Bwd. Method	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Uniform	Uniform

Notes: Table 2 reports the first stage coefficients based on the *main sample* defined as in Section 4.2. (equation 2). These coefficients measure the effect of obtaining a poverty score above the eligibility threshold in the first application (Z_i^{1st}) on the probability of ever participating in the program. In this case, the outcome variable takes the value of 100 if an individual ever received the benefits of the program, while takes the value of 0 otherwise. Column (1) reports the estimates for the baseline specification that uses an optimal MSERD bandwidth (Calonico, Cattaneo, Farrell, and Titiunik, 2018), a triangular kernel function, and a linear local polynomial. Columns (2) through (4) reports a series of estimates based on alternative specification as robustness tests. Column (2) changes the polynomial degree from linear to quadratic, column (3) uses a uniform kernel function and a linear local polynomial, while column (4) uses a uniform kernel function and a quadratic local polynomial. In addition to the point estimate, this table reports robust standard errors clustered at the household level, the robust p-value, the effect size measured as a percent of the mean for the ineligible group within the optimal bandwidth, the optimal bandwidth, and the total number of effective observations used in the estimation.

Table 2.3: Balance of Baseline Covariates - Main Sample

	Ineligible Intercept	Eligible Intercept	Difference (2) - (1)	p-value Robust	Sharpened FDR q-value
	(1)	(2)	(3)	(4)	(5)
Predicted Eligibility	0.66	0.66	0.003	0.635	0.651
HH - Avg. Age	25.11	25.41	0.306	0.113	0.276
HH - Avg. age adults	39.92	40.53	0.607	0.016	0.118
HH - Capital City	0.19	0.17	-0.023	0.021	0.118
HH - Number of people	4.27	4.20	-0.073	0.126	0.276
HH - Number of children	2.35	2.32	-0.030	0.358	0.416
HH - Single Parent	0.53	0.55	0.024	0.091	0.276
HHH - Income (IHS)	4.32	4.33	0.010	0.690	0.651
HHH - Employed	0.61	0.62	0.008	0.517	0.603
HHH - Years of Educ.	6.98	7.06	0.080	0.348	0.416
Age at 1st. App.	13.40	13.41	0.011	0.645	0.651
Age (Dec. 31, 2019)	29.04	28.91	-0.130	0.010	0.118
Number of Apps.	2.80	2.81	0.012	0.896	0.916
Female	51.17	50.40	-0.768	0.367	0.416
Number of HH.	1.64	1.61	-0.026	0.098	0.276

Notes: Table 3 is the RDD-analogous to a balance table in an experimental design. All variables included in the Table are measured at the baseline, i.e., at the moment of application, and correspond to the *main sample* as defined in Section 4.2.. For each variable included in the Table, I replicate the baseline estimation procedure used on the main outcomes and test whether they are continuous at the threshold. In all cases, estimates are based on an optimal MSERD bandwidth (Calonico, Cattaneo, Farrell, and Titiunik, 2018), a triangular kernel function, and local linear regressions. Column (1) reports the intercept of the regression function estimated on the ineligible side of the running variable. Column (2) does the same but for observations in the eligible side. Column (3) reports the difference between columns (2) and (1), i.e, the sharp RDD estimate. Column (4) reports the robust p-value of the continuity test. Column (5) reports the sharpened FDR q-values that adjust for multiple hypotheses testing. The variable predicted eligibility correspond to predicted values estimated based on a probit model where the dependent variable is defined as 1 if the individual has a score in their first application that is above the eligibility threshold, and 0 otherwise. All the other variables in the table are included as regressors in this probit model and used to calculate the predicted eligibility.

Table 2.4: Intention to Treat Effects

	18 years old			23 years old			30 years old	
	Fertility	Education	Labor	Fertility	Education	Labor	Fertility	Labor
				a. Dep. Var.: Dummy Variable				
Ever Treated	-0.031***	0.008	0.004	-0.023*	0.005	0.020**	0.011	0.006
	(0.011)	(0.007)	(0.007)	(0.014)	(0.005)	(0.008)	(0.022)	(0.012)
Robust p-value	0.005	0.204	0.559	0.078	0.445	0.022	0.822	0.632
Effect Size (%)	-13.52%	1.06%	1.77%	-4.14%	5.83%	3.05%	1.49%	0.75%
Bwd.	[0.050;0.050]	[0.063;0.063]	[0.055;0.055]	[0.042;0.042]	[0.048;0.048]	[0.059;0.059]	[0.071;0.071]	[0.068;0.068]
Observations	24,078	61,225	46,474	20,292	45,194	49,828	6,504	17,964
				b. Dep. Var.: Number of Events				
Ever Treated	-0.038***	0.086**	-0.009	-0.060**		0.816**	-0.041	2.001
	(0.013)	(0.039)	(0.076)	(0.026)		(0.395)	(0.068)	(1.540)
Robust p-value	0.003	0.029	0.865	0.016		0.045	0.381	0.145
Effect Size (%)	-14.93%	3.41%	-0.36%	-7.36%		3.65%	-2.96%	3.72%
Bwd.	[0.052;0.052]	[0.046;0.046]	[0.060;0.060]	[0.040;0.040]		[0.054;0.054]	[0.057;0.057]	[0.052;0.052]
Observations	25,258	43,514	49,987	19,294		44,910	5,245	13,560
Parameter Selection:								
Pol. Degree	1	1	1	1	1	1	1	1
MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Notes: Table 4 reports the effects of obtaining a score of the first application above the eligibility threshold on a series of outcomes of interest. These must be interpreted as intention to treat effects (ITT), or reduced form effects and are based in the specification described in equation 3. Columns (1) through (3) correspond to ITT effects of the program on outcomes measured at the age of 18, columns (4) through (6) correspond to outcomes measured at the age of 23, and columns (7) and (8) correspond to outcomes measured at the age of 30. Panel a. reports a series of estimates based on binary outcome variables, as defined in Figure 5. Panel b. reports estimates on a series of continuous outcome variables as defined in Figure 6. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), the bandwidth used is selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights and robust standard errors clustered at the household level. For transparency and consistency between the point estimates reported in this table and the graphical representation in Figures 5 and 6, estimates are based on a specification that does not include any additional covariates. In addition to the point estimate, this table reports robust standard errors clustered at the household level, the robust p-value, the effect size measured as a percent of the mean for the ineligible group within the optimal bandwidth, the optimal bandwidth, and the total number of effective observations used in the estimation.

Table 2.5: Local Average Treatment Effects

	≤ 18 years old			≤ 23 years old			≤ 30 years old	
	Fertility	Education	Labor	Fertility	Education	Labor	Fertility	Labor
	a. Dep. Var.: Dummy Variable							
Ever Treated	-0.094**	0.017	0.021	-0.047	0.013	0.064*	0.043	0.016
	(0.039)	(0.026)	(0.034)	(0.042)	(0.015)	(0.032)	(0.045)	(0.028)
Robust *p*-value	0.015	0.469	0.313	0.193	0.526	0.062	0.288	0.570
Effect Size (%)	-41.19%	2.21%	9.36%	-8.52%	13.72%	9.69%	5.62%	1.98%
Bwd.	[0.044;0.044]	[0.042;0.042]	[0.044;0.044]	[0.050;0.050]	[0.072;0.072]	[0.065;0.065]	[0.040;0.040]	[0.068;0.068]
Observations	20,033	37,050	33,594	22,711	66,259	51,206	3,343	16,875
	b. Dep. Var.: Number of Events							
Ever Treated	-0.108**	0.253**	0.250	-0.137*		4.401***	-0.039	1.259
	(0.044)	(0.114)	(0.382)	(0.078)		(1.871)	(0.114)	(2.674)
Robust *p*-value	0.015	0.027	0.253	0.067		0.005	0.635	0.324
Effect Size (%)	-41.95%	10.07%	9.84%	-16.86%		19.77%	-2.78%	2.31%
Bwd.	[0.049;0.049]	[0.041;0.041]	[0.041;0.041]	[0.046;0.046]		[0.041;0.041]	[0.046;0.046]	[0.082;0.082]
Observations	22,536	36,257	31,203	20,776		31,602	3,760	19,810
Parameter Selection:								
Pol. Degree	1	1	1	1	1	1	1	1
MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Notes: Table 5 reports the local average treatment effects of being ever treated on a series of outcomes of interest. These estimates are based on the specification described by equation 1 estimated using 2SLS where eligibility based on the score obtained in the first application is used as an instrument for the endogenous variable of ever treatment. Columns (1) through (3) correspond to the effects of the program on outcomes measured at the age of 18, columns (4) through (6) correspond to outcomes measured at the age of 23, and columns (7) and (8) correspond to outcomes measured at the age of 30. Panel a. reports a series of estimates based on binary outcome variables, as defined in Figure 5. Panel b. reports estimates on a series of continuous outcome variables as defined in Figure 6. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), the bandwidth used is selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights and robust standard errors clustered at the household level. All estimates correspond to a model that includes the pre-treatment variables described in Table 3 (except predicted eligibility) as control variables. In addition to the point estimate, this table reports robust standard errors clustered at the household level, the robust p-value, the effect size measured as a percent of the mean for the ineligible group within the optimal bandwidth, the optimal bandwidth, and the total number of effective observations used in the estimation.

Table 2.6: LATE Effects, by Age - Estimates With Covariates - Male

	18 years old		23 years old		30 years old
	Education	Labor	Education	Labor	Labor
	a. Dep. Var.: Dummy Variable				
Ever Treated	0.035	-0.005	-0.006	0.044	0.016
	(0.034)	(0.054)	(0.018)	(0.046)	(0.031)
Robust p-value	0.341	0.702	0.782	0.268	0.538
Effect Size (%)	4.55%	-2.09%	-6.30%	6.70%	1.96%
Bwd.	[0.052;0.052]	[0.052;0.052]	[0.064;0.064]	[0.066;0.066]	[0.122;0.122]
Observations	23,189	19,451	29,094	25,170	12,197
	b. Dep. Var.: Number of Events				
Ever Treated	0.280**	-0.253		-0.764	-4.698
	(0.135)	(0.607)		(2.314)	(3.665)
Robust p-value	0.031	0.894		0.975	0.263
Effect Size (%)	10.89%	-10.03%		-3.40%	-8.49%
Bwd.	[0.061;0.061]	[0.050;0.050]		[0.070;0.070]	[0.138;0.138]
Observations	27,878	18,780		26,603	13,116
Parameter Selection:					
Pol. Degree	1	1	1	1	1
Bwd. Method	mserd	mserd	mserd	mserd	mserd
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular

Notes: Table 6 reports the local average treatment effects of being ever treated on a series of outcomes of interest for the sub-sample of men. These estimates are based on the specification described by equation 1 estimated using 2SLS where eligibility based on the score obtained in the first application is used as an instrument for the endogenous variable of ever treatment. Columns (1) and (2) correspond to the effects of the program on outcomes measured at the age of 18, columns (3) and (4) correspond to outcomes measured at the age of 23, and column (5) correspond to outcomes measured at the age of 30. As explained in Section 4, due to data limitations, there is no reliable information available about men's fertility decisions. Panel a. reports a series of estimates based on binary outcome variables, as defined in Figure 5. Panel b. reports estimates on a series of continuous outcome variables as defined in Figure 6. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), the bandwidth used is selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights and robust standard errors clustered at the household level. All estimates correspond to a model that includes the pre-treatment variables described in Table 3 (except predicted eligibility) as control variables. In addition to the point estimate, this table reports robust standard errors clustered at the household level, the robust p-value, the effect size measured as a percent of the mean for the ineligible group within the optimal bandwidth, the optimal bandwidth, and the total number of effective observations used in the estimation.

Table 2.7: LATE Effects, by Age - Estimates With Covariates - Female

	18 years old			23 years old			30 years old	
	Fertility	Education	Labor	Fertility	Education	Labor	Fertility	Labor
				a. Dep. Var.: Dummy Variable				
Ever Treated	-0.094**	0.005	0.021	-0.047	0.019	0.112*	0.043	0.032
	(0.039)	(0.036)	(0.044)	(0.042)	(0.031)	(0.060)	(0.045)	(0.052)
Robust p-value	0.015	0.807	0.706	0.193	0.717	0.051	0.288	0.462
Effect Size (%)	-41.19%	0.65%	9.73%	-8.52%	20.76%	16.98%	5.62%	4.12%
Bwd.	[0.044;0.044]	[0.050;0.050]	[0.047;0.047]	[0.050;0.050]	[0.050;0.050]	[0.046;0.046]	[0.040;0.040]	[0.045;0.045]
Observations	20,033	22,751	18,913	22,711	22,778	18,375	3,343	5,740
				b. Dep. Var.: Number of Events				
Ever Treated	-0.108**	0.216	0.455	-0.137*		5.920***	-0.039	5.855
	(0.044)	(0.166)	(0.435)	(0.078)		(2.417)	(0.114)	(5.237)
Robust p-value	0.015	0.231	0.209	0.067		0.009	0.635	0.215
Effect Size (%)	-41.95%	8.60%	17.98%	-16.86%		26.50%	-2.78%	10.93%
Bwd.	[0.049;0.049]	[0.045;0.045]	[0.048;0.048]	[0.046;0.046]		[0.047;0.047]	[0.046;0.046]	[0.041;0.041]
Observations	22,536	20,375	19,214	20,776		18,756	3,760	5,210
Parameter Selection:								
Pol. Degree	1	1	1	1	1	1	1	1
MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD	MSERD
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular

Notes: Table 7 reports the local average treatment effects of being ever treated on a series of outcomes of interest for the sub-sample of women. These estimates are based on the specification described by equation 1 estimated using 2SLS where eligibility based on the score obtained in the first application is used as an instrument for the endogenous variable of ever treatment. Columns (1) through (3) correspond to the effects of the program on outcomes measured at the age of 18, columns (4) through (6) correspond to outcomes measured at the age of 23, and columns (7) and (8) correspond to outcomes measured at the age of 30. Panel a. reports a series of estimates based on binary outcome variables, as defined in Figure 5. Panel b. reports estimates on a series of continuous outcome variables as defined in Figure 6. Following Calonico, Cattaneo, Farrell, and Titiunik (2018), the bandwidth used is selected by minimizing the mean squared error (MSE). Estimates reported are based on local linear regressions using triangular weights and robust standard errors clustered at the household level. All estimates correspond to a model that includes the pre-treatment variables described in Table 3 (except predicted eligibility) as control variables. In addition to the point estimate, this table reports robust standard errors clustered at the household level, the robust p-value, the effect size measured as a percent of the mean for the ineligible group within the optimal bandwidth, the optimal bandwidth, and the total number of effective observations used in the estimation.

CHAPTER 3

How do Top Earners Respond to Taxation? Evidence from a Tax Reform in Uruguay

Marcelo Bergolo, Universidad de la Republica and IZA

Gabriel Burdin, Leeds University Business School and IZA

Mauricio de Rosa, Universidad de la Republica and Paris School of Economics

Matias Giaccobasso, University of California, Los Angeles [1]

Martin Leites, Universidad de la Republica

Horacio Rueda, University of Houston

Abstract

We present new evidence on how top income earners respond to personal income taxation. An unprecedented combination of exhaustive administrative records and variation in the tax rates for the top 1% income earners allows us to uncover the different margins of responses. Exploiting a tax reform implemented

[1]As described in the acknowledgments page, Chapter 3 is a version of a submitted article. The latest version of this article as well as the complement online appendix, can be found here. All listed co-authors are principal investigators and contributed in equal shares in the elaboration of the article. We have benefited from comments and suggestions from Anne Brockmeyer, Guillermo Cruces, Jim Hines, Eckhard Janeba, Claus Thustrup Kreiner, David Lopez, Andreas Peichl, Roberto Hsu Rocha, Alisa Tazhitdinova, Mazhar Waseem, Nicolas Werquin and numerous seminar participants. We also wish to thank the Direccion General Impositiva (DGI) by providing us with the administrative tax records. This research was funded by the Agencia Nacional de Investigacion e Innovacion (ANII) - grant FCE 1 2014 104284

in Uruguay in 2012, we estimate an intensive margin elasticity of 0.577, partially explained by a real labor supply adjustment. Responses on the extensive margin are larger (semi-elasticity of 2.479), driven mainly by labor-to-corporate income shifting (semi-elasticity of -1.967). The efficiency costs of the reform represent 31% of the projected tax revenue.

3.1 Introduction

A substantial increase in income and wealth concentration (Atkinson et al., 2011; Alvaredo et al., 2018), combined with an increasing need to finance public spending, has revitalized the debate about the appropriate level of taxation to top income earners (TIEs).[2] The size and the type of TIEs' behavioral responses to taxation, e.g., due to changes in labor supply decisions (Feldstein, 1995) or tax avoidance/evasion responses (Slemrod and Kopczuk, 2002; Slemrod, 2019a), and their implied efficiency costs should be key inputs in this discussion (Slemrod, 1995, 2001; Piketty et al., 2014).

A prolific strand of empirical literature has already shown that income taxation induces substantial behavioral responses among high-income taxpayers, mostly through tax avoidance rather than real labor supply adjustments (Saez et al. 2012 and Auten et al. 2016). However, the empirical analysis of how TIEs react to tax rate changes presents at least two critical challenges that have rarely been addressed simultaneously. First, it requires an exogenous change in the tax rate

[2]Recently, to name a few in the context of the United States, authors in academic (see, e.g., New York Times on Jan. 22, 2019) and political (see, e.g., Bloomberg on Jan. 14, 2019) circles have called on the U.S. Congress to raise the top marginal income tax rates substantially. Furthermore, the recent tax reform proposed by President Biden includes an increase in the top marginal tax rate from 37% to 39.6% that would only affect the top 1% (see, e.g., New York Times on Apr. 22, 2021).

among TIEs. Most existing studies have to rely on tax reforms that affect individuals across the whole earnings distribution, limiting the comparability of the control groups. Moreover, many of these tax reforms affected both the tax rates and the tax base, which makes the interpretation of the results more intricate. Second, very few studies are able to analyze the behavioral responses on a comprehensive set of income sources. Since TIEs have numerous opportunities for advanced tax planning, accounting for mechanisms such as income shifting is important to precisely estimate the welfare effects of tax reforms (Slemrod, 1995; Chetty, 2009) and to offer policy recommendations (Piketty et al., 2014).

This paper contributes with novel empirical evidence that addresses these limitations in a unified setting. We exploit a unique reform to Uruguay's progressive personal labor income tax (PLIT) schedule that took place in 2012. This reform generated variation in the tax rates that exclusively affected the top 1% of the labor income distribution and left unchanged all other relevant tax bases. This has two relevant implications. First, some TIEs experienced an increase in their tax rates, while other TIEs did not. Second, the reform affected the tax rate differential between tax bases, increasing the incentives to shift between tax bases, especially for self-employed individuals. The tax reform was salient, fairly simple to understand, and similar in size to the policy variation used by previous studies (e.g., Saez 2017).

Our empirical analysis leverages a rich set of administrative records to describe TIEs responses on three different margins: intensive, extensive, and income-shifting. We use individual-level tax records for the universe of taxpayers between 2009 and 2015 containing information, for instance, on earned income, deductions, and tax withholdings, both from tax returns and third-party reports. We link this information at the individual level with corporate income tax (CIT), personal income tax on capital (PITC) returns, and employer-level information. Furthermore, we complement this with employer-employee social security data on earned income and hours worked. The combination of a rare tax reform that affected the upper portion of the income distribution exclusively and such detailed individual-level microdata creates a unique setting to dig into the individual responses to income taxation.

In the first part of the paper, we develop a simple model where individuals can choose between alternative tax bases to illustrate how changes in the PLIT rates can affect their reporting decisions across tax bases. In the second part, we implement a difference-in-difference design to estimate the three main elasticities described by our model: intensive margin, extensive margin, and income-shifting elasticities.[3] The key identification assumption in our research design is that the outcome of interest, e.g., reported labor income, would have evolved similarly for treated and untreated TIEs in the absence of the tax reform. We present non-parametric evidence that supports this assumption. Finally, we use our model and the estimated elasticities to compute the efficiency costs of the 2012 tax reform.

Our empirical analysis yields three main results. First, we document a decline in the labor income reported by the TIEs, implying a sizable intensive margin elasticity of 0.577. This response is similar for wage earners and self-employed and can be explained by TIEs at the very top of the earnings distribution. Importantly, we obtain similar elasticities at the intensive margin under an alternative instrumental variable strategy in which we instrument the log change of the marginal net-of-tax rate by using a predicted tax rate constructed from income lagged two and three periods prior to the base year (Weber, 2014). Our results align with the findings

[3]See Saez et al. (2012) for a thorough discussion of the different strategies that typically have been used in the modern public economics literature to estimate elasticities in settings similar to ours.

of a meta-regression analysis (Neisser, 2021), which describes a typical range between 0 and 1 for this type of parameter.[4] Our analysis of mechanisms suggests that the response observed in the intensive margin is partly driven by changes in the number of hours worked by wage earners, while unilateral income underreporting does not seem to play a role. Second, we estimate a strong extensive margin response: a 1% decline in the average net-of-tax rate in the PLIT base leads to a 2.479 percentage points reduction in the probability of reporting earnings in that tax base. The response for self-employed individuals (4.471) is twice that of wage earners (1.914). The nature of this response is also different between these groups. While wage earners who respond to the reform vanish completely from the tax records, self-employed taxpayers shift their personal income to other tax regimes. This leads to our third main result. We find evidence of large income-shifting responses (-1.967) where taxpayers, mainly self-employed, shift their income from the PLIT base to the CIT base. This is not surprising given that Uruguay's tax code allows self-employed individuals to choose between labor or corporate income taxation. Moreover, we document that some self-employed TIEs anticipate the tax rate increase by moving their earned income across tax bases during the year when the reform was announced.[5]

Based on our theoretical model, we estimate that the welfare cost of the tax reform accounts only for 31.3% of the projected (mechanical) increase in revenues. Efficiency costs are mostly explained by extensive margin responses (81.1%) com-

[4]While our estimated intensive margin elasticity is slightly larger (0.577 compared to 0.287), it falls in the middle of the range of 0-1 estimated by this meta-study. Our larger estimates might be explained by our focus on TIEs, which are usually considered to have more possibilities to respond.

[5]This is consistent with Foremny et al. (2018), who show evidence of a large intertemporal income-shifting by liberal professionals——who mostly earn self-employed income——just before the implementation of a major tax reform in Uruguay in 2007.

pared to intensive margin responses (18.9%). Fiscal externalities such as income-shifting offset about 64.4% of this welfare cost. These results illustrate the importance of considering the widest arrange of possible behavioral responses to draw accurate conclusions about the welfare effects of tax reforms. From a tax policy perspective, our findings yield two lessons. First, the 2012 tax reform was not an inefficient strategy to increase tax revenues. Second, the tax administration efforts should be concentrated on reducing incentives to switch between, or even disappear from, tax bases (e.g., minimize tax loopholes and eliminate opportunities for arbitrage).

Our study contributes to the empirical literature that uses exhaustive tax administrative records to uncover behavioral responses to income taxation (e.g., Saez et al. 2012), and in particular to the analysis of TIEs' responses (e.g., Auten et al. 2016).[6] Our contribution to this strand of literature is two-fold. First, from a methodological perspective, the unusual variation in the top marginal tax rates within the top 1% creates a quasi-experimental setting with a clean control group helping to shield our empirical strategy from recurring threats that other studies face, such as mean reversion and heterogeneous income trends.[7] Furthermore, our setting allows us to isolate the effect of a change in the tax rates from changes in other characteristics of the tax regimes, such as the tax bases (see Kopczuk 2005 for a discussion of this issue). Second, from an empirical perspective, combining a natural quasi-experimental setting with access to detailed tax administrative records presents several advantages. To illustrate them, it is important to note

[6]Online Appendix, we discuss how our preferred estimates align with those from related studies.

[7]For instance, Giertz 2010 uses U.S. data and shows the elasticity of taxable income is highly sensitive to different model specifications that seek to control for mean reversion and divergence within the income distribution.

that the bulk of the existing literature on how TIEs react to personal income taxation has focused on specific margins of responses such as (reported income) *intensive margin* (e.g., Feldstein 1995; Auten and Carroll 1999; Gruber and Saez 2002; Kleven and Schultz 2014; Auten et al. 2016; Saez 2017; Miao et al. 2020), *income shifting* (e.g., Slemrod 1995; Goolsbee 2000; Kreiner et al. 2016; Gordon and Slemrod 2000; Pirttilä and Selin 2011; Alstadsæter and Jacob 2016; Harju and Matikka 2016), or the *extensive margin* (e.g., Kleven et al. 2013; Bastani et al. 2021), and that most of these studies have not exploited clean sources of exogenous variation and have relied on survey data for the empirical analysis. While all of them provide significant advances to our understanding of TIEs' behavioral responses to taxation, in our case, the use of rich administrative records allows us to overcome some of the critical measurement issues related to the use of survey data, especially to describe top-income earners (Barrett and Hamermesh, 2019). It also allows us to provide evidence of TIEs' responses across several margins in a unified framework. As we show in our welfare assessment of the reform, a narrow analysis of the behavioral responses might yield misleading conclusions.[8] Finally, having longitudinal information on individuals' incomes across different tax bases also enables us to dig into the drivers of these responses (e.g., labor supply vs. misreporting, anticipation, etc.).

Our paper also contributes to a burgeoning literature that uses quasi-experimental designs and administrative data to uncover behavioral responses to taxation in developing countries. While most of these studies focus on firm behavior, research

[8] In a paper related to ours, Waseem (2018) does investigate behavioral responses to personal income taxation in several margins of response and also quantifies the welfare costs of the reform. However, while Waseem (2018) analyzes firm behavior, we focus on a different agent: high-income individuals. Our paper is also closely related to Hermle and Peichl (2018) who propose a theoretical model of jointly optimal taxes for different types of income and provides evidence of differential responses to taxation depending on the income source (e.g., capital income or labor income) using administrative records from Germany.

examining the responsiveness of individuals to taxation remains scarce (Pomeranz and Vila-Belda 2019), and only a handful of studies have focused on TIEs (e.g., Tortarolo et al. 2020; Jouste et al. 2021).[9] This is particularly important in many developing countries that have made recent efforts to strengthen the redistributive capacity of their tax structures (Martorano, 2018; ECLAC, 2013) but that have been historically affected by high and persistent income inequality and limited enforcement capacity (Alvaredo et al., 2018; Gordon and Li, 2009).

The paper is organized as follows. Section 3.2 describes the Uruguayan tax system and the reform that gave rise to the top tax rates. Section 3.3 discusses the expected behavioral responses to the tax reform and presents the key aspects of the model we use to guide the empirical analysis. Section 3.4 describes all the sources of information used in the empirical analysis as well as the identification strategy. Section 3.6 analyzes the efficiency costs of the tax reform, and Section 3.7 concludes.

3.2 Institutional Background

3.2.1 The Pre-Reform Tax Structure

Uruguay is an upper-middle-income country with a population of about 3.5 million and a GDP per capita of USD 20,200 (PPP, 2015). The overall tax burden measured as a percentage of the GDP is 27.3%, which is relatively large compared to the Latin American and the Caribbean average (22.6%), but it is still lower than the OECD average (33%).[10] During the period of analysis, direct taxes

[9] In related research, Londoño-Vélez and Mahecha (2021) study behavioral responses of high-net-worth individuals to personal wealth taxes in Colombia.

[10] OECD.stats: https://stats.oecd.org/

represented around 35% of total tax collection and most of the remaining 65% correspond to a widespread VAT.[11] Tax collection from direct taxation can be divided into five components: (i) dual personal income tax where labor and capital income are taxed separately - for simplicity, we will call them Personal Labor Income Tax (PLIT) and Personal Income Tax on Capital (PITC) respectively -, (ii) corporate income tax (CIT), (iii) property tax, (iv) retirement income tax, and (v) non-resident personal income tax. Personal and corporate income taxes are the two major components of direct taxation, and both together account for about three-quarters of direct tax collections in equal parts.

Personal Labor Income Tax (PLIT). The PLIT progressively taxes all sources of individual labor income, both for wage earners and self-employed individuals. It comprises a *labor income tax part* and a *tax deduction part*, where the final tax liability is calculated as the difference between the two. The *labor income tax part* is the result of passing the total gross labor income through a set of income brackets with progressive marginal tax rates. Panel (a) in Table 3.1 reports the six income brackets and marginal tax rates (that range between 0% and 25%) for the pre-reform year 2011.[12] Figure 3.1 plots the PLIT structure and the gross labor income distribution, and illustrates that only middle- or high-income individuals pay PLIT.[13] The top two income brackets - the ones affected by the reform - overlap almost perfectly with the top 1% of the gross labor income distribution. The *tax deduction part* is reported in the second part of Panel (a) in Table 3.1. Deductions

[11]See Online Appendix for more details about Uruguay's tax structure.

[12]Since brackets are adjusted annually by CPI, we should not expect "bracket creep" (Saez, 2003).

[13]According to Uruguay's Tax Agency, only about 30% of registered employees paid PLIT between 2009-2015. A large exemption threshold is not atypical in developing countries, as shown in Jensen (2022)

are comprised of itemized and non-itemized deductions. Altogether, these are passed through a progressive deduction rate schedule (with rates between 10% and 25%). The resulting amount is subtracted from the income tax part, which yields the final tax liability.

Personal Income Tax on Capital (PITC). The PITC taxes all sources of individual capital income. It is based on a set of proportional tax rates, and deductions are not allowed. Table 3.1 in the Appendix, Panel (b), shows the 2011 tax rates. The tax code distinguishes twelve capital income items, which can be grouped into three broad categories: interest from deposits (taxed at a 3% rate), dividends and other financial income (7%), and real estate rents (12%).[14] [15] The tax code only requires capital income earners to file a tax return if they have not been subject to withholding. Furthermore, individuals do not need to report capital gains from bank-deposit interests (due to bank secrecy rules) or distributed dividends from anonymous companies. In these cases, taxes are withheld by the financial institutions and paid to the tax agency anonymously. Hence, a substantial share of capital income cannot be linked to specific individuals. This prevents us from using micro-level data to estimate the effects of the tax reform on reported capital income. However, we conduct some exploratory analysis using aggregate data to illustrate PITC trends before and after the reform.

Corporate Income Tax (CIT). Self-employed workers operating as unincorporated firms (sole proprietorship or partnerships) and earning less than a certain

[14]Interest from deposits includes all cash or in-kind rents coming from bank deposits and other financial assets. Other financial income includes dividends and royalties, among others.

[15]It is worth noting that dividends and other financial income are taxed at the corporate level at a 25% rate before being distributed to individuals. Hence, the effective rate for this type of income is about 30%.

threshold (UYU 8,948,000 in 2011) can choose to pay CIT instead of PLIT.[16] Panel (c) in Table 3.1 shows the structure of the CIT. It consists of a statutory 25% tax rate applied to business profits. For self-employed workers who opted for the corporate tax regime, the 25% tax rate is applied to 48% of their (gross) earned income. This results in an effective tax rate of 12%. Exercising the option between PLIT and CIT has no major administrative costs other than filing an application form. However, once a self-employed worker opts for CIT, she is prevented from switching back to PLIT for three years. Figure 3.2, panel (b) shows the total number of workers within the top 1% of the gross labor income distribution that were allowed to choose between labor and corporate taxes in 2011 and the total number of workers who ended up opting for the latter (about 15%, on average). This figure also depicts how the gap between the average corporate and labor income tax rates closes when the gross self-employment income increases until a point where the average CIT rate becomes smaller than the average PLIT rate. This explains why 50% of self-employed workers located in the upper-income bracket choose the CIT regime compared to a 2% share for top income earners with lower income.[17]

3.2.2 Top Income Earners in Uruguay

Throughout the world, TIEs capture a large share of total income. In developed countries, this is reported in the seminal studies by Piketty (2003); Atkinson (2007); Atkinson et al. (2011). This is also the case in Latin America, where the burgeoning literature shows even higher concentration levels (Alvaredo, 2010;

[16]Henceforth, we express all money metrics in 2011 constant UYU. For reference purposes, the UYU/USD exchange rate in 2011 was 19.02.

[17]It is important to note that there is not a clear-cut threshold valid for all individuals as a result of different sets of deductions they may apply, or because the decision was made in previous periods.

Alvaredo and Londoño Velez, 2014; Flores et al., 2020; Morgan, 2017). Recent estimates illustrate a very similar picture for Uruguayan TIEs who receive about 15% of all income (Burdín et al., 2022). While among the lowest in Latin America, this share is similar to that of other higher-income countries such as the U.S.. Capital and business income have played a major role as drivers of income concentration both in the U.S. and other rich countries (Piketty et al., 2018), as in Latin America (De Rosa et al., 2020). Uruguay is not the exception: business income for individuals in top fractiles can reach up to 40% of earnings after considering all sources of capital income and imputing a large amount of the anonymous capital income (Burdín et al., 2022).[18]

3.2.3 The 2012 Tax Reform: Changes in the PLIT

The PLIT was first introduced in 2007 but was subject to important changes in 2012 that exclusively affected individuals in the top 1% of the labor income distribution (Figure 3.2). More specifically, the tax reform split the top two brackets into three, with a new top marginal tax rate of 30% compared to the former top tax rate of 25%. The reform created four groups of TIEs based on their pre-reform income:

[18] A more detailed discussion about the characteristics of top income earners in Uruguay, and in particular about the composition of their income, can be found in Online Appendix

		Marginal tax rates	
Group	Income	Pre-reform	Post-reform
G1	$\in [1.4M, 2M]$	22%	22%
G2	$\in (2M, 2.7K]$	22%	25%
G3	$\in (2.7M, 3.1M]$	25%	25%
G4	$\in (3.1M, y^{max}]$	25%	30%

Figure 3.2, panel (a) depicts the pre- and post-reform labor income tax schedule and these four groups (G1, G2, G3, and G4). Groups G2 and G4 faced an increase in their marginal tax rates from 22% to 25% and from 25% to 30%, respectively. This implies a decrease in the marginal net-of-tax rates of about 4% and 7%, respectively. Groups G1 and G3 did not face any change in their marginal net-of-tax rates. Figure 3.2 panels b. and c. illustrate the pre- and post-reform *average* tax rates for self-employed and wage earners separately. While the tax reform did not affect the PLIT average rate for taxpayers in the G1 group, it increased it for individuals in G2, G3, and G4. The magnitude of this change is increasing in income and can reach up to 3% at the very high end of the labor income distribution. In contrast, the average CIT rate remained unchanged. This is important for self-employed workers who can choose to pay either PLIT or CIT.

Some questions may arise about whether this is an adequate setting to estimate behavioral responses to taxation. One concern is whether the changes introduced are large enough to estimate behavioral responses to taxation (Chetty, 2012). Even though the change in the marginal net-of-tax rates appears to be small (see Table 1 in Chetty 2012 for a comparison), the changes in the marginal and average tax rates are not. For instance, the 2012 tax reform implied an average increase of

17% and 8% in the marginal and average tax rates for the top 1%, respectively, relative to the 2011 tax rates. Another concern could be that the reform was not salient. However, the political debate surrounding the 2012 tax reform, the media coverage, and the type of workers affected made the reform highly salient.[19] In addition to salient, the changes introduced by the reform were simple, especially for TIEs who are arguably sophisticated taxpayers. This is important because simpler reforms could more likely trigger behavioral responses than reforms involving more intricate changes. In sum, the tax reform was salient, fairly simple to understand, and similar in size to other reforms studied in the literature (e.g., Saez 2017). This creates a compelling setting to understand how TIEs react to taxation.

Finally, it is also important to note that taxpayers might have anticipated some elements of the reform since it was announced a year before its implementation. Figure 3.3 depicts the timeline. The reform was announced on September 12^{th}, 2011 and approved in May 25^{th}, 2012 (Law 18.910). However, it was applied retroactively, starting in January 1^{st}, 2012. The ruling party had an absolute majority of votes in both houses of the Parliament. Hence, once the tax reform was officially announced, taxpayers might have been convinced that the reform would be approved, although not at the precise moment. Thus, it could be argued that some TIEs might have anticipated the tax reform a (fiscal) year ahead of its implementation. We will return to this point in Section 3.5.4, where we discuss potential anticipatory effects.

[19]For instance, the party in government included multiple references to changes in the upper part of the PLIT schedule in its political platform. This became a significant topic in the presidential and legislative elections. Furthermore, the debates about the tax reform generated a political confrontation between the president and vice president that drew the general public's attention. Nationwide, TV channels and newspapers exhaustively covered this. See Online Appendix for more details about the salience of the tax reform

3.3 Conceptual Framework

In this section, we discuss the intuition of how income reporting and tax base choices may be affected by changes to the top income tax rates of one tax base. This is formalized in a simple model that combines elements from Waseem (2018) and Zawisza (2019), and we describe in detail in Online Appendix. Next, we provide an intuition of the basic characteristics of the model and the predictions that will guide our empirical analysis.

In our model, heterogeneous individuals decide how much to consume, the number of hours worked, the amount of unreported income, and one of the three mutually exclusive options for tax reporting: PLIT base (L), CIT base (C), or a broadly defined outside option (O) that includes, for instance, retirement, labor informality (Waseem, 2018), but also more sophisticated responses such as fiscal migration (Kleven et al., 2013).[20] For each tax base, there is a tax schedule that defines the individual tax liability under each tax base choice. For simplicity, we assume linear schedules for all the alternatives.[21] There are also increasing base-specific utility costs of working and evading taxes that account for differences in the processes that generate the income associated to each tax base.[22]

[20]As described in Section 3.2.1, we cannot reliably estimate the effects of the tax reform on PITC reporting. To simplify and better connect our model to the empirical analysis, we omit PITC as one of the alternative tax bases. The only implication, if anything, is that our welfare analysis will provide an upper bound of the actual welfare costs of the reform since we are excluding one margin of response that would reduce the welfare losses associated with the PLIT rates increase.

[21]While the Uruguayan PLIT consists of a progressive schedule, one could think of this model as a model of decision conditional on being on a given bracket

[22]It is important to note that evasion costs account for the time and effort expended by individuals to evade taxes or other less traditional costs such as psychological or moral costs of illegal behavior. In any case, we assume that these are mostly resource costs that are not transferred between economic agents, such as penalties or fines. See Chetty (2009) for a detailed discussion about resource and transfer costs of evasion. It is also important to acknowledge that in a more general model - e.g., with poor and rich people - it would make sense to introduce

Our model can be solved in two steps. First, the individual evaluates the optimal labor and evasion choices conditional on the tax base. In equilibrium, the marginal cost of working must equal the utility gain in consumption, i.e., the net-of-tax rate. Similarly, the marginal cost of under-reporting income must equal the marginal benefit of evading taxes, i.e., the marginal tax rate. Second, individuals decide between alternative tax bases. Hence, the equilibrium comprises an income supply function (a combination of labor supply and evasion decisions) and a tax base choice rule. Next, we describe the predicted effects of a change in the PLIT rate for each relevant response margin.

Intensive Margin Response. An increase in the PLIT rate reduces the marginal net-of-tax rate for those who choose the PLIT, which can lead to multiple types of response in the intensive margin (i.e., conditional on reporting in the PLIT base). First, individuals may decide to reduce the number of hours dedicated to labor-market activities (real labor supply response). Second, it could also incentivize tax evasion practices. Moreover, some taxpayers may decide to re-classify part of their labor income into other forms of income whose tax rates were unaffected by the tax reform.[23] The predicted direction of the effect is the same for any of these mechanisms: an increase in the PLIT rate is expected to reduce the reported income to the PLIT tax base.

Extensive Margin Response. An increase in the PLIT rate could also induce responses in the extensive margin. Decisions in the extensive margin require comparing the total tax liabilities across all potential bases. Hence, unlike for intensive

---

heterogeneous evasion technologies.

[23]In addition, some responses might also include changing the extent of rent-seeking behavior. Some individuals (e.g., managers) may have substantial pay-setting power, and a higher marginal tax rate would make their rent-seeking less profitable.

margin responses, here, the relevant tax rate is the *average* tax rate instead of the *marginal*.[24] There are several types of extensive margin responses. First, some individuals could exit the PLIT base towards other relatively more attractive tax bases (i.e., income shifting). Second, taxpayers could also switch entirely to the local informal sector and hide completely from the Tax Agency. Third, they could also move toward temporary inactivity by engaging in income re-timing, which leverages tax rates' differences over time. Finally, individuals could also engage in international fiscal migration or outright tax evasion by using offshore accounts. While all of these represent potential extensive margin responses, only some of them are plausible and can be tested empirically. We discuss this in detail in Section 3.5.

Income-Shifting Response. Income-shifting responses are the flip side of the extensive margin responses. In particular, the type of income-shifting responses considered in this model correspond to TIEs who switch their *entire* income from the PLIT base to the CIT base to benefit from more favorable *average* tax rates. [25]

Efficiency cost of taxation. Based on this simplified model, and as proved in Online Appendix, the overall change in welfare due to the tax reform is described by

[24]While this distinction is irrelevant in this simplified model because we assumed a linear tax schedule, it will be important for the empirical estimates that will be used to assess the welfare effects of the tax reform.

[25]While our model only allows for mutually exclusive tax base choices, the Uruguayan tax code actually allows individuals to report income in multiple tax bases. This introduces an additional type of response: Individuals can shift only *part* of their income (intensive margin income-shifting response). However, it is worth noting that self-employed workers who decide to shift to the CIT base must change some legal aspects of their business organization (i.e., to create a firm as a sole proprietorship or partnership). Since these are fixed costs, the type of income-shifting response we expect and focus on is extensive margin income shifting. Indeed, in our empirical setting, only 4% of self-employed TIEs——those who can choose between CIT and PLIT taxation——report income in both tax bases.

equation 3.1 where s indexes the tax base $\{L, C, O\}$ and Y_s denotes the aggregate income reported to tax base s:

$$\frac{dW}{d\tau_L} = \frac{dB}{d\tau_L} = \tau_L \times \frac{dY_L}{d\tau_L} + \tau_C \times \frac{dY_C}{d\tau_L} \tag{3.1}$$

It is important to note that $\tau_L \times \frac{dY_L}{d\tau_L}$ can be thought of as the combination of the intensive margin response - i.e., the revenue loss associated with a reduction in reported income to the PLIT base, conditional on staying in this tax base - and the extensive margin response - i.e., the revenue loss from workers that leave the PLIT base -. The behavioral response $\frac{dY_C}{d\tau_L}$ captures the income-shifting response. Similar to Waseem (2018), we can prove that:

$$\frac{dW}{d\tau_L} = -\frac{\tau_L}{1-\tau_L} \times Y_L \times [\bar{\varepsilon}_L + \bar{\mu}_L + \bar{\eta}_{L,C}] \tag{3.2}$$

which shows that the welfare change due to a small increase in τ_L can be written as a combination of responses in the intensive, extensive, and income-shifting margins. In particular, $\bar{\varepsilon}_L$ is the income-weighted average intensive margin elasticity of the reported income y_L with respect to changes in the tax rate τ_L, across individual types. This term captures the reduction in total tax revenues from the PLIT base due to reduced reported labor earnings. $\bar{\mu}_L$ and $\bar{\eta}_{L,C}$ represent the revenue-weighted average (semi-)elasticities in the extensive and income-shifting margin, respectively. These (semi-)elasticities capture the change in the share of individuals reporting to tax bases L or C, with respect to changes in τ_L.[26]

Finally, it is important to recall here that, conceptually, the relevant tax rate for intensive margin responses is the *marginal* tax rate, while for extensive and

[26] Online Appendix includes the proof of this result and presents more details about the weights required to calculate the weighted-average elasticities.

income-shifting responses, the relevant rate is the *average* tax rate. While in this simplified model, both *marginal* and *average* PLIT tax rates are the same (τ_L), with more complex tax schedules (e.g., progressive tax rates), these will be different. Since the 2012 Uruguayan tax reform that we analyze in this paper is based on a progressive PLIT schedule, the empirical analysis will be conducted using either the change in the marginal tax rate or the change in the average tax rate, depending on the margin of response considered.

3.4 Data and Research design

3.4.1 Data Sources and Sample Restrictions

Our analysis is based on administrative records provided by Uruguay's Tax Agency (*Direccion General Impositiva*) and the Social Security Administration (*Banco de Prevision Social*, henceforth, SSA) that can be linked to each other at the individual level. The richness of the tax records allows us to explore the three margins of response discussed in the previous section and to analyze in greater depth whether reporting or labor supply decisions can explain potential changes in reported income. Next, we describe in detail the information from each data source that will be used in the empirical analysis.

Administrative Tax Records

Administrative tax records cover the universe of registered wage earners and self-employed individuals between 2009-2015. They include information on all possible personal income tax bases before (2009-2011) and after the tax reform (2012-2015): 1) PLIT records, 2) CIT records, 3) PITC records, 4) employer statements on employee activities, and 5) firm-level data.[27] First, the information contained in PLIT records is similar to the information reported in the 1040-form in the U.S.: earned income, tax withholdings, deductions, and every single item used to determine the final tax liability. In addition, it also includes gender and date of birth. Second, the CIT records are comprised of the annual tax returns of all self-employed individuals that opted for the CIT base instead of PLIT base. The only difference with PLIT records is that income is reported as firm profits instead of wages or self-employed income. Third, PITC records contain information on capital income, such as profits, dividends, and real estate rents. As explained before, some of these items cannot be attributed to a specific individual (e.g., dividends from non-nominative shares). Fourth, similar to W-2 forms in the U.S., employer statements on employees' activities contain similar information to PLIT records but are reported by employers. Finally, firm-level records contain information on the number of employees, the firm's age, location, industry, and sector.

We linked this information at the individual level to create a longitudinal dataset that tracks individuals' income sources between 2009-2015. Our analysis sample is restricted to the subset of individuals whose gross labor income was at least in the fifth income bracket - the first income bracket affected by the reform - in all the pre-reform years (2009-2011). This group represents approximately the top 1% of the labor income distribution (see Figure 3.2). Because the 2012 reform targeted individuals in the top 1% of the *labor* income distribution, we exclude TIEs whose income before the reform is comprised of *business* or *capital* income almost exclusively. Regardless, 98% of the sample used in this study belongs to the 1% of the *total* income distribution. Online Appendix provides a more detailed discussion and description of different types of TIEs in Uruguay. In addition, we

[27] Online Appendix contains detailed information about each one of these.

also exclude individuals younger than 25 years old or older than 70 years old, and individuals who pay retirement income tax. Finally, we exclude individuals who switched treatment status in pre-reform years to improve similarity across treatment groups in terms of the volatility and growth of their past earned income. The resulting dataset (hereafter referred to as *TAX* sample) is an unbalanced panel of 28,434 observations for the 2009-2015 period corresponding to about 4,062 TIEs with a pre-reform earned income (y_{it}) in the range [1,336,000-y^{top}]and post-reform earnings in the range [0-y^{top}].

SSA Records of Earnings and Employment.

We complement our analysis with employer-employee individual-level administrative records from Uruguay's SSA. These records contain earnings, hours, and days worked for the universe of workers who were registered with the SSA for at least one month in 2007-2014. The lack of information on hours or days worked is an important limitation of the empirical literature on labor supply responses to taxation. We use this information to conduct additional exploratory analysis on the reasons explaining reported income changes. Furthermore, SSA's records allow us to extend the pre-reform period by two years (2007-2008), which will help assess the validity of our empirical strategy. SSA records can be linked at the individual level to a subset of the *TAX* sample using a masked version of the national identification number. However, this is only possible for individuals who have a partner or children covered by the National Health Insurance. Hence, single individuals without underage children are excluded in the analyses using SSA records.[28] The resulting matched *TAX-SSA* records database accounts for around 62% of the *TAX*

[28]We also excluded individuals with inconsistent information across datasets. For instance, individuals with zero or missing hours or incomes in the SSA record but positive income reported in the TAX data.

sample and 45% after applying the additional filters.[29]

3.4.2 Identification Strategy

The 2012 tax reform induced variation in tax rates across taxpayers within the top 1% labor income earners. This is extremely rare, and most existing studies have relied on broader tax reforms that affect individuals across the whole earnings distribution.[30] Furthermore, having alternate treatment (G2, G4) and control groups (G1, G3) also allow us to address reasonable concerns about heterogeneity within the 1% itself. Second, even within the TIEs, the tax reform creates a "pure" control group where the marginal and average tax rates remain unchanged. This is not the case in tax reforms affecting tax rates across the whole income distribution. Finally, the 2012 tax reform did not affect the PLIT base definition or any other features of alternative tax bases. Hence, our estimates can be interpreted directly as causal effects of changes in tax rates, as opposed to broader reforms that change several features of the tax schedule.

We estimate the effects of the 2012 tax reform based on a difference-in-differences design that compares TIEs that faced an increase in the tax rates on the PLIT versus TIEs that did not, over time.[31] Because of PLIT's progressivity, the outcomes

[29]More details about the TAX-SSA matched sample are reported in Online Appendix. Specifically, we present summary statistics and more descriptive details for the main and secondary analysis samples based on the *TAX* and *TAX-SSA* records, respectively.

[30]For instance, typical approaches in the literature use the top 5%-less top 1% or the top 10%-less top 1% as control groups, which are arguably less comparable to the top 1%. In Online Appendix, we document that the income shares of the groups affected by the reform remained stable over the period of analysis, while the top 5% and 10% income shares continuously declined. This illustrates the issues that could arise when individuals in the 90-99 or 95-99 percentiles are used as the control groups for individuals at the top 1% and highlight one of the main strengths of our empirical setting

[31]This is a widely used approach in labor and public economics literature that estimates the sensitivity of the tax base to changes in the tax rates. Saez et al. (2012) provides a thorough

of interest (e.g., reported labor income) and the tax rates are defined simultaneously. Hence, as is typical in the related literature, we rely on the pre-reform labor income to define the treatment status (Saez et al., 2012). Moreover, our definition of treatment requires individuals to be in the same group (treatment or control) in the three consecutive pre-reform years (2009-2011). This allows us to reduce concerns associated with the natural volatility of the income variable (e.g., mean reversion or economic cycle effects).

As discussed in Section 3.3, to analyze intensive margin responses, the relevant rate is the *marginal* tax rate. Therefore, our treatment group ($treat_i^{MTR}$) consists of TIEs that, given their pre-reform labor earnings, faced an increase in their marginal tax rate (i.e., G2 and G4 groups in Figure 3.2, panel (a)). Analogously, the control group is defined by taxpayers whose marginal tax rates remain unchanged (i.e., G1 and G3). The relevant treatment variable for extensive margin or income shifting responses is the *average* tax rate. In this case, the treatment group ($treat_i^{ATR}$) consists of TIEs that, given their pre-reform labor earnings, faced an increase in their average tax rate (i.e., G2, G3, G4 in Figure 3.2).[32]

The key identification assumption is that the outcomes of treated and control TIEs would have evolved similarly in the absence of the tax reform. This assumption would be violated if non-tax-related shocks affected the trends in the outcomes of interest for treated and control groups differently. Indeed, because we define the treatment status based on the pre-reform labor income, this might be somewhat plausible (Saez et al., 2012). However, we present several pieces of evidence to

survey of related studies

[32]While the tax reform did not affect the *marginal* tax rates for G3, it did affect their *average* tax rate, and therefore G3 is not a "pure" control group in the intensive margin analysis. We provide additional estimates that test the robustness of our results to the exclusion of this group from the intensive margin estimates.

show this is not true. Next, we discuss broadly some of the most general potential concerns about our identification strategy and how we rule them out. In addition, throughout Section 3.5, we report compelling graphical evidence that supports the parallel trends assumption in the pre-treatment period, and in Online Appendix, we discuss additional robustness tests for each margin of response.

A first concern could be the existence of non-tax-related changes in the income distribution between 2009-2016 that are correlated with income and, therefore, could have affected treatment and control groups differently. In our context, this is highly unlikely since the top 1%, 0.5%, and 0.1% income shares - i.e., income groups affected by the reform - remained stable over the period. This is consistent with findings reported in Burdín et al. (2022).[33]

Another concern could be that the parallel trends assumption holds by construction, given that our sample of interest comprises individuals that were at least in the fifth income bracket of the tax schedule in all the pre-reform years. This could select individuals with relatively more stable income into our final sample, affecting estimates of the intensive margin response. Furthermore, in the extensive margin analysis, the parallel trends assumption might hold exclusively by construction since we require individuals to be in the PLIT base during the whole pre-treatment period. Online Appendix shows that this should not be a major concern since there is still compelling evidence that supports the parallel trends assumption even when using the *TAX-SSA* sample that contains two additional pre-treatment years that were not used to define the treatment variable in the *TAX* sample.

An alternative threat to identification could be endogenous changes in the

[33]Online Appendix contains more details as well as the supporting evidence used for the discussion

composition of treatment and control groups. The use of longitudinal data and the fact that we construct the treatment variable based on the three pre-reform years should minimize these concerns. In addition, Online Appendix shows that even though treatment and control groups are not perfectly stable over time, a large share of individuals (for instance, 80% in the first post-treatment year) remain in the same group throughout the period. Moreover, among those who move, there does not seem to be evidence of differential dynamics between groups, which is reinforcing given our difference-in-differences approach. Our empirical analysis also reports that our results hold when we exclude switchers from our sample.

Finally, as typical in this literature, one might still be worried that treatment assignment is based on the outcome variable. In this regard, for all estimates in the main analysis, we report results where treatment is defined based on an exogenously *predicted* labor income rather than *actual* labor income. Our results hold qualitatively, although more imprecisely estimated. We also implement the alternative instrumental variable strategy proposed by Weber (2014), which yields similar results (see Online Appendix).

3.5 Results

3.5.1 Intensive Margin Response: Main Results

Our main outcome variable in the analysis of intensive margin responses is gross labor income. The gross labor income measure is closer to the concept of broad income (i.e., before deductions) than to the taxable income (i.e., after deductions) and allows us to disregard changes in behavior that are driven by changes in deductions.[34] Since our focus in this section is on the behavioral responses in the intensive margin, we restrict our sample to individuals that reported positive income in the PLIT base.

The reduced-form effect of the tax reform can be obtained using the following regression that estimates a difference-in-differences parameter:

$$\Delta \log y_{it} = \alpha + \beta treat_i^{MTR} \times \mathbb{1}(year = 2012) + X_i\delta \ + \ \lambda_t + u_{it} \tag{3.3}$$

where $\Delta \log y_{it}$ is the log change in the gross labor income for individual i between years t and $t-1$, $treat_i^{MTR}$ indicates whether taxpayer i belongs to a group that faced a marginal tax rate increase (i.e., G2 or G4), and $\mathbb{1}(year = 2012)$ is an indicator variable for the year of the tax reform, 2012. The coefficient of interest β identifies the reduced-form effect of the 2012 tax reform on reported gross labor income. To improve precision, we include a set of time-invariant individual-level control variables (X_i) defined in the pre-reform period and year fixed effects λ_t.[35] As discussed in Section 3.3, all regression estimates on intensive margin responses are weighted by income. Standard errors (u_{it}) are clustered at the individual level.

To obtain the intensive margin elasticity, we follow the standard approach in related literature (e.g., Auten and Carroll 1999; Gruber and Saez 2002; Kleven and Schultz 2014) and estimate the following model using two-stage least-squares

[34] Moreover, under certain conditions, the elasticity of broad income is the relevant statistic to define the top optimal tax rate and for welfare analysis(Chetty, 2009; Doerrenberg et al., 2017; Saez et al., 2012).

[35] The control variables include age, gender, employment sector, an indicator for receiving profits and dividends, firm size and type (public/private), and, sometimes, a self-employment income indicator.

regressions:

$$\Delta\log y_{it} = \alpha + \varepsilon\Delta\log(1 - \tau_{it}^{MTR}) + X_i\delta \ + \ \lambda_t + v_{it} \tag{3.4}$$

where $\Delta\log(1 - \tau_{it}^{MTR})$ is the log-change in the marginal net-of-tax rate between t, and $t-1$.[36] In this case, the parameter of interest (ε) represents the intensive margin elasticity, i.e., the percentage change in gross labor income as a result of a 1% change in the marginal net-of-tax rate. To address the endogeneity between the log changes in the marginal net-of-tax rate and the gross labor income, we exploit the 2012 tax reform and instrument $\Delta\log(1 - \tau_{it}^{MTR})$ using the difference-in-differences (DinD) interaction term: $treat_i^{MTR} \times \mathbb{1}(year = 2012)$. Again, all estimates are weighted by income such that ε matches $\bar{\varepsilon}_L$ in Equation 3.2. Since treatment status is based on individuals' pre-reform earnings, this specification estimates the intention-to-treat short-run intensive margin elasticity.[37] Finally, it is important to note that our treatment definition is given by a dichotomous variable, and all units are simultaneously affected. Hence, the interpretation of our estimates should not be subject to recent concerns affecting TWFE and DinD estimates in contexts of heterogeneous effects (e.g., De Chaisemartin and D'Haultfoeuille 2022).

Graphical evidence. Panel a. in Figure 3.4 depicts the evolution of the log labor income - normalized to 0 in 2011 - for treatment and control groups separately, while panel b. depicts the year-by-year difference between these two groups. These

[36]Because the sample in the analysis of PLIT responses is restricted to individuals who report positive income in the PLIT tax base, the log-specification does not exclude any observation.

[37]The ITT estimate will be a lower bound for the treatment on the treated (TOT) estimand if TIEs switched tax treatment status across years (see Kawano et al. 2016 for a further discussion). Figure 3.6 shows that almost 80% of the TIEs remain in the same income ranges that defined their treatment status over the analysis period.

figures are based on unweighted estimates to illustrate the reduced-form effects of the reform as transparently as possible. Two conclusions can be drawn from this simple but intuitive graphical analysis. First, there is no evidence of differential pre-trends, which is critical to validate the DinD approach. Moreover, in Online Appendix, we show that this holds even when extending the pre-treatment period with years that are not used to define treatment and control groups. This allows us to rule out parallel trends being held by construction.

Second, a strong diverging pattern in the reported income between the two groups arises immediately after the reform. This reduced-form effect indicates that the tax reform negatively affected the income reported by TIEs that faced increases in their marginal tax rates relative to TIEs in the control group. Figure 3.5 replicates the analysis for wage earners and self-employed individuals separately, showing that both groups seem to respond in the same direction and magnitude. In the literature, self-employed workers are usually considered the most responsive group for different reasons, such as more flexibility to adjust labor supply and increased tax planning and evasion opportunities. In the next section, we explore this further and show that the similarly sized responses that we observe here are due to a composition effect and that using a balanced sample of individuals who always reported to the PLIT yields the typical larger responses on self-employed workers.

Regression evidence. Table 3.2 reports our baseline elasticity estimates based on the econometric analysis. Panel A. focuses on the reduced form effect, while panels B. and C. report the first-stage estimates and the labor income elasticities, respectively. In columns, we present the results for three different samples: all TIE workers, wage earners, and self-employed. Columns 1, 4, and 7 report estimates from our baseline specification based on Equations 3.3 and 3.4, while other columns present additional results for robustness. First, and consistent with the graphical evidence, Table 3.2 shows a reduced-form effect of the 2012 tax reform of -0.026 log points for the full sample (p-value <0.05). In addition, Table 3.2 shows that TIEs affected by the reform saw their marginal net-of-tax rates decrease by 0.045 log points compared to TIEs unaffected by the reform. Together, these results imply an intensive margin elasticity (ε) of 0.577 (p-value <0.05). This result can be interpreted as follows: a 1% decrease in the marginal net-of-tax rate reduces the reported labor income of TIEs by 0.577%. Consistent with the preliminary evidence discussed in the graphical analysis, there are no differences in the direction and size of the responses by type of worker.[38]

For robustness, column 2 reports an alternative specification where the treatment variable is defined based on the predicted labor income, while column 3 combines our basic DinD specification with a non-parametric coarsened exact matching (Iacus et al., 2012) that creates more comparable treatment and control groups.[39] Both strategies yield qualitatively similar results, although the effects reported in column 2 are slightly larger in size. This is mostly explained by a first-stage coefficient that is considerably smaller, which is reasonable given that the relevant groups are now defined based on a predicted rather than observed income. The patterns observed in the robustness tests for the full sample are also observed when splitting the sample by wage earners and self-employed workers.

[38]While there are some differences in the statistical significance of the estimates, these are most likely due to the substantially smaller sample size of the self-employed group.

[39]The intuition behind this approach is that matching improves the comparability based on observable characteristics, while the DinD will "difference-out" unobserved differences (see, e.g., Blundell and Dias, 2009). Specifically, we first match individuals coarsely using pre-reform year characteristics and determine the matching weights in one year of data since it provides a more stable formula for determining matching weights than the full period. The pre-reform characteristics are age, gender, filing a tax return, number of jobs, firm size, sector of employment, public/private employment, and receiving self-employment income. We match about 80% of the original analysis sample.

In Online Appendix, we conduct several robustness exercises aimed at addressing concerns related to 1) the use of pre-treatment labor income to define the treatment variable (mean reversion), 2) compositional changes (i.e., entries and exits from the PLIT base), 3) responses driven by the control group and individuals with very specific characteristics (e.g., extreme TIEs or TIEs who can engage in *partial* income-shifting), 4) confounding effect of heterogeneous macro shocks at the industry level, 5) alternative definition of the treatment variable with a more flexible definition of the treatment group, and 6) instrument selection (Weber, 2014). Our main findings are broadly robust to these additional checks but with three qualifications. First, including additional income controls reduces the magnitude of the estimated elasticities. This is somewhat common in previous related studies and suggests that too many base year control variables with only three pre-treatment years absorb much of the variation in the tax rates (Giertz, 2010; Saez et al., 2012). Second, the estimated intensive margin responses are considerably smaller and statistically insignificant when we replicate our baseline estimates excluding top 0.1% TIEs (about 20% of individuals in the G4 group), indicating that the aggregated response could mask important heterogeneity across individuals. Consistent with this, we also report evidence from an exploratory triple-difference strategy that shows that individuals in G4 seem to have a stronger response than individuals in G2. However, this analysis is underpowered. A potential explanation for this finding is that individuals in G4 were more strongly affected by the reform, as explained in Section 3.2.3. Finally, estimated intensive margin elasticities for self-employed workers are larger when using a balanced sample of individuals who typically report to the PLIT base. This helps us to reconcile the similarly-sized responses observed in our baseline specification with the typical finding in the literature that self-employed workers tend to be more responsive than wage earners.

These estimates suggest that self-employed individuals that, for some reason, are "trapped" in the PLIT base indeed seem to respond more than wage earners.

3.5.2 Unpacking Intensive Margin Response: Real Labor Supply Responses and Misreporting Effects

In this section, we leverage the *TAX-SSA* data to explore some potential mechanisms behind the intensive margin responses. In particular, we discuss whether TIEs respond through tax avoidance/evasion or real labor supply decisions. As described in Section 3.4.1, while the *TAX-SSA* contains additional valuable information, it does not perfectly coincide with the *TAX* sample. Hence, for comparison, columns 1, 4, and 5 in Table 3.3 replicate our preferred estimates in the *TAX-SSA* sample. Table 3.3 shows that individuals in the *TAX-SSA* sample are more responsive compared to the *TAX* sample, which could be explained by an over-representation of high-intensity treated individuals (G4) in the *TAX-SSA* sample. Despite these limitations, given the lack of studies using quasi-experimental variation to identify real labor supply responses to taxation, the results reported in this section still represent a step toward a better understanding of the mechanisms behind the reported income responses. [40]

[40] Some early studies have analyzed the labor supply responses of high-income individuals to tax reforms in the U.S. - e.g., (Eissa, 1995; Moffitt and Wilhelm, 2000) - finding mixed results. Most of this early literature cannot identify labor supply responses cleanly and usually rely on survey data, which for TIEs is more likely to be affected by measurement error and may lead to biased elasticity estimates (Barrett and Hamermesh, 2019). More recent work has tried to address some of these issues by using administrative micro-level records and has found evidence of large responses in part-time and secondary employment of low-income individuals in Germany - (Tazhitdinova, 2020b, 2022) - and a small labor supply elasticity on the overtime hours margin in response to a tax holiday for high-wage earners in Argentina (Tortarolo et al., 2020). In contrast to these studies, we identify tax-driven labor supply responses using cleaner quasi-random variation in the top marginal tax rates within the top 1% exclusively, and we address measurement error issues by using information about hours worked from a unique administrative dataset that links tax records with employer-employee social security data.

First, we explore if intensive margin responses could be explained by income under-reporting. One way in which taxpayers can misreport income is by manipulating their tax returns. While this behavior is typically flagged and prevented by third-party reporting, its effectiveness is limited by the institutional capacity of enforcement.[41] Columns 2, 5, and 8 in Table 3.3 replicate our preferred specification but use the income reported by employers to the SSA. Responses estimated using self- and third-party reported income are similar in size and magnitude, which is inconsistent with the idea of unilateral underreporting. For self-employed workers, who are presumably less subject to third-party controls, the intensive margin response in the tax records is indeed slightly larger, but the difference is statistically insignificant. Overall, these results indicate that unilateral underreporting is far from being the main channel of intensive margin response.[42]

Using SSA records, we can also test whether intensive margin responses could be explained, at least partially, by real labor supply adjustments. Figure 3.7 and Table 3.3 replicate our main graphical and econometric analysis using the log weekly hours as the outcome variable.[43] For the pooled sample, regression estimates report evidence of a small and statistically insignificant reduced-form effect of -0.009, which translates into an intensive margin elasticity of 0.181. This result masks some heterogeneity by income source. For wage earners, the elasticity

[41] Indeed, some related studies for Uruguay have shown evidence of unilateral income under-reporting by comparing self- and third-party reported income (Bergolo et al. 2020a, 2021).

[42] An alternative mechanism for income under-reporting is through collusion with employers. While employers have no direct incentives to collude, TIEs are likely to have high-profile positions and some degree of influence on how the firm reports earned income to the Tax Agency. For instance, Kreiner et al. (2016) highlight that cooperation from employers - who report employees' earnings to the tax authorities - emerges as an important factor in explaining their findings of inter-temporal shifting responses in wage income to a tax reform in Denmark. Unfortunately, we are unable to test this hypothesis with our current data

[43] Online Appendix describes how this variable is constructed, discusses concerns about its quality, and presents evidence supporting the reliability of this measure in our setting.

is 0.711 and statistically significant at a 10% level. On the contrary, the elasticity has the opposite sign for self-employed individuals but is very imprecisely estimated and uninformative, given the extremely reduced sample size.[44]

In sum, labor income responses of TIEs to the tax reform may have been driven at least partially by changes in the number of hours worked in the case of wage earners, while for self-employed TIEs, the evidence is inconclusive. Three additional points are worth to be made when interpreting these results. First, because of how the hours worked variable is constructed (see Online Appendix), these effects can be explained by a reduction in hours or days worked within a job or a reduction in the number of jobs. Second, it is possible that part of the change in reported hours is due to a collusive employer-employee arrangement. Given Uruguay's labor market regulations that do not allow reductions in workers' compensation without changes in hours worked, collusive employer-employee agreements should mechanically translate into changes in reported hours. Finally, one potential concern is that hours worked might not be informative about the earned income of TIEs and, hence, a somewhat irrelevant margin of response at the high end of the income distribution. However, it is important to note that in the case of Uruguay, unlike other richer countries, liberal professionals and workers in the health sector represent a large share of the income earners top 1% of the income distribution (Burdín et al., 2022). For these, hours worked might indeed be a relevant margin of response.

[44] Online Appendix presents additional visual evidence about the raw trends in the hours variable, splitting the sample into wage earners and self-employed individuals.

3.5.3 Extensive Margin Response

Our main outcome variable in the analysis of extensive margin responses is a binary variable $\mathbb{1}(y_{it} > 0)$ that takes the value of 1 if a TIE reports to the PLIT base in year t and 0 otherwise (i.e., either if he reported income to the CIT or PITC bases, or did not report income at all). Since $\mathbb{1}(y_{it} > 0)$ is observed for all individuals, our estimates are based on a balanced sample. To recover the extensive margin semi-elasticity, we estimate the following model:

$$\mathbb{1}(y_{it} > 0) = \alpha + \beta treat_i^{ATR} + \mu \log(1 - \tau_{it}^{ATR}) + X_i\delta \ + \ \lambda_t + v_{it} \tag{3.5}$$

where the treatment variable now refers to the *average* tax rate (τ_{it}^{ATR}) instead of the *marginal* tax rate. The coefficient of interest, μ, represents the extensive margin semi-elasticity and measures the change in the probability of reporting to the PLIT base when the *average* net-of-tax rate changes by 1%. For identification, we exploit the policy reform and instrument $\log(1 - \tau_{it}^{ATR})$ with the interaction variable $treat_i \times \mathbb{1}(year \geq 2012)$. All regressions are estimated using revenue weights, so μ corresponds exactly to the parameter $\bar{\mu}_L$ in equation 3.2. Standard errors are clustered at the individual level.

In the post-treatment period, some individuals do not report income to the PLIT base at all. Consequently, their labor income and *average* PLIT tax rate are unobserved. Hence, to analyze their extensive and income-shifting responses, we need to impute their hypothetical PLIT and tax rates. To do this, we extrapolate the income growth rate of individuals in G1 - who did not face changes in their *marginal* or *average* tax rates - and apply it to 2011's income of individuals who did not report PLIT. To make the tax rates comparable within groups, we also replicate this procedure on individuals who reported positive income.[45]

Graphical evidence. Figure 3.8 provides graphical evidence of the reduced-form effect of the 2012 Tax Reform on the extensive margin.[46] First, it is important to note that as explained in Section 3.4.1, by construction, all individuals in our sample reported to the PLIT base pre-treatment. Second, Figure 3.8 illustrates that the share of TIEs reporting to the PLIT income declines gradually over time, but this decline is larger for individuals in the treatment group. Three years after the reform, the percentage of taxpayers who report to the PLIT base is about 6% lower for TIEs who faced a tax-rate increase relative to those who did not.[47]

Figure 3.9 breaks down the analysis in wage earners and self-employed workers.[48] The response is larger for self-employed TIEs and for this group is observed as early as in 2012. For wage earners, we also observe a negative response. However, this response is smaller and takes at least two years to build up. Some specific features of Uruguay's tax code that apply to wage earners could accommodate this result. In particular, the fact that once a wage earner has earned a dollar in wages, it is tied to the PLIT base in the extensive margin, at least for that fiscal year. This implies that while responses on the intensive margin can happen immediately, responses in the extensive margin can take more time to realize fully.

Regression results. Table 3.4 reports our baseline estimates for extensive margin responses. Consistent with the graphical evidence, it shows that TIEs' PLIT

---

[45] Online Appendix describes this procedure in detail.

[46] See section 3.5.1 for additional details on how figures are constructed

[47] To rule out the already mentioned concerns about mechanical parallel pre-trends we replicate the analysis based on the *TAX-SSA* sample. See Online Appendix for additional details

[48] Online Appendix provides further details, including the raw trends for wage earners and self-employed workers separately

reporting decisions are highly sensitive to changes in the average tax rate. For the whole sample, the reduced-form estimates show that the reform led to a decrease of 6.6% in the share of TIEs reporting to the PLIT base, which implies a large extensive margin semi-elasticity of about 2.5. When splitting the sample by type of worker, self-employed individuals show a semi-elasticity of 4.4 compared to a semi-elasticity of about 2 for wage earners. Table 3.4 shows that these results are generally robust to the alternative specifications and follow the same patterns discussed for intensive margin elasticities. Columns 4, 8, and 12 report the results of an additional triple difference specification. As discussed in Section 3.4.2, one concern from our sample selection procedure is that it might create parallel pre-trends by construction. In the previous section, we reported some graphical evidence that mitigates these concerns based on an extension of the pre-treatment period for a restricted sample. Here, we report additional results that compare high- vs. low-intensity of treatment (i.e., G4 vs. G2). If our estimates are indeed capturing responses to changes in the tax rates, we should observe a larger response in the more intensively treated group. Indeed, the semi-elasticities obtained using the triple-difference specification are similar in direction and moderately larger than our baseline estimates. Finally, as for the intensive margin analysis, we also conduct a series of additional robustness tests that are reported in Online Appendix.

While our setting and data do not allow us to disentangle between all the possible mechanisms that could explain TIEs leaving the PLIT base completely, we can explore the importance of income-shifting behavior, discussed in the following section, relative to the other alternative explanations. To do this, we exploit the fact that, contrary to income-shifting responses, some types of extensive margin responses entail disappearing completely from the tax records. Figure 3.10 reports estimates for an alternative outcome variable that indicates if the TIEs reported either to the PLIT or to the CIT base. For comparison purposes, we also report the evolution of our baseline extensive margin outcome variable. Estimates for the full sample suggest that at least part of the effects observed on the extensive margin are driven by TIEs who disappear completely from the tax records, regardless of the tax base considered. When splitting the sample by wage earners and self-employed, the figure reveals that transitions from PLIT to CIT are purely explained by self-employed workers. On the contrary, wage earners cannot shift their income to the CIT base and therefore respond by purely exiting the PLIT.

To understand these results better, discussing how "pure" exits from the PLIT base should be interpreted is important. First, international migration, transition to informality, or simply moving to retirement are consistent with disappearing completely from the post-treatment tax records. However, the Uruguayan labor market structure, being highly-paid employees, and an average age of 50 years old make these very unlikely. One alternative explanation is that TIEs may switch to other forms of compensation, such as dividends or stock options (see, e.g., Goolsbee, 2000; Saez, 2017). While these are reported to the PTIC base by the employer, as explained before, legal restrictions prevent the Tax Agency from linking some of these income sources to individual taxpayers. Nevertheless, we conducted some exploratory analysis based on aggregate data reported by the Tax Agency. Figure 3.11 depicts the aggregate dividends distributed by firms as a share of GDP. The overall amount of dividends increased during the whole period, but starting in 2012, this trend is driven exclusively by non-nominative dividends. This preliminary evidence suggests that the PLIT to PITC mechanism cannot be ruled out as one of the mechanisms to explain the observed extensive margin response.[49]

[49] As a complementary analysis, we augmented the alternative extensive margin variable to

3.5.4 Income Shifting Responses

In this section, we dig into the analysis of income-shifting responses to the 2012 tax reform. It is important to recall that based on Uruguay's tax legislation, the relevant type of response regarding income shifting are complete transitions from one tax base to another, not partial income shifts. Hence, the analysis of income-shifting responses replicates the strategy used to estimate extensive margin responses but on an outcome variable that indicates whether a TIE reports income to the CIT base. The only change is that we augment equation 3.5 to allow for anticipatory responses to changes in the *average* tax rate. The reason is that the graphical evidence reported below reveals a clear anticipatory effect that took place the year before the reform. If anticipatory effects were important but not accounted for in the empirical model, the estimated elasticity would be downward biased, and the welfare costs of the reform would be overvalued. More specifically, we estimate the following equation:

$$\mathbb{1}(y_{it}^{CIT} > 0) = \alpha + \beta treat_i^{ATR} + \eta_1 \log(1 - \tau_{it}^{ATR}) + \eta_2 \log(1 - \tau_{it+1}^{ATR}) + X_i\delta \ + \ \lambda_t + v_{it} \tag{3.6}$$

where $\mathbb{1}(y_{it}^{CIT} > 0)$ is an indicator function for reporting positive earnings on the CIT returns, and τ_{it}^{ATR} and τ_{it+1}^{ATR} correspond to individuals' PLIT average tax rates in years t and $t + 1$. All the remaining variables are defined as in equation (3.5). In this model, there are two coefficients of interest. η_1 captures the *contemporaneous* income-shifting responses to changes in the current log *average*

include reporting to nominative PTIC items that can be linked at the individual level, and the results remain unchanged. This finding suggests that, unlike the non-nominative items, PTIC items that can be linked at the individual level are not an important driver of the extensive margin response.

net-of-tax rate, while η_2 captures the anticipatory component of the response.[50] We refer to the longer-term income-shifting semi-elasticity as η, which is given by the sum of η_1 and η_2 and represents the relevant parameter to be used to assess the efficiency costs of the reform. As in the intensive and extensive margin analyses, we exploit the exogenous variation in the average tax rate as a consequence of the 2012 tax reform, and we instrument the endogenous variables τ_{it}^{ATR} and τ_{it+1}^{ATR} with the DinD interaction terms $treat_i \times \mathbb{1}(year \geq 2012)$ and $treat_i \times \mathbb{1}(year \geq 2011)$, respectively. Regressions are revenue-weighted so that the longer-term income-shifting semi-elasticity η derived from the specification 3.6 corresponds to the parameter $\bar{\eta}_{L,B}$ in equation 3.2.

Graphical evidence. Figure 3.12 illustrates the changes in the percentage of taxpayers reporting income to the CIT base for treatment and control groups. The graphical analysis yields three main findings. First, very few TIEs in our sample (2% and 4%, respectively) report income both to the PLIT and CIT bases in the pre-treatment period. Second, there is compelling evidence about parallel pre-trends, at least until 2010. Third, the tax reform seems to have increased the share of TIEs that report to the CIT base. Figure 3.13 splits the sample into wage earners and self-employed individuals and shows that individuals who report to the PLIT and CIT bases in the pre-treatment period are almost exclusively self-employed TIEs, who are also driving the overall response.[51] Figure 3.12 also reveals some additional features of the dynamics of the effects. On the one hand, the effects of the tax reform build gradually over time, stabilizing about three years

[50] Future net-of-tax rate has also been used to capture anticipatory effects in the public economics literature. See Goolsbee (2000); Holmlund and Söderström (2011); Auten and Kawano (2014), among others.

[51] Online Appendix provides more detailed evidence about the time patterns by treatment and control groups.

after the reform. On the other hand, there seems to be a relevant anticipatory effect that manifests as early as 2011.[52] This finding is not necessarily surprising. Indeed, Foremny et al. (2018) already documented an anticipatory effect of a tax reform carried out in Uruguay in 2007 for a subgroup of workers that is included in our definition of self-employed - i.e., the liberal professionals: lawyers, public notaries, architects, engineers, and accountants, among others -. The authors find a large shifting response of business income the month before the reform occurred.

Regression Results. Table 3.5 reports the regression estimates of the income-shifting semi-elasticity based on Equation 3.6. Each panel reports the reduced-form, first stage, and 2SLS estimates. Panel A. measures the contemporaneous effect, panel B. reports the anticipation effects, and Panel C. reports the aggregated response, corresponding to the long-run semi-elasticity. In addition, panel C. also reports the corresponding p-value test for the null hypothesis that the long-run elasticity is zero. Three main findings are worth noting. First, consistent with the graphical reduced-form evidence, the behavioral responses in the income-shifting margin to the 2012 tax reform are large and statistically significant. Panel C., column (1) shows that once we account for anticipation effects, the long-run semi-elasticity is -1.967 for the full sample and statistically significant (p-value < 0.01). This result can be interpreted as follows: a 1% increase in the *average* net-of-tax rate on income reported to the PLIT base reduces by 2p.p. the probability of reporting income on the CIT base. Second, considering the full sample of TIEs, the anticipation effect accounts for more than 35% of the long-run semi-elasticity (-0.714 out of -1.967). Third, the income-shifting response is mostly explained by adjustments of self-employed individuals. For this specific group of TIEs, we

[52] Ideally, to assess the anticipation effects, one would like to look at the entry rates to the CIT base using monthly or daily data. Unfortunately, we currently do not have access to this type of granular data.

estimate a three times larger semi-elasticity than the full sample. It is important to note that although the size of the effect might seem implausibly large, the share of TIEs that report income to the CIT base in the pre-reform period is relatively low (approximately 15%).[53] Similar to what we observe for the intensive and extensive margin, the estimates obtained using the predicted labor income instead of the actual income are considerably higher. However, the instrument in the 2SLS estimation performs poorly and the associated F-statistics of the first-stage regressions are very low, which might lead to a weak instrument problem.[54]

3.6 Welfare Analysis

In Section 3.3 we showed that the efficiency costs of a small increase in $d\tau_L$ are a combination of the revenue losses associated with the behavioral responses on the intensive, extensive and income shifting margins. In this section we use the elasticities estimated in Section 3.5 to quantitatively assess them. As is typical in the literature, to make the interpretation easier, we express the efficiency costs described in Equation 3.2 in terms of the projected mechanical increase in tax revenues $\frac{dM}{d\tau_L} = Y_L \times d\tau_L$. Hence:

$$\frac{dW}{dM} = \frac{dB}{dM} = -\frac{\tau_L}{1-\tau_L} \times [\bar{\varepsilon}_L + \bar{\mu}_L + \bar{\eta}_{L,B}] \tag{3.7}$$

Intuitively, this ratio shows the welfare loss associated with the tax reform expressed as a fraction of the projected mechanical increase in tax revenue. Since the 2012 tax reform affected two of the marginal tax rates in the pre-reform pro-

---

[53]See Online Appendix

[54]Further robustness tests are reported in Online Appendix.

gressive schedule, we calculate τ_L as the weighted average of the two pre-reform tax rates (20% and 25%) affected by the reform, resulting in a value of 22.34%. From our preferred specifications in Section 5, $\bar{\varepsilon} = 0.577$ (row C, col. 1, Table 3.2), $\bar{\mu}_L = 2.479$ (row C, col. 1, Table 3.4) and $\bar{\eta}_{L,B} = -1.967$ (row C, col. 1, Table 3.5). By plugging these values into equation (3.7), we estimate that the efficiency costs of the 2012 tax reform are about 31.3% of the projected mechanical increase in tax revenues.[55]

This result has several implications. First, the efficiency costs of taxation are smaller than the projected increase in tax revenues. This indicates that the new top tax rates are on the "correct" side of the Laffer curve and that the reform had a positive net effect on total tax revenues. Second, the magnitude of the efficiency costs associated with this reform is comparable to estimates from recent reforms that increased tax rates on high-income earners in developed (e.g., Saez 2017) and developing countries (e.g., Jouste et al. 2021).

Finally, a step-by-step calculation of the efficiency costs of taxation provides a very clear indication of the importance of considering all relevant margins of responses when analyzing the efficiency costs of policy reforms. If we consider only the intensive margin responses, the efficiency costs associated with the 2012 tax reform would be 15%. This would ignore the fact that some people completely

[55]One concern is that we could be underestimating the welfare loss of the tax reform by not considering changes in VAT revenues associated with changes in the income reported by self-employed workers, who are also subject to this tax. In this regard, there are two things that are worth to be mentioned. First, self-employed workers are subject to VAT both in the PLIT and in the CIT base. Figure 3.10, and Tables 3.4 and 3.5 provide evidence that most self-employed workers actually transition from PLIT to CIT, and do not exit all tax bases. Therefore, we should not expect a change in VAT revenues associated with these mechanisms. Second, while it is true that intensive margin responses by self-employed workers will affect VAT collection, we believe that including this change would barely modify our estimates. First, because intensive margin responses are considerably smaller, and second, because self-employed workers only represent a 25% of our sample of interest.

abandoned the PLIT base and would underestimate the welfare costs of the tax increase. When introducing the extensive margin responses, the efficiency costs go up to 103%. In this case, the behavioral responses due to the increase in the marginal tax rate would more than offset the increase in tax revenues, and the reform would have a negative effect on total tax revenues. Finally, when we introduce the income-shifting responses, we incorporate in our calculations the fact that some individuals left the PLIT base but switched to the CIT, increasing the taxes collected by this tax base and partly offsetting the efficiency costs associated with the intensive and extensive margin responses. An alternative way of interpreting these results is that the efficiency costs of the tax reform are mostly explained by extensive margin responses - 81.1% versus 18.8% explained by intensive margin responses -. However, fiscal externalities such as income-shifting offset about 64.4% of these efficiency costs. This has strong policy implications since it suggests that tax administration efforts should be concentrated on limiting the possibilities of extensive margin responses by reducing incentives to shift between tax bases (e.g., minimize tax loopholes, and eliminate opportunities for arbitrage).

In sum, this simple but illustrative analysis shows that a full assessment of the effects of a tax reform requires a careful analysis of all the possible margins of response available. If this is not done properly, the conclusions obtained can be misleading and may even suggest that a tax increase had a negative effect on total revenues, when it actually had the opposite effect.

3.7 Conclusions

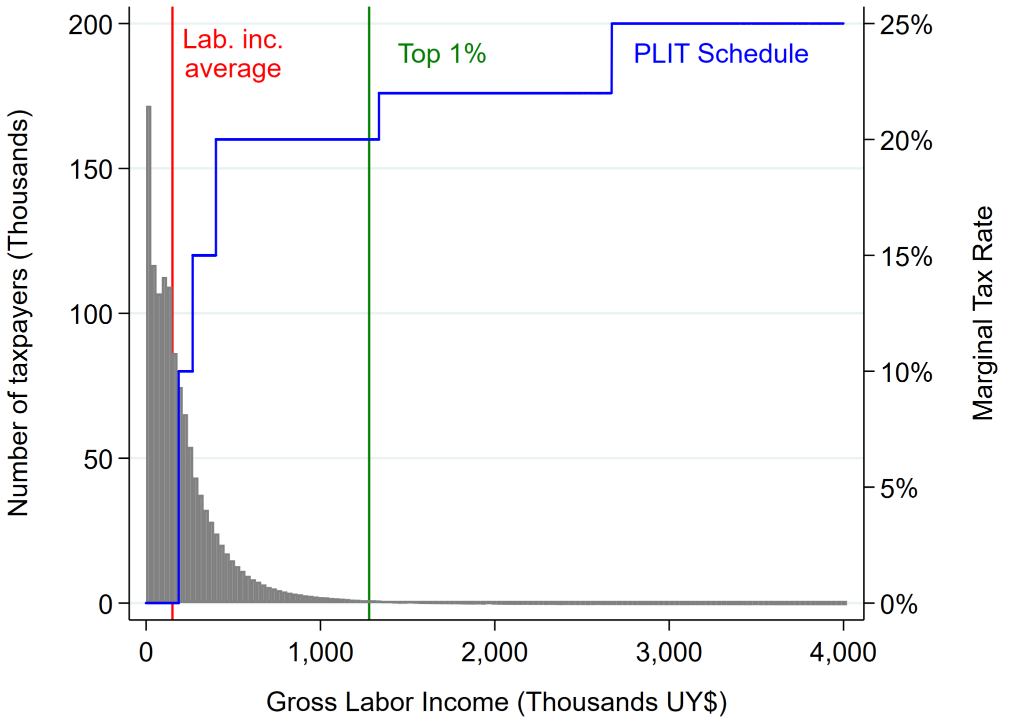
Using a unique policy experiment induced by a tax reform in Uruguay, this paper analyzed the behavioral responses of TIEs along different margins. To guide the empirical analysis and estimate the welfare cost of the reform, we developed a simple theoretical model where individuals can choose between alternative tax bases to report their earned income. According to this model, individuals can respond to tax changes by reducing reported income and hours worked (intensive margin), by exiting the labor income tax base (extensive margin), and by reallocating income from the labor income tax base to the corporate income tax base (income-shifting margin). To identify these three elasticities, we exploited exogenous variation in tax rates within the top 1% of the labor income distribution in Uruguay using a DinD design.

Our estimates suggest a moderate intensive margin response with an implied elasticity of 0.577 in our preferred specification. In the case of wage earners, we find evidence that responses on labor income are partially explained by real labor supply adjustment through fewer hours worked. We also document responses in the extensive margin (semi-elasticity ≈ 2.5), which are mostly driven by self-employed individuals. Finally, we find evidence supporting the hypothesis that individuals reallocate income from labor to the corporate income tax base. The elasticity associated with this income-shifting channel is roughly -2. This effect is driven by both contemporary and anticipatory behavioral responses of self-employed individuals. Our welfare analysis suggests that the losses associated with the 2012 tax reform account for 31% of the projected increase in revenue, implying that the new tax rate schedule is on the left-hand side of the Laffer curve. Hence, the specific tax reform studied in this paper was not an inefficient strategy to raise tax revenues.

At this point, it is worth mentioning some of the limitations of our study. First, while unilateral misreporting does not seem to explain the intensive-margin responses documented in the paper, we cannot rule out the potential role of more sophisticated forms of tax evasion involving collusive employee-employer agreements (Bjørneby et al., 2021). Second, data restrictions prevented us from disentangling further the channels behind the extensive margin response of wage earners. For instance, we were unable to study theoretically plausible yet empirically unlikely channels such as international migration of top earners (Kleven et al., 2020). More importantly, legal restrictions in how the data is reported to the tax agency prevented us from analyzing responses that involve shifting from the labor income tax base to non-nominative capital income. Third, our research design is only suitable for identifying behavioral responses to taxation in the short and medium run. Hence, our study is silent about mechanisms that may arise in the long run, such as career effects of work effort, entrepreneurial activities and innovation (Best and Kleven, 2012; Akcigit et al., 2022). Finally, we did not investigate potential responses along the bargaining margin, which might be particularly relevant in the case of individuals employed in top managerial positions. For instance, it has been argued that tax changes may affect the pay-setting bargaining power of these individuals and alter the private returns from rent-seeking activities within firms (Piketty et al., 2014; Rothschild and Scheuer, 2016). The caveats mentioned above suggest potential extensions of this study. Future research could also analyze responses of TIEs to income and wealth taxes using an integrated framework. Moreover, it would be interesting to analyze whether certain individual traits and social preferences moderate the responses of TIEs to taxation.

Our study has potentially relevant policy implications for tax design in low- and middle-income countries where taxing TIEs and wealth holders is relatively difficult. While tax enforcement capacities are usually weaker in these countries compared to developed countries, TIEs seem to be as sophisticated as their counterparts in advanced economies in terms of their access to tax-planning instruments (Londoño-Vélez and Mahecha, 2021). Developing countries face the challenge of improving their fiscal capacity and enhancing tax progressivity without introducing major distortions in economic incentives. Moreover, the presence of a wide range of avoidance opportunities may also put some limits to the redistributive capacity of tax systems in developing country contexts. Our estimates suggest that 81.1% of the efficiency costs of the tax reform are associated with extensive margin responses, and 18.2% are associated with intensive margin responses. However, fiscal externalities such as income-shifting offset about 64.4% of these costs. Given the large size of this type of response, tax agencies should devote more effort to enhancing tax administration practices, reducing tax loopholes, and restricting arbitrage opportunities between different tax bases. Making further progress in understanding the behavior of TIEs will also require continuous improvements in tax data transparency and cooperation between researchers and tax authorities.

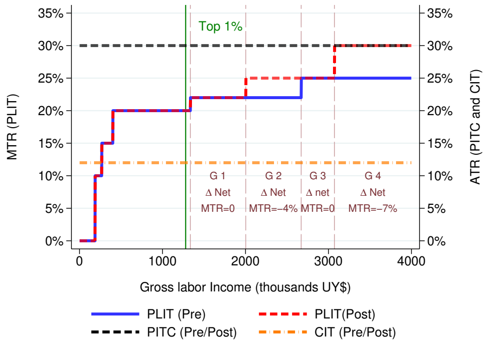
Figure 3.1: Labor Income Distribution and the 2011 PLIT schedule (2011)



Notes: This figure shows the labor income distribution in the *TAX* sample (gray bars) and the PLIT schedule for the year 2011. The solid red line indicates the average labor income. The solid green line indicates the lower limit of the top 1% of formal labor income earners (aged 25-70 years old). The solid blue line represents the marginal tax rates in the PLIT schedule, also for year 2011 (pre-reform period). Gross labor income is expressed in thousands of constant 2011 UYU.

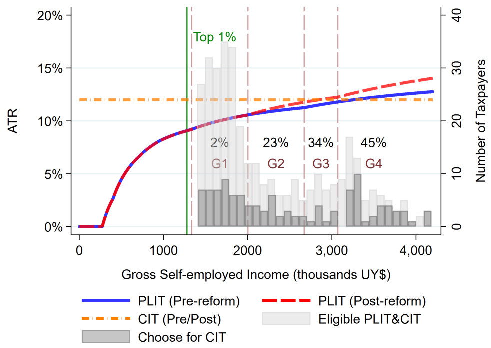
Figure 3.2: Tax Variation Created by the 2012 Tax Reform

a. Marginal Tax Rates Before and After the Tax Reform



Simulated Average Tax Rates, Before and After the Reform

b. Self-Employed



c. Wage Earners

Notes: This figure shows the tax variation created by the 2012 tax reform. All panels depict red dashed vertical lines that correspond to the treatment groups defined by the tax reform based on the pre-reform gross labor income (see Section 3.2.3 for additional details). Panel a. shows the change in the PLIT schedule introduced by the reform. The solid blue line indicates the pre-reform PLIT marginal tax rates, while the red dashed line indicates the post-reform marginal tax rates. The change in log marginal tax rate is reported for each group separately. For comparison purposes we also report the effective corporate tax rate (dashed orange line) and the effective dividends tax rate as an example of one of the capital income components taxed by PITC (orange dot-dashed line). See more details in Table 3.1. Panel b. shows the simulated average tax rate for PLIT before (blue solid line) and after the reform (red dashed line), as well as the average CIT rate (orange dot-dashed line) for a self-employed professional worker. For PLIT, tax rates were computed over $\frac{1}{0.7}$ of the gross labor income because self-employed workers have an automatic 30% deduction on their reported income. Mechanical deductions (e.g., payroll taxes) are already considered in the simulation, as well as an itemized deduction for a single child. The histograms report the number of self-employed workers eligible to choose the CIT base - light gray bars - and the number of workers who actually choose it - dark gray bars-. A note reports the percentage of self-employed workers that choose the CIT base within each treatment/control group. The solid green line indicates the lower limit of the top 1% of formal labor income earners (aged 25-70 years old). Panel c. also depicts the average PLIT tax rates before and after the reform but for the case of wage earners. For this type of workers, the tax rates are computed directly based on the tax pre- and post-reform tax schedule, and assuming the same set of deductions as in Panel b.. In all cases, gross labor income is expressed in thousands of constant 2011 UYU.

Figure 3.3: Timeline of the 2012 tax reform

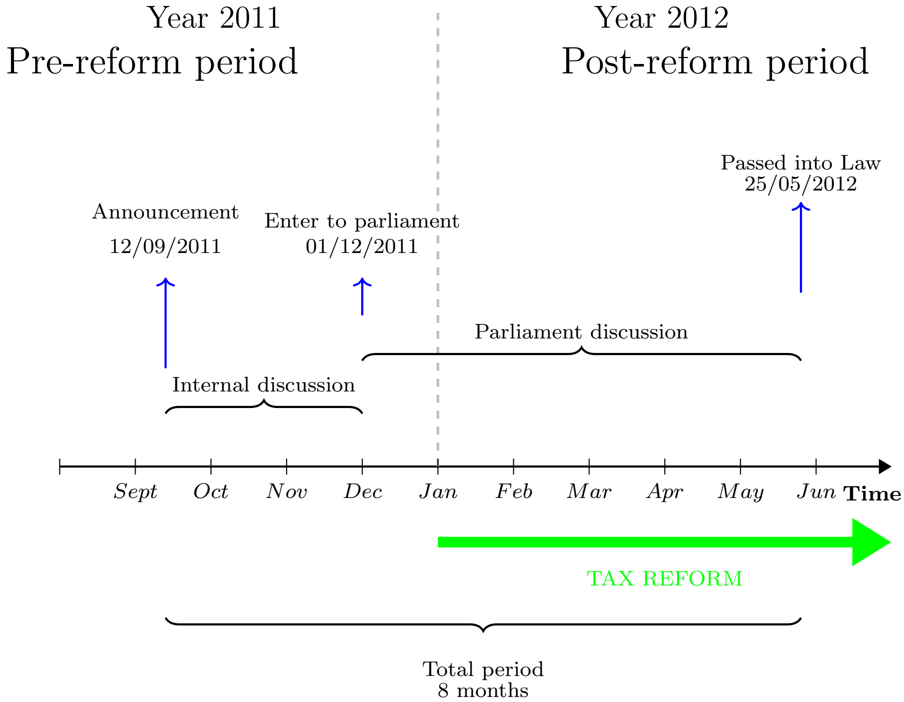
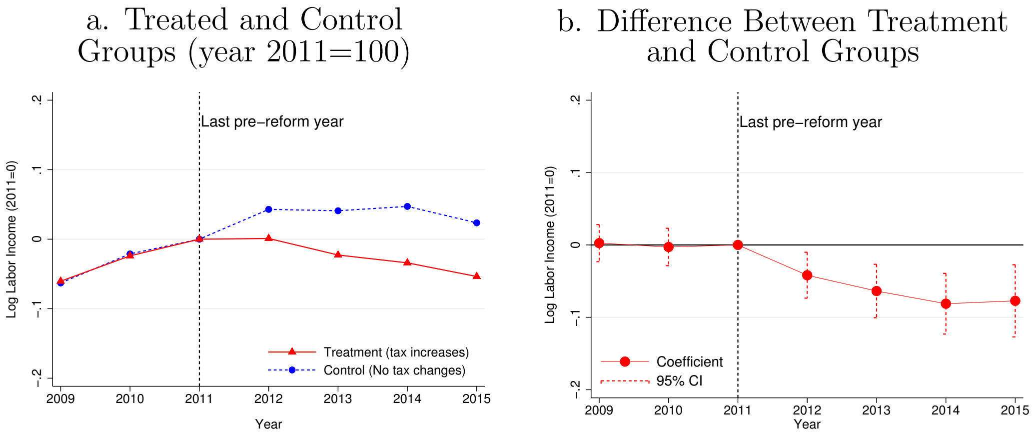


Figure 3.4: Labor Income Response to the 2012 Tax Reform: Graphical Evidence

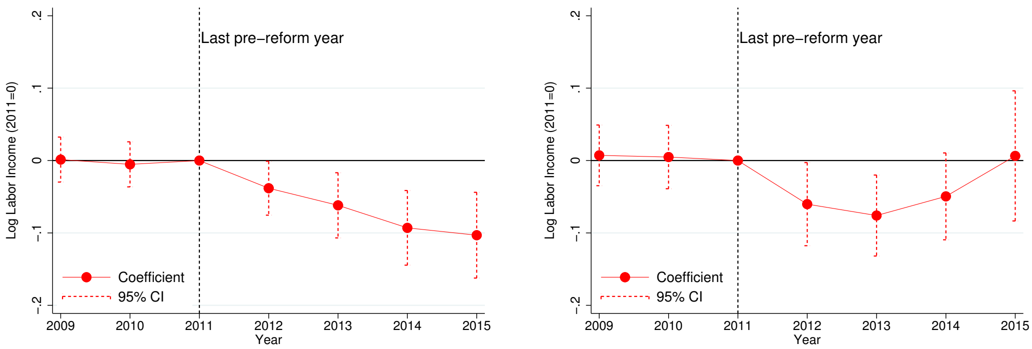


Notes: This figure shows the effects of the 2012 tax reform on gross labor income reported to the PLIT base. The figure compares the evolution of log labor income for TIEs that were and were not affected by the increase in the marginal tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an unbalanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in (0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Panel a. depicts the evolution of the mean log gross labor income over time for treated and control groups, normalized to zero in the pre-reform year 2011. Panel b. depicts the raw differences between these two time series. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust to heteroskedasticity standard errors.

Figure 3.5: Labor Income Response to the 2012 Tax Reform: Difference Between Treatment and Control Groups, by Group of Workers

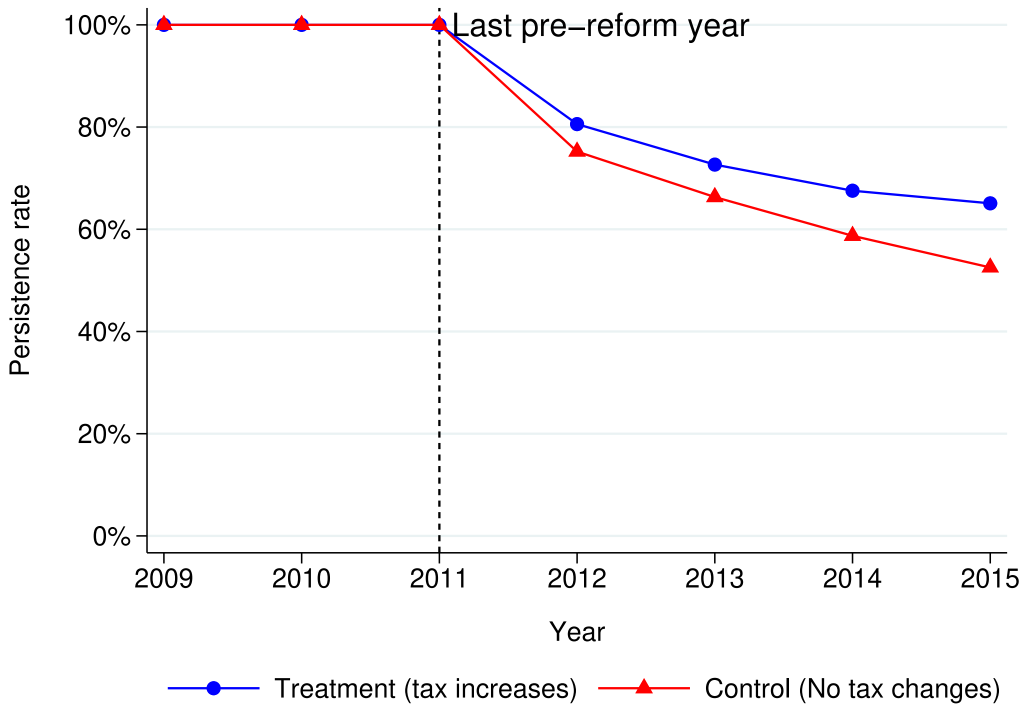
a. Wage Eaerners

b. Self-Employed



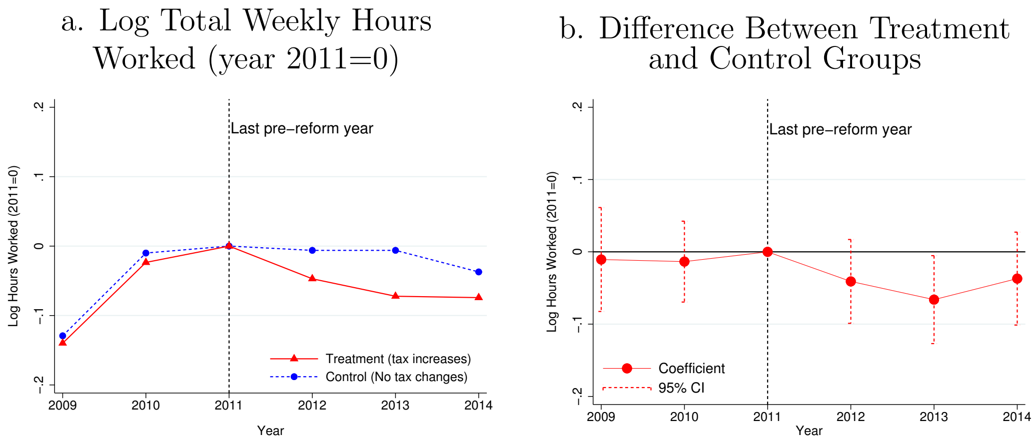
Notes: This figure shows the effects of the 2012 tax reform on gross labor income reported to the PLIT base for wage earners and self-employed workers separately. The figure compares the evolution of log labor income for TIEs that were and were not affected by the increase in the marginal tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an unbalanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in (0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Each panel depicts the raw differences between treatment and control groups normalized to 0 in the pre-reform year. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust to heteroskedasticity standard errors. Online Appendix plots the raw trends for treatment and control groups for wage earners and self-employed workers separately.

Figure 3.6: Tax Bracket Persistence Rate for Treatment and Control Group



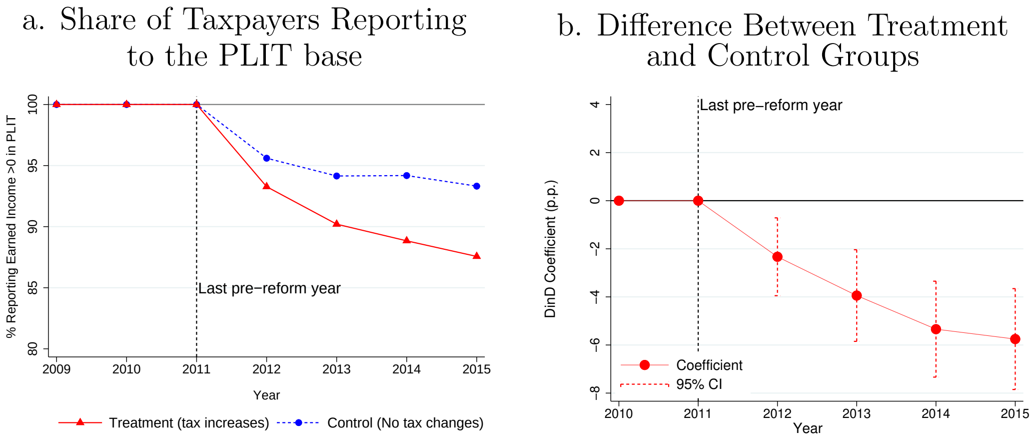
Notes: This figure shows the persistence of taxpayers in the control and treated group throughout the period of analysis. The assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The persistence rate is calculated as the ratio of taxpayers who stay in the same group from one year to the next over the total of taxpayers in each group.

Figure 3.7: Labor Supply Response to the 2012 Tax Reform: Graphical Evidence Based on TAX-SSA Sample



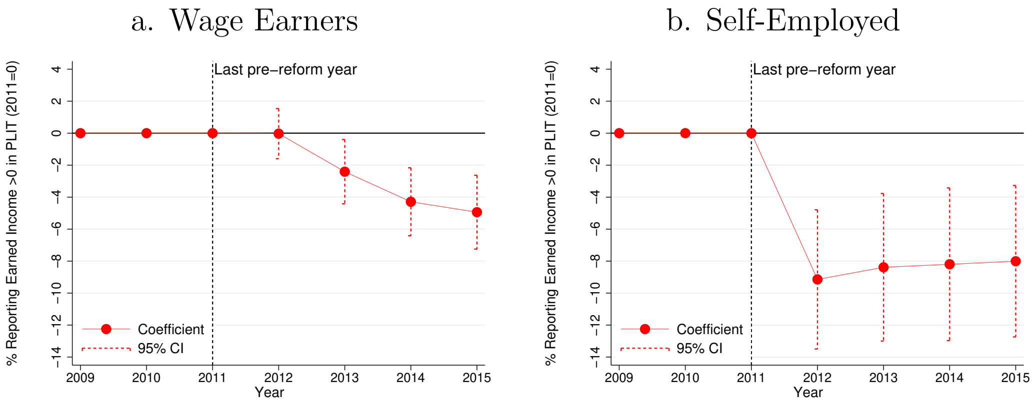
Notes: This figure shows the effects of the 2012 tax reform on log weekly hours worked reported to the PLIT base. The figure compares the evolution of log weekly hours worked for TIEs that were and were not affected by the increase in the marginal tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an unbalanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in (0, y^{max}]$ in the post-reform years. However, since the log weekly hours worked variable is based on information reported in the SSA records, the sample is restricted to those individuals that can be matched to the SSA data (i.e., the *TAX-SSA* sample described in detail in 3.4.1) and information is reported for the 2007-2014 period. Panel a. depicts the evolution of the mean log weekly hours worked over time for treated and control groups, normalized to zero in the pre-reform year 2011. Panel b. depicts the raw differences between these two time series. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust to heteroskedasticity standard errors.

Figure 3.8: Extensive Margin Response to the 2012 Tax Reform: Graphical Evidence



Notes: This figure shows the effects of the 2012 tax reform on the percentage of TIEs reporting positive earned income to the PLIT base. The outcome variable is a dummy variable that indicates whether an individual reports positive income in the PLIT base in a given year. All the remaining cases are coded as 0 and may correspond to individuals who report income in the CIT or PITC bases, or simply not report income at all. The figure compares the evolution of the share of TIEs reporting positive income to the PLIT tax base for TIEs that were and were not affected by the increase in the average tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an balanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in [0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Panel a. depicts the evolution of the share of TIEs reporting positive income to the PLIT tax base for treated and control groups. Since the sample of analysis requires that an individual received positive income in the 2009-2011 period, all pre-treatment shares are 100% by construction, either for the treatment and control groups (See Section 3.4.1 for a more detailed explanation). Panel b. depicts the raw differences between these two time series. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust standard errors clustered at the individual level.

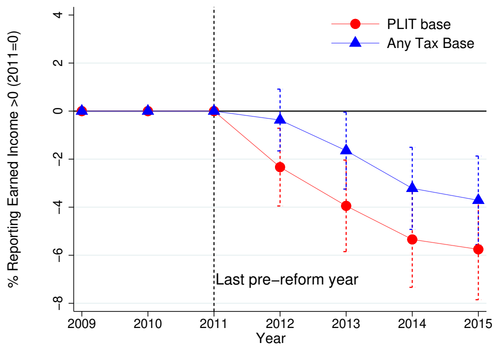
Figure 3.9: Extensive Margin Response to the 2012 Tax Reform: Difference Between Treatment and Control Groups, by Group of Workers



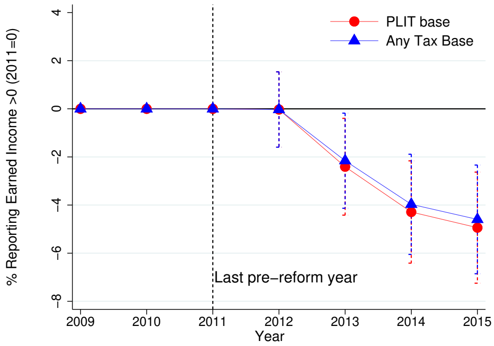
Notes: This figure shows the effects of the 2012 tax reform on the percentage of TIEs reporting positive earned income to the PLIT base for wage earners and self-employed workers separately. The outcome of interest is a dummy variable that indicates whether an individual reports positive income in the PLIT base in a given year. All the remaining cases are coded as 0 and may correspond to individuals who report income in the CIT or PITC bases, or simply not report income at all. The figure compares the evolution of TIEs reporting positive earned income to the PLIT base for TIEs that were and were not affected by the increase in the average tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is a balanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in [0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Since the sample of analysis requires that an individual received positive income in the 2009-2011 period, all pre-treatment shares are 100% by construction, either for the treatment and control groups (See Section 3.4.1 for a more detailed explanation). Each panel depicts the raw differences between treatment and control groups. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust standard errors clustered at the individual level. Online Appendix plots the raw trends for treatment and control groups for wage earners and self-employed workers separately.

Figure 3.10: Extensive Margin Response to the 2012 Tax Reform - Reporting Earnings in PLIT *vs* in Any Tax Base. Difference Between Treatment and Control Groups

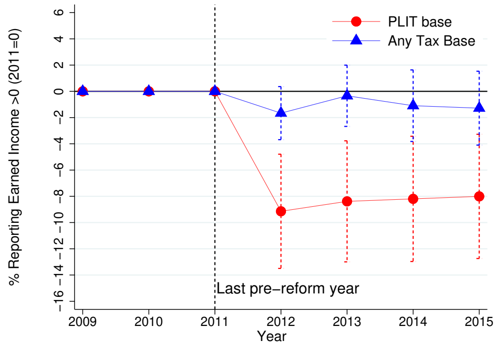
a. All Workers



b. Wage Earners

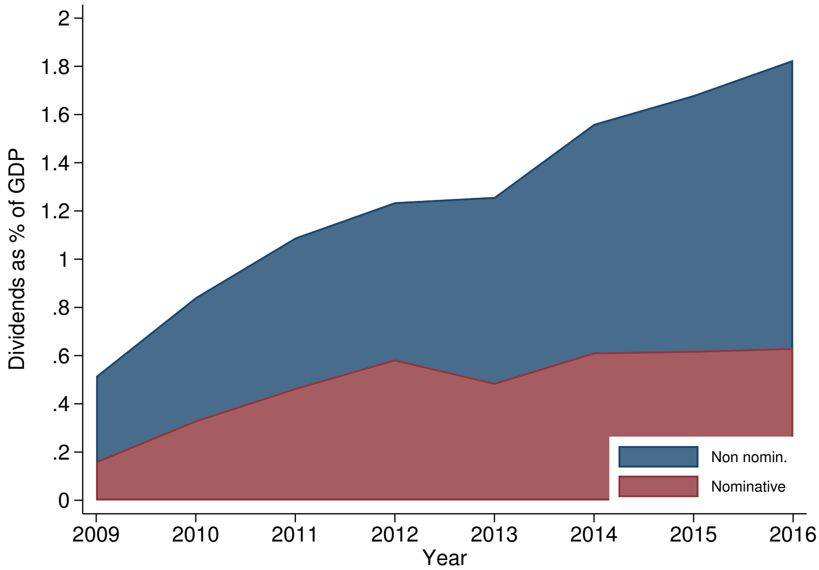


c. Self-Employed



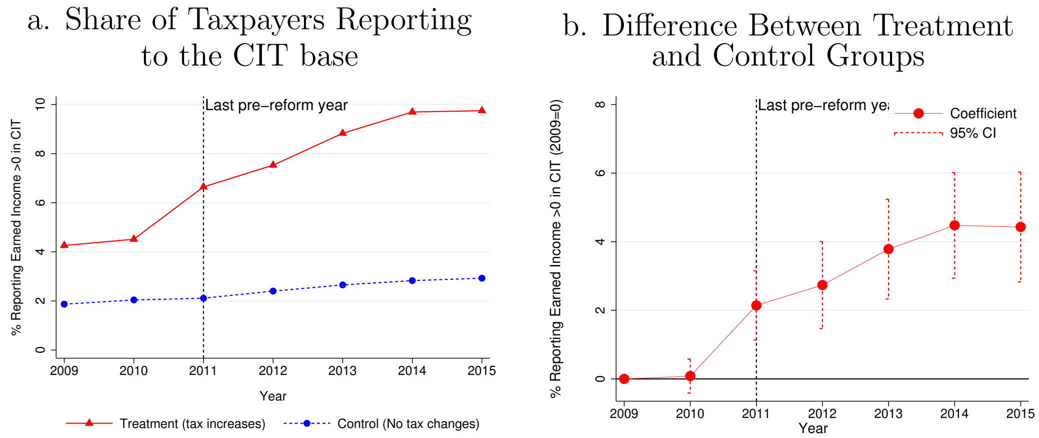
Notes: This figure shows the effects of the 2012 tax reform on the percentage of TIEs reporting positive earned income to the PLIT (red circles) and on the percentage of TIEs reporting positive earned income to any tax base (blue triangles). We report these results for all workers (panel a.), wage earners (panel b.), and self-employed workers (panel c.). For estimates of the effects on reporting to PLIT, the outcome of interest is a dummy variable that indicates whether an individual reports positive income in the PLIT base in a given year. All the remaining cases are coded as 0 and may correspond to individuals who report income in the CIT or PITC bases, or simply not report income at all. For estimates of the effect on reporting to any tax base, the outcome of interest is a dummy for reporting either to the PLIT or the CIT bases. The figure compares the evolution of these two outcomes for TIEs that were and were not affected by the increase in the average tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an balanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in [0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Since the sample of analysis requires that an individual received positive income in the 2009-2011 period, all pre-treatment shares are 100% by construction, either for the treatment and control groups (See Section 3.4.1 for a more detailed explanation).Each panel depicts the raw differences between treatment and control groups. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust standard errors clustered at the individual level.

Figure 3.11: Evolution of Dividends Income Share (2009-2016)



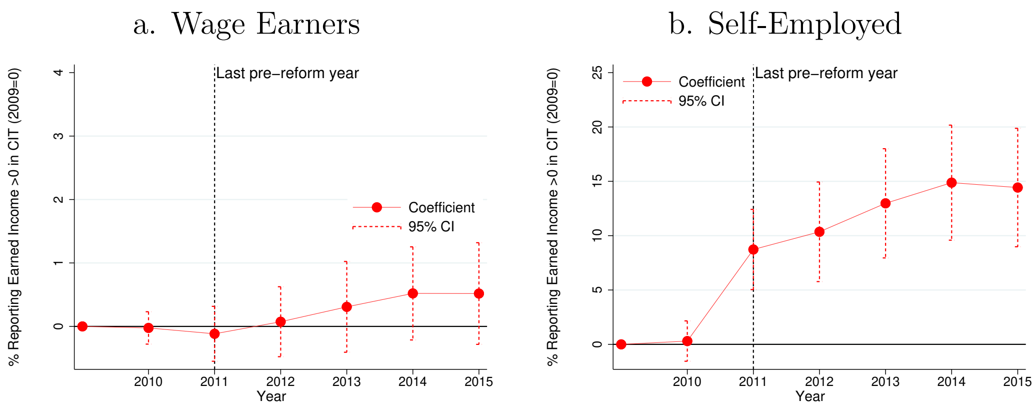
Notes: This figure shows the total revenue from nominative and non-nominative dividends as a share of GDP. Nominative dividends are taxed at the individual level and hence the dividend-receiver can be identified when PITC is paid, whilst in the case of non-nominative dividends individual PITC is withheld at the firm level. Aggregate dividends by category provided by the Tax Agency (DGI). For details, see Burdín et al. (2022).

Figure 3.12: Income-Shifting Margin Response to the 2012 Tax Reform: Graphical Evidence



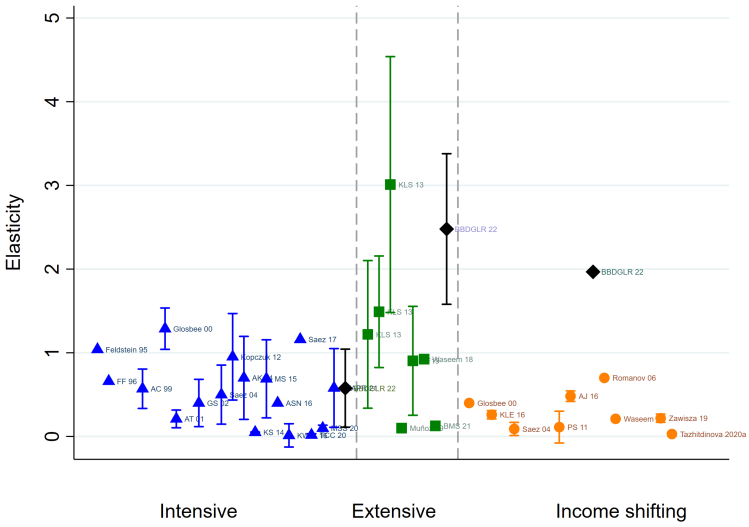
Notes: This figure shows the effects of the 2012 tax reform on the percentage of TIEs reporting positive income to the CIT base. The outcome variable is a dummy variable that indicates whether an individual reports positive income in the CIT base in a given year. All the remaining cases are coded as 0 and may correspond to individuals who report income in the PLIT or PITC bases, or simply not report income at all. The figure compares the evolution of the share of TIEs reporting positive income to the CIT tax base for TIEs that were and were not affected by the increase in the average tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an balanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in [0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Panel a. depicts the evolution of the share of TIEs reporting positive income to the CIT base for treated and control groups. Panel b. depicts the raw differences between these two time series. Because of the anticipation effect, in this case we normalize the difference to be 0 in year 2009. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust standard errors clustered at the individual level.

Figure 3.13: Income-Shifting Margin Response to the 2012 Tax Reform: Difference-in-Differences Estimates by Group of Workers



Notes: This figure shows the effects of the 2012 tax reform on the percentage of TIEs reporting positive earned income to the CIT base for wage earners and self-employed workers separately. The outcome of interest is a dummy variable that indicates whether an individual reports positive income in the CIT base in a given year. All the remaining cases are coded as 0 and may correspond to individuals who report income in the PLIT or PITC bases, or simply not report income at all. The figure compares the evolution of TIEs reporting positive earned income to the CIT base for TIEs that were and were not affected by the increase in the marginal tax rates. Assignment to treatment and control groups is based on pre-reform gross labor income (see Section 3.4.2). The sample is an balanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in [0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Each panel depicts the raw differences between treatment and control groups normalized to 0 in 2009. For a more transparent interpretation of the trends observed in the figures, these are based on unweighted estimates. 95% confidence intervals are reported based on robust standard errors clustered at the individual level. Online Appendix plots the raw trends for treatment and control groups for wage earners and self-employed workers separately.

Figure 3.14: Elasticities



Notes: This figure plots the elasticities (point estimate) and their confidence intervals –when available– for the review of the consulted literature. The labels are as follow: Feldstein 95 Feldstein (1995), FF 96 Feldstein and Feenberg (1996), AC 99 Auten and Carroll (1999), Glosbee 00 Goolsbee (2000), AT 01 Aarbu and Thoresen (2001), GS 02 Gruber and Saez (2002), Saez 04 Saez (2004), Kopczuk 12 Kopczuk (2012), AK 14 Auten and Kawano (2014), KS 14 Kleven and Schultz (2014), MS 15 Milligan and Smart (2015), ASN 16 Auten et al. (2016), KWW 16 Kawano et al. (2016), Saez 17 Saez (2017), TCC 20 Tortarolo et al. (2020), MSS 20 Miao et al. (2020), JKAPR 21 Jouste et al. (2021), KLS 13 Kleven et al. (2013), Muñoz 19 Muñoz (2019), AF 19 Agrawal and Foremny (2019), Waseem 18 Waseem (2018), BMS 21 Bastani et al. (2021), Glosbee 00 Goolsbee (2000), KLE 16 Kreiner et al. (2016), Saez 04 Saez (2004), PS 11 Pirttilä and Selin (2011), AJ 16 Alstadsæter and Jacob (2016), Romanov 06 Romanov (2006), Waseem 18 Waseem (2018), Zawisza 19 Zawisza (2019), Tazhitdinova 2020a Tazhitdinova (2020a). BBDGLR 22 correponds to our estimations.

Table 3.1: Income Taxation to Individuals in Uruguay, 2011 Tax Schedule

Panel A. Personal Labor Income Tax (PLIT)

Gross labor income brackets (Annual)	Rate
0-186	0%
186-267	10%
267-400	15%
400-1,336	20%
1,336-2,671	22%
Over 2,671	25%

Annual deductions	Rate
0-80	10%
80-213	15%
213-1148	20%
1148-2484	22%
Over 2484	25%

Panel B. Personal Income Tax on Capital (PITC)

Income source	Statutory tax rate	Effective tax rate
Interests from deposits	3%	3%
Dividends and other financial incomes	7%	30%
Real estate rent	12%	12%

Panel C. Corporate Income Tax (CIT)

Annual corporate income	Statutory tax rate	Effective tax rate
0-8,948	25%	12%
Over 8,948	25%	-

Notes: This table reports the statutory tax rates faced by taxpayers according to the personal income tax schedule in 2011, provided by the Tax Agency (DGI). Panel (a) shows the marginal tax rate schedule faced by individuals reporting income to the PLIT base, while panel (b) shows the statutory (and effective) tax rates under the PITC base. Finally, panel (c) shows the statutory (and effective) tax rates for individuals reporting earned income on their CIT returns. All the values are expressed in thousands of 2011 UYU.

Table 3.2: Intensive Margin Responses on Labor Income to the 2012 Tax Reform. Reduced Form, First Stage and Elasticities

	All Workers			Wage Earners			Self Employed		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Reduced Form									
$d\Delta \log y$	-0.026**	-0.020*	-0.025**	-0.027**	-0.019	-0.023*	-0.027	-0.027	-0.031
	(0.010)	(0.012)	(0.011)	(0.012)	(0.013)	(0.013)	(0.019)	(0.030)	(0.022)
B. First Stage									
d $\Delta\log(1-\tau^{MTR})$	-0.045***	-0.011***	-0.045***	-0.045***	-0.010***	-0.046***	-0.042***	-0.016***	-0.043***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)
C. 2SLS DinD									
$\varepsilon = \frac{d\Delta \log y}{d\Delta\log(1-\tau^{MTR})}$	0.577**	1.876	0.544**	0.605**	2.009	0.512*	0.652	1.678	0.734
	(0.238)	(1.300)	(0.258)	(0.276)	(1.633)	(0.294)	(0.480)	(2.095)	(0.533)
F-statistic	2,283	55	1,747	1,892	35	1,436	401	21	317
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predicted Income	No	Yes	No	No	Yes	No	No	Yes	No
Matching	No	No	Yes	No	No	Yes	No	No	Yes
Observations	23,068	16,808	18,376	17,754	13,608	14,544	5,314	3,200	3,832

Notes: This table reports estimates of the intensive margin responses to the tax reform. Each cell corresponds to one regression estimate. The sample used for the regression analysis is an unbalanced panel of TIEs observed in the PLIT records during the 2009-2015 period with earnings in the range of $y_{it} \in (0, y^{max}]$ in the post-reform years. All variables use information reported in the *TAX* records. Row A. shows the results of a reduced-form DinD estimation of specification based on Equation 3.3, where the dependent variable is the yearly log change of labor income. Row B. shows the first stage estimates associated to the 2SLS-DinD model in Equation 3.4. In this case, the dependent variable is the yearly log change of the marginal net-of-tax rate. Row C. shows the 2SLS-DinD estimated elasticity to the marginal net-of-tax rate based on Equation 3.4, where the dependent variable is the yearly log change of labor income and log change of the marginal net-of-tax rate is instrumented with the DinD interaction between a treatment dummy and a 2012 year dummy. All regressions are income-weighted. Hence, the elasticity reported in row C. corresponds to the parameter $\bar{\varepsilon}_L$ in equation (3.7). The "F-statistic" row refers to the F-statistic from the first-stage regression used to evaluate the the strength of the instrument. "Controls" refers to the baseline specification that only adds socio-economic covariates. When corresponding, the set of control variables include gender, age (a dummy equal to 1 if the age range is from 25 to 54 years and 0 otherwise), number of jobs, dummies for activity sector (10) and dummies for firm size (terciles according to the number of employees), and a set of dummies for being a public employee, an earner of capital income and an earner of business income. Columns present estimates that use different strategies. "Predicted Income" refers to a specification that uses predicted labor income rather than actual labor income in defining treatment status. "Matching" refers to a regression estimation of a coarsened exact-matched sample of re-weighted treatment and control individuals. Results in columns are grouped by different samples: "All Workers", "Wage Earners", and "Self-Employed". Robust standard errors are clustered at the individual level (shown in parentheses).

Table 3.3: Intensive Margin Responses on Labor Income and Weekly Hours Worked to the 2012 Tax Reform: TAX-SSA Sample

	All Workers			Wage Earners			Self-Employed		
Dep. Var:	Lab. Inc. TAX (1)	Lab. Inc. SSA (2)	Hours Worked (3)	Lab. Inc. TAX (4)	Lab. Inc. SSA (5)	Hours Worked (6)	Lab. Inc. TAX (7)	Lab. Inc. SSA (8)	Hours Worked (9)
A. Reduced Form									
$d\Delta \log y$	-0.045**	-0.055***	-0.009	-0.044**	-0.056***	-0.034**	-0.062*	-0.049	0.104
	(0.018)	(0.019)	(0.022)	(0.021)	(0.022)	(0.018)	(0.037)	(0.032)	(0.095)
B. First Stage									
$d\Delta \log(1-\tau)$	-0.049***	-0.049***	-0.048***	-0.049***	-0.049***	-0.049***	-0.048***	-0.048***	-0.048***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)
C. 2SLS DinD									
$\varepsilon = \frac{d\Delta \log y}{d\Delta \log(1-\tau)}$	0.929**	1.131***	0.181	0.911**	1.157**	0.711*	1.302	1.017	-2.188
	(0.390)	(0.408)	(0.453)	(0.441)	(0.466)	(0.366)	(0.828)	(0.709)	(1.969)
F-statistic	982	981	964	805	805	794	163	163	158
Observations	8,181	8,177	8,181	6,391	6,389	6,391	1,790	1,788	1,790

Notes: This table reports estimates of the intensive margin responses to the tax reform on labor income and hours worked. Each cell corresponds to one regression estimate. The regression analysis is based on the panel of individuals in the TAX-SSA sample observed in the 2009-2014 period and on information from the Tax and SSA records. Row A. shows the results of a reduced-form DinD estimation of specification based on Equation 3.3. Row B. shows the first stage estimates associated to the 2SLS-DinD model in Equation 3.4. Row C. shows the 2SLS-DinD estimated elasticity to the marginal net-of-tax rate based on Equation 3.4, where the dependent variable is the yearly log change of labor income and log change of the marginal net-of-tax rate is instrumented with the DinD interaction between a treatment dummy and a 2012 year dummy. Columns present estimates that use different dependent variables. "Lab. Inc. TAX" refers to yearly log change of labor income based on tax record information. "Lab. Inc. SSA" refers to yearly log change of labor income based on SSA record information. "Hours Worked" refers to yearly log change of weekly hours worked based on SSA record information. Regressions that use log labor income are income-weighted. Hence, the elasticity reported in row C. corresponds to the parameter $\bar{\varepsilon}_L$ in equation (3.7). The "F-statistic" row refers to the F-statistic from the first-stage regression used to evaluate the the strength of the instrument. "Controls" refers to the baseline specification that only adds socio-economic covariates. When corresponding, the set of control variables include gender, age (a dummy equal to 1 if the age range is from 25 to 54 years and 0 otherwise), number of jobs, dummies for activity sector (10) and dummies for firm size (terciles according to the number of employees), and a set of dummies for being a public employee, an earner of capital income and an earner of business income. Columns present estimates that use different strategies. "Predicted Income" refers to a specification that uses predicted labor income rather than actual labor income in defining treatment status. "Matching" refers to a regression estimation of a coarsened exact-matched sample of re-weighted treatment and control individuals. Results in columns are grouped by different samples: "All Workers", "Wage Earners", and "Self-Employed". Robust standard errors are clustered at the individual level (shown in parentheses).

Table 3.4: Extensive Margin Responses on PLIT base to the 2012 Tax Reform. Reduced Form, First Stage and Elasticities

	All Workers				Wage Earners				Self Employed			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Reduced Form												
dP	-0.066***	-0.046***	-0.061***	-0.090***	-0.051***	-0.037***	-0.046***	-0.073***	-0.115***	-0.069**	-0.115***	-0.155***
	(0.012)	(0.010)	(0.013)	(0.015)	(0.012)	(0.010)	(0.013)	(0.015)	(0.032)	(0.029)	(0.035)	(0.041)
B. First Stage												
$d\log(1-\tau^{ATR})$	-0.026***	-0.011***	-0.027***	-0.032***	-0.027***	-0.011***	-0.027***	-0.033***	-0.026***	-0.011***	-0.025***	-0.030***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)
C. 2SLS DinD												
$\mu = \frac{dP}{d\log(1-\tau^{ATR})}$	2.479***	4.195***	2.259***	2.811***	1.914***	3.276***	1.689***	2.238***	4.471***	6.248**	4.573***	5.113***
	(0.459)	(1.055)	(0.479)	(0.482)	(0.462)	(1.007)	(0.472)	(0.477)	(1.174)	(2.682)	(1.297)	(1.288)
F-statistic	814	66	673	1,202	611	57	530	938	217	16	155	305
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predicted Income	No	Yes	No	No	No	Yes	No	No	No	Yes	No	No
Matching	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Triple Difference	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Observations	28,434	28,434	23,086	28,434	21,749	21,749	18,060	21,749	6,685	6,685	5,026	6,685

Notes: This table reports estimates of the extensive margin responses to the tax reform. Each cell corresponds to one regression estimate. The regression analysis is based on the panel of individuals in the *TAX* sample observed in the 2009-2015 period and on information from the tax records. Row A. shows the results of the reduced-form version of Equation 3.5, where the dependent variable is an indicator variable for reporting positive earned income to the PLIT base. Row B. shows the first stage estimates. In this case, the dependent variable is the yearly log change of the marginal net-of-tax rate. Row C. shows the 2SLS-DinD estimated elasticity to the marginal net-of-tax rate, where the dependent variable is an indicator variable for reporting positive earned income to the PLIT base and the log change of the marginal net-of-tax rate is instrumented with the DinD interaction between indicators of treatment status and post-reform period. All regressions are revenue-weighted. Hence, the semi-elasticity reported in row C. corresponds to the parameter $\bar{\mu}_L$ in equation (3.7). The "F-statistic" row refers to the F-statistic from the first-stage regression used to evaluate the the strength of the instrument. "Controls" refers to the baseline specification that only adds socio-economic covariates. When corresponding, the set of control variables include gender, age (a dummy equal to 1 if the age range is from 25 to 54 years and 0 otherwise), number of jobs, dummies for activity sector (10) and dummies for firm size (terciles according to the number of employees), and a set of dummies for being a public employee, an earner of capital income and an earner of business income. Columns present estimates that use different strategies. "Predicted Income" refers to a specification that uses predicted labor income rather than actual labor income in defining treatment status. "Matching" refers to a regression estimation of a coarsened exact-matched sample of re-weighted treatment and control individuals. "Triple Difference" refers to a specification where we interact the DinD variable with an indicator for high intensity of the average tax change. Results in columns are grouped by different samples: "All Workers", "Wage Earners", and "Self-Employed". Robust standard errors are clustered at the individual level (shown in parentheses).

Table 3.5: Income-Shifting Margin Responses to the 2012 Tax Reform. Reduced Form, First Stage and Elasticities

	All Workers			Wage Earners			Self Employed		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. *Contemporaneous Effect*									
Reduced Form (dP)	0.017***	0.012***	0.018***	0.006**	0.004**	0.006*	0.041**	0.034**	0.049**
	(0.005)	(0.004)	(0.006)	(0.003)	(0.002)	(0.004)	(0.019)	(0.017)	(0.020)
First Stage ($d\log(1-\tau^{ATR})$)	-0.023***	-0.009***	-0.023***	-0.023***	-0.009***	-0.024***	-0.022***	-0.012***	-0.021***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
2SLS DinD ($\eta_1 = \frac{dP}{d\log(1-\tau^{ATR})}$)	-0.714***	-1.740**	-0.764***	-0.302**	-0.771**	-0.277*	-1.868**	-0.778	-2.401**
	(0.265)	(0.721)	(0.273)	(0.148)	(0.378)	(0.161)	(0.919)	(2.388)	(1.087)
B. *Anticipation Effect*									
Reduced Form (dP)	0.026***	0.019***	0.026***	-0.000	0.005*	0.000	0.122***	0.057***	0.117***
	(0.007)	(0.004)	(0.008)	(0.002)	(0.003)	(0.003)	(0.029)	(0.018)	(0.030)
First Stage ($d\log(1-\tau^{ATR})$)	-0.021***	0.001*	-0.021***	-0.001	0.000	-0.021***	-0.000	0.004*	-0.022***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
2SLS DinD ($\eta_2 = \frac{dP}{d\log(1-\tau^{ATR})}$)	-1.252***	-2.930***	-1.167***	0.032	-0.715*	0.011	-5.450***	-9.724**	-5.182***
	(0.347)	(0.896)	(0.374)	(0.083)	(0.429)	(0.129)	(1.314)	(4.521)	(1.368)
C. *Implied Semi-elasticity*: $\eta_1+\eta_2$	-1.967	-4.670	-1.931	-0.270	-1.487	-0.266	-7.318	-10.502	-7.584
p-value	0.000	0.000	0.000	0.087	0.019	0.158	0.000	0.010	0.000
F-statistic	298	14	253	208	11	189	92	4	68
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Predicted Income	No	Yes	No	No	Yes	No	No	Yes	No
Matching	No	No	Yes	No	No	Yes	No	No	Yes
Observations	23,793	23,793	19,350	18,140	18,140	15,089	5,653	5,653	4,261

Notes: This table reports results of the income-shifting response across tax bases to the tax reform. Each cell corresponds to one regression estimate. The regression analysis is based on the panel of individuals in the TAX sample observed in the 2009-2015 period and on information from the tax records. Panel A. ("Contemporaneous Effect") and B. ("Anticipation Effect") present estimates for contemporaneous and anticipatory income-shifting responses, respectively. Each panel reports the reduced-form effects, first stage, and 2SLS estimates based on equation 3.6. Panel C. ("Implied semi-elasticity") presents the long-term income-shifting semi-elasticity given by the sum of η_1 and η_2, and the corresponding *p-value* statistic for the null hypothesis of null semi-elasticity. Regressions are revenue-weighted, so that the elasticity reported in Panel C. corresponds to the parameter $\bar{\eta}_{L,C}$ in equation (3.7). The first row in Panels A. and B. show the results from a reduced-form DinD estimation, where the dependent variable is an indicator variable for reporting positive earned income on the CIT returns. The second row shows results the first stage estimates where the dependent variable is the log of the (labor income) average net-of-tax rate. The third row shows the results from a 2SLS-DinD estimation, where the dependent variable is an indicator variable for reporting positive earned income on the CIT returns and the log of the average net-of-tax rates in years t and $t+1$ are instrumented with the DinD interaction between indicators of treatment status and 2011 and 2012 post-period years, respectively. The "F-statistic" row refers to the F-statistic from the first-stage regression used to evaluate the the strength of the instrument. "Controls" refers to the baseline specification that only adds socio-economic covariates. When corresponding, the set of control variables include gender, age (a dummy equal to 1 if the age range is from 25 to 54 years and 0 otherwise), number of jobs, dummies for activity sector (10) and dummies for firm size (terciles according to the number of employees), and a set of dummies for being a public employee, an earner of capital income and an earner of business income. Columns present estimates that use different strategies. "Predicted Income" refers to a specification that uses predicted labor income rather than actual labor income in defining treatment status. "Matching" refers to a regression estimation of a coarsened exact-matched sample of re-weighted treatment and control individuals. Results in columns are grouped by different samples: "All Workers", "Wage Earners", and "Self-Employed". Robust standard errors are clustered at the individual level (shown in parentheses).

CHAPTER 4

Where Do My Tax Dollars Go? Tax Morale Effects of Perceived Government Spending

Matias Giaccobasso, University of California, Los Angeles [1]

Brad Nathan, Columbia University

Ricardo Perez-Truglia, University of California, Berkeley

Alejandro Zentner, The University of Texas, at Dallas

Abstract

Do perceptions about how the government spends tax dollars affect the willingness to pay taxes? We designed a field experiment to test this hypothesis in a

[1] As described in the acknowledgments page, Chapter 4 is a version of a submitted article. The latest version of this article as well as the complement online appendix can be found here. All listed co-authors are principal investigators and contributed in equal shares in the elaboration of the article. We are thankful for excellent comments from Raj Chetty, Matthew Weinzierl, Austan Goolsbee, Steve Levitt, James Poterba, Dario Tortarolo, Sutirtha Bagchi and seminar participants at the NBER-Public Economics, University of Michigan, University of Chicago, University of Chicago-Advances in Field Experiments, RIDGE, IIPF, Journees LAGV, and NOVAFRICA. This project was reviewed and approved in advance by the Institutional Review Board at The University of Texas at Dallas. The field experiment was pre-registered in the AEA RCT Registry (#0007483). To prevent contamination of the subject pool (e.g., that subjects could read about the hypotheses being tested), we posted the RCT pre-registration immediately after the deadline to file a protest had passed, but before conducting any analysis of the data. After the study is accepted for publication, we will share all the code and data through a public repository. Xinmei Yang provided superb research assistance.

natural, high-stakes context and via revealed preferences. We measured how taxpayers perceive the destination of their tax dollars, such as the percentage of their property taxes that funds public schools. We find that even though accurate information is available, taxpayers still hold substantial misperceptions. We use an information-provision experiment to induce exogenous shocks to these perceptions. Using administrative data on property tax appeals, we measure the causal effect of perceived government spending on the willingness to pay taxes. We find that perceptions about government spending have a significant effect on the probability of filing a tax appeal and in a manner that is consistent with reciprocal motivation: individuals are more willing to pay taxes if they believe that the government services funded by those taxes will be of greater personal benefit to them. We discuss implications for the study of tax morale.

4.1 Introduction

Why is tax compliance higher in some countries than in others? Why are some individuals more willing to pay their taxes than others? There are two schools of thought that offer potential explanations: institutions and tax morale. Abundant research shows that institutions have a large effect on tax compliance (Slemrod, 2019b). For example, the introduction of withholding and third-party reporting caused a massive increase in tax compliance (Bagchi and Dušek, 2021). In contrast, little causal evidence shows that tax morale actually matters (Luttmer and Singhal, 2014). In this paper, we attempt to advance our understanding of tax morale by means of a natural field experiment, in a high-stakes context and via revealed-preferences.

Tax morale encompasses various potential mechanisms. We focus on one specific mechanism: our hypothesis is that individuals are more willing to pay taxes if they believe that the government services funded by those taxes will be of greater personal benefit to them. Our hypothesis is related to what Luttmer and Singhal (2014) calls *reciprocal motivation*: "the willingness to pay taxes in exchange for benefits that the state provides to them (...) even though their pecuniary payoff would be higher if they didn't pay taxes." Our hypothesis also relates to a normative principle known as *benefit-based taxation*, which can be briefly described as the "idea of basing tax liabilities on how much an individual benefits from the activities of the state" (Weinzierl, 2018). To test our hypothesis, we conducted an experiment to determine how taxpayers' *perceptions* of how the government spends tax dollars affect their willingness to pay taxes.

Our experiment leverages the context of property taxes, which represents an important source of revenue for governments in the United States and around the world. For instance, U.S. property tax revenues in 2019 were estimated at $577 billion (Tax Policy Center, 2021a), nearly three times higher than the corporate income tax.[2] In the United States, virtually all counties rely heavily on property taxes to fund key government services such as schools, parks, and roads. School funding typically makes up the largest component of property taxes.

This context offers two key advantages to test our hypothesis of reciprocal motivation. First, our research design leverages the straightforward path between property taxes and the government services they fund, allowing to directly identify who benefits from what. For instance, households with children enrolled in local public schools benefit directly from publicly funded education, whereas households with no children enrolled in local public schools do not. For the sake of brevity,

[2]For reference, the 2019 federal income tax generated $1.717 trillion in revenue and corporate income tax generated $230 billion (Tax Policy Center, 2021b).

we refer to households with children enrolled in public schools as "households *with* children" and those without as "households *without* children." The second advantage of this setting is that we can study the willingness to pay taxes via revealed preferences using households' decisions to file property tax appeals, also known as tax protests (Nathan et al., 2020). Filing an appeal is a consequential, high-stakes action that households can take to reduce the amount they have to pay in property taxes.[3] In a nutshell, households can use the subjective nature of the appraisal process in their favor. If they feel like their taxes are too high, they can file a tax appeal to reduce their tax burden.[4]

We conducted a field experiment in Dallas County, Texas. We focus on this county primarily because, from a logistical perspective, it is more practical to implement a field experiment in a single location. However, to the extent that property tax appeals work similarly in other places, the results can be extrapolated to other settings. Dallas County is the second-largest county in Texas, with an estimated population of about 2.6 million in 2020 (U.S. Census Bureau, 2021) – indeed, Dallas County alone has a larger population than 15 of the 50 U.S. states. The county also is diverse along many dimensions, such as ethnicity, and has a relatively even distribution of Democrat and Republican supporters.[5]

We sent a letter to a sample of households in Dallas County inviting them to participate in an online survey. Our main subject pool comprises 2,110 respon-

---

[3]When studying attitudes towards taxation, economists and other social scientists rely primarily on survey data. However, survey data have some well-known limitations, such as social desirability bias. For example, some individuals may *say* that they are willing to pay more in taxes but would *choose* otherwise when facing real stakes.

[4]For more details about how tax protests work, see the discussion in Section 4.2.3 and also Nathan et al. (2020) and Jones (2019).

[5]For example, in the 2012 presidential election, Barack Obama received 57% of the votes in Dallas County, whereas Mitt Romney received 42% (the remaining 1% of votes went to third-party candidates).

dents who completed the survey between April and May of 2021, during which subjects could file a protest of their property taxes with the county. Our survey elicited whether the household has children enrolled in public schools to identify which subjects benefit directly from public school spending and which do not. We conducted an information-provision experiment a few weeks before households faced the opportunity to file a tax appeal. We then matched survey responses to administrative records from the county assessor's office. The rich administrative data allowed us to determine, among many other things, if the survey respondent subsequently filed a tax appeal.

Our experimental design can be summarized as follows. First, we measure respondents' perceptions about the share of their own property taxes that corresponds to school taxes and thus funds public school spending. For brevity, in the remainder of the paper we refer to this percentage as the household's "school share." The school share for the average household in Dallas County is about 49.78%.[6] We can measure the respondents' misperceptions about where their tax dollars go by comparing their guesses about the school share to the true estimates from administrative records. To study the causal effect of beliefs about government spending, the survey embeds an information-provision experiment. After eliciting respondents' prior beliefs, we inform a random half of them about the true value of their respective school shares. By doing so, we can assess how that information affects their posterior beliefs, as measured by our survey, and their decisions to file a tax appeal, as measured by administrative data.

The information-provision experiment creates exogenous variation in respondents' posterior beliefs about the fraction of their property taxes that funds local

[6] This average is calculated over 400,192 properties in Dallas County and excludes commercial properties and other non-owner-occupied residencies (for details, see Online Appendix). Unless explicitly stated otherwise, all statistics about Dallas County are based on this sample.

schools. To illustrate, a subject who perceives her or his school share amount to be 30% may be informed that the actual share is 50%. According to the reciprocal motivation and as noted in the randomized control trial (RCT) pre-registration, the expected effects of the information shock depend on whether the household has children enrolled in public schools. Upon learning that the school share is higher than initially thought, households *with* children should become less likely to file a tax appeal because they learn that they benefit more from government services than they originally believed. Conversely, households *without* children enrolled in public schools should become more likely to file a tax appeal because they learn that they benefit less from government services than they originally thought.

The principle of reciprocal motivation could have implications for tax redistribution. When taxpayers learn that their tax dollars are being spent in communities other than their own, they may be less willing to pay taxes because they do not receive benefits from the taxes they pay. We explore this additional hypothesis using a second treatment arm. Specifically, we leverage the significant redistribution of property taxes across school districts that occurs in some states. In Texas, for example, this redistribution is dictated by legislation often referred to by the media as the "Recapture Plan" or the "Robin Hood Plan."[7] Thus, in the second treatment arm, we measure households' perceptions about the share of their school funding that is redistributed away from, or towards, their own school district. For the sake of brevity, in the remainder of the paper, we refer to this as the "recapture share." For example, a recapture share of 50% would imply that half of the school tax revenue from an advantaged district is transferred to disadvantaged school districts.

We can measure the causal effects of the perceived recapture share using the

[7]For the full history of property tax recapture in Texas, see for example Villanueva (2018).

information-provision experiment. According to the mechanism of reciprocal motivation, the belief about recapture share should not affect the decision to file a tax appeal of households *without* children, because the diverted funds are being used for a service that does not benefit them directly anyways. By contrast, households *with* children should be more likely to protest upon learning that some of their tax payments are being diverted to other districts, because they were benefiting directly from the diverted funds.

The average subject in our sample owns a home worth $349,988 and pays $7,738 in annual property taxes.[8] Households show significant variation in the extent to which they benefit from public education, which is important for our research design. For example, households *with* children accounted for 25.5% of the sample, and households *without* children accounted for the remaining 74.5%.[9] We also find significant variation in how the recapture system affects school districts in our sample, with some school districts diverting as much as 57% of their school districts' property taxes and others receiving as much as 23% additional funds from other districts.[10] Owners can protest "directly" on their own, which is the main focus of this paper, or they can hire an agent to protest on their behalf. For reference, 30.1% of the homeowners in the control group (i.e., those who did not receive any information on school taxes or on recapture) protested directly in 2021. These tax protests are consequential. For instance, 65.4% of protests led to a decrease in assessed home value, resulting in average tax savings of $579 in the first year alone.

[8]These estimated taxes are prior to any adjustments resulting from tax appeals.

[9]We show below that our school taxes information treatment is statistically significant for both of these types of households.

[10]The reported numbers refer to net transfers.

The results from the first treatment arm indicate that even though the information is publicly available and easily accessible, most households have misperceptions about their respective school shares. When provided with factual information, we observe that households strongly update their beliefs. We leverage the information shocks from the experiment to estimate the causal effects of these beliefs and find effects that are consistent with predictions of the framework of reciprocal motivation. Upon learning that their school shares are higher, households *with* children become *less* likely to protest, whereas households *without* children become *more* likely to protest. The effects of the perceptions about government spending are statistically and economically significant. Our baseline estimates imply that increasing the (perceived) school share by 10 percentage points (pp) would cause a drop of 3.67 pp in the probability of filing a protest among households *with* children and an increase of 2.78 pp in the probability of protesting among households *without* children. The effects amount to 11% and 10% of the corresponding baseline protest rates, respectively. These results are robust to a host of alternative specifications and falsification tests.

To assess whether the results were surprising or predictable, we conduct a forecast survey using a sample of 56 experts, most of whom are professors researching related topics. After receiving a brief explanation of the experiment, the experts are asked to forecast the experimental findings. Only a few of them were able to accurately predict the experimental findings. Most experts predicted that the beliefs on school share would have no effect on the likelihood of filing a tax appeal.

The results of the second treatment arm, about share of funds being recaptured, are unfortunately very imprecisely estimated and thus largely inconclusive. We find that respondents have significant misperceptions about the recapture share and that they update their beliefs significantly when provided with information in the experiment. Both the levels of misperception and updating, however, are smaller relative to the corresponding findings for the school share. As a result, the information shocks for the recapture share are not nearly as strong as those for the school share. Thus, the causal effects of the beliefs about the recapture share are very imprecisely estimated. It is important to note that the level of misperceptions and belief updating is difficult to anticipate before conducting the experiment. So, while ex-ante we expected to be well-powered for both treatment arms, ex-post we found out this was not the case for the second treatment arm. In an effort to mitigate publication bias (DellaVigna and Linos, 2022), we still report the analysis for the second treatment arm. Consistent with the hypothesis of reciprocal motivation, the belief about recapture share does not have significant effects on the decision to file a tax appeal among households *without* children – although this finding must be taken with a grain of salt due to the lack of sufficient statistical power. We do not find evidence of significant positive effects for households *with* children – however, the coefficient is so imprecisely estimated that we cannot rule out large positive effects.

Property taxes work almost identically in other counties in Texas and similarly across the country (Dobay et al., 2019; World Bank, 2019; Nathan et al., 2020).[11] These similarities imply that our results from Dallas County can be reasonably generalizable to other U.S. counties. Moreover, replicating our field experiment in other U.S. counties would be straightforward. Indeed, we propose the use of property tax protests as a novel context to study taxpayers' preferences and tax compliance.[12] We provide detailed accounts of the implementation and

[11] For instance, property taxes provide a significant source of school funding in most of the U.S. (Chen, 2021), and other states also redistribute property taxes across school districts, similar to Texas' recapture system (Youngman, 2016).

[12] One notable advantage of our setting is that it uses publicly available data, which facilitates

data sources that other researchers can follow, and we are happy to share data, code, tips, and additional resources.

Our study relates and contributes to the literature on the role of tax morale in tax compliance decisions. Unlike the vast amount of causal evidence showing that institutions matter, there is little causal evidence showing that tax morale matters (Luttmer and Singhal, 2014). We contribute to this literature by providing novel evidence that tax morale can be a significant factor in practice. Moreover, we provide methodological innovations that other researchers can follow to better explore the role of tax morale.

Indeed, far from showing that tax morale matters for tax compliance, the existing causal evidence seems to suggest that its role is negligible. For example, a growing literature on correspondence experiments studies how moral suasion affects tax payments (Slemrod, 2019b). These experiments consist of sending messages to taxpayers highlighting that paying taxes is the right thing to do and then measuring the effects of those messages on subsequent tax compliance. The moral suasion messages have been found to be largely ineffective (Slemrod, 2019b). By comparison, messaging about institutions (e.g., audits, penalties) have been found to be highly effective. Based on this evidence, one natural interpretation is that tax morale is not important and that only institutions matter for tax compliance. Our results challenge this view: tax morale matters, but the existing correspondence experiments are ill-conceived to uncover the effects of tax morale.

Two innovations in our methodology allow us to shed light on tax morale, both of which are possible largely due to the novel research design linking data from a survey experiment to administrative tax data at the individual level. This

---

replication efforts and avoids potential conflict of interests in partnerships with government organizations.

approach is new to this stream of literature and rare even in broader economics research (Bergolo et al., 2020b). First, moral suasion messaging in previous research typically has sought to affect individuals' tax morale by influencing individuals' preferences. However, such preferences are based on historical life experiences and may be too hard to change with a simple message (e.g., "it is important to contribute your part"). Instead of trying to influence preferences, we propose to study tax morale by inducing changes in beliefs. Indeed, a large literature shows that simple information-provision experiments can have significant and long-lasting effects on perceptions and expectations in a range of topics such as macroeconomic expectations (Cavallo et al., 2017) and salary perceptions (Cullen and Perez-Truglia, 2022).

The second innovation is our ability to measure heterogeneous effects via the linkage between the survey data and the administrative data. In the context of tax morale, there is scope for highly heterogeneous effects of information. As illustrated by our results on the school share, the same piece of information can have effects in opposite directions for different groups of subjects (i.e., households *with* children vs. households *without* children). It is possible that these large effects across different groups cancel each other out, on average, which would lead to the erroneous conclusion that tax morale is irrelevant for tax compliance. Using survey data to identify which households have kids enrolled in public schools and which do not, we can measure the effects of the information separately for each group.[13] Another reason to expect heterogeneous effects relates to how subjects update their beliefs in response to new information. Households that underestimate their school share may adjust their beliefs upward when given accurate information, whereas

[13] Castro and Scartascini (2015) is a notable exception, as they provide direct evidence on treatment heterogeneity.

households that overestimate their school share may adjust their beliefs downward when provided with the same information. Again, it is possible that these large effects across different groups cancel each other out, on average, which would lead to the erroneous conclusion that tax morale is irrelevant for tax compliance. Our survey allows us to measure prior and posterior beliefs, thus allowing us to fully elucidate the effects of the information.

Our findings are related to a few other studies. Some of the previously mentioned correspondence experiments are especially relevant because they include a treatment arm with a message related to the importance of taxes for the provision of community services (Blumenthal et al., 2001; Castro and Scartascini, 2015; Bott et al., 2020; De Neve et al., 2021; Bergolo et al., 2021).[14] As described above, our methodological innovations allow us to unpack effects that would have been otherwise hidden. Our study is also related to Cullen et al. (2020), who provide suggestive evidence of tax morale by showing that tax compliance can change with the partisan alignment of the government.[15] In another related study, Carrillo et al. (2021) treated a sample of 400 taxpayers from an Argentine municipality with a joint intervention that recognized them publicly for their good behavior and awarded them with the construction of a sidewalk near their homes. They provide evidence that this joint intervention decreased subsequent tax delinquency, although it is unclear whether the effects are due to the public recognition, the

[14]The evidence is also mixed. Blumenthal et al. (2001) find their message does not have a significant effect on tax evasion. Bergolo et al. (2021) and Bott et al. (2020) find significant negative effects in the first year, but the effects do not persist after a year. De Neve et al. (2021) find that information about government spending increases knowledge and appreciation about how taxes are spent but does not affect tax compliance. Castro and Scartascini (2015) find insignificant average effects.

[15]Huet-Vaughn et al. (2019) provide related laboratory evidence showing that the ideological match between the taxpayer and specific tax expenditures affects the willingness to pay taxes.

sidewalk, or both.[16] We contribute to this literature by disentangling a potential mechanisms at play: reciprocal motivation.[17] Beyond tax compliance, recent quasi-experimental evidence demonstrates how the salience of government spending can affect electoral outcomes (Huet-Vaughn, 2019; Ajzenman and Durante, 2022).

Our study also relates to a small but growing literature on the interplay between tax policy and normative considerations. The normative principle of benefit-based taxation was a prominent, and at times leading, approach among tax theorists in the early twentieth century (Seligman, 1908; Musgrave, 1959). However, it has been largely ignored by the modern optimal taxation literature, which instead focuses solely on efficiency aspects of taxation (Weinzierl, 2018; Scherf and Weinzierl, 2020). Moreover, a growing body of work seeks to incorporate other normative considerations into tax policy design (Mankiw and Weinzierl, 2010; Weinzierl, 2014; Saez and Stantcheva, 2016).[18] This literature is new and mostly theoretical, with empirical evidence limited to survey data, such as asking individuals to choose between hypothetical tax policies (Weinzierl, 2014; Saez and Stantcheva, 2016; Weinzierl, 2017). We contribute to this literature by providing the first revealed-preference evidence on this topic from a natural, high-stakes context.

We conclude by discussing some policy implications. Our evidence highlights

[16] Relatedly, Kresch et al. (2023) provide evidence showing that households with access to the city sewer system are more likely to pay property taxes.

[17] Also related to our study, Nathan et al. (2020) provide evidence that perceptions about the average tax rate affects households' decisions to file a protest. Although this result does not pertain to perceptions of government spending, it constitutes consistent evidence that fairness concerns play a significant role in the decision to file a tax appeal. Cait et al. (2018) provide evidence from a laboratory experiment showing that tax payments increase when participants have the opportunity to voice their preferences for how their tax dollars are to be spent.

[18] For instance, the normative considerations related to equality of opportunity or poverty alleviation.

the complexity of the use of transparency policies and communication strategies to boost tax compliance. For instance, messages sharing information on government services may have mixed effects on tax compliance based on whether the recipient of the message benefits from the advertised service. Indeed, our framework can explain the mixed results in the moral suasion literature, and our findings suggest some strategies that could raise average tax compliance.

The rest of the paper proceeds as follows. Section 4.2 describes the institutional context. Section 4.3 presents the conceptual framework. Section 4.4 discusses the experimental design and implementation. Sections 4.5 and 4.6 present the results. The last section concludes.

4.2 Institutional Context

4.2.1 Property Taxes and Public Schools

In Dallas County, property taxes fund various public services, such as schools, parks, roads, and police and fire departments. In 2021, the average home in Dallas County was worth $327,690, and the average estimated property tax bill was $6,370, implying an effective tax rate of 1.94%. Texas does not have a state income tax. To compensate, revenues from property taxes fund a greater share of local government services in Texas than in many states. School districts receive the largest share of a household's property tax, accounting for nearly half (49.78%) of the average total property tax bill.[19] The second-highest component is the city tax (accounting for roughly 28% of property taxes), followed by hospital (10%),

[19] Variation in the school share across households ranges from 13.2% (1st percentile of the distribution) to 90.8% (99th percentile).

county (8%), college (4%), and special district (¡1%) taxes.[20]

Dallas County has 14 main Independent School Districts (ISDs).[21] Homeowners who live within the geographical boundaries of a given ISD jurisdiction are subject to the tax rate for that ISD.[22] Households also have the right to send their children to the K–12 public school(s) in their ISD. All households must pay school taxes, regardless of whether they have children enrolled in public schools. The public schools in Dallas County are generally of good quality, although significant differences exist.[23] Alternatively, homeowners can send their children to private schools, conduct homeschooling, or enter a lottery for a chance to send their children to a charter school.[24] Sending children to private schools can be expensive, however. The average tuition cost for private schools in Dallas County is $12,374 per student as of 2022.[25] According to data from the 2020 U.S. Census, about 90% of K–12 students in Dallas County attend a public school.[26]

[20] See Online Appendix for more details.

[21] The total number of ISDs is sixteen, but two of them are extremely small and thus are excluded from the analysis. See Online Appendix for more details.

[22] School districts in Texas can set their own tax rates, but they must abide by certain state regulations. See Online Appendix for more details.

[23] For example, according to www.GreatSchools.org, 100% of the schools in the Highland Park ISD have above-average ratings in Texas, whereas 43% of schools in the Mesquite ISD have below-average ratings (data accessed on November 4, 2021).

[24] Charter schools are tuition-free public schools that receive funding directly from the state and do not receive funding from property taxes.

[25] Data accessed from https://www.privateschoolreview.com/exas/dallas-county on January 5, 2022.

[26] More precisely, 89% of kindergarten students and 92.5% of students in grades 1–12.

4.2.2 Property Tax Recapture

To make public school funding more equitable across school districts, Texas enacted a redistribution system in 1993, called the "Recapture Plan" or "Robin Hood Plan", to divert school tax funds from "property-wealthy" districts to "property-poor" districts.[27] Due to the large tax amounts involved, the recapture system has been a topic of heated debate among politicians and the general public (Dallas Morning News, 2018). The recapture system has been amended several times since its inception, including a change in 2019 that slowed down the strong growth in the amount recaptured. Nevertheless, redistribution amounts remain substantial under the current recapture formula (Texas Education Agency, 2021).[28]

In this paper, we focus on the *net* redistribution, which is the difference between the taxes recaptured by the state from the district (if any) and the amount distributed from that state pool to the districts (for specifics on the recapture formula and this calculation, see Online Appendix). Wide variation in the recapture share occurs across the 14 ISDs that we study. Four ISDs are net givers: the highest giver is Highland Park ISD, which has 57.3% of its school taxes diverted. The remaining ten districts are net receivers: the highest receiver is Mesquite ISD, which receives an additional 23.3% in funding from property taxes diverted from other districts.

4.2.3 Tax Protests

Each year, the Dallas Central Appraisal District (DCAD) conducts market value appraisals for all homes in the county. Each appraisal results in a "proposed value"

[27] This system was the result of poor school districts legally challenging the system of state school finances in the late 1980s and early 1990s on state constitutional grounds.

[28] See Online Appendix for more details.

for the home, which is an estimate of the home's market value as of January 1 of that year. The DCAD makes this information available to all homeowners through its website and/or by mail.[29] The notice includes additional information, such as the estimated taxes due based on the property's proposed values and how property taxes are allocated across jurisdiction types (e.g., school taxes, city taxes). After the notifications are sent, households have a month from the notification date to file a protest if they disagree with the proposed value. In 2021, the DCAD notified the proposed values on April 16; as a result, the deadline to protest was May 17.

Homeowners can file a protest by mail using a form included with their mailed notice, or they can file a protest online using a simple tool called uFile.[30] After reviewing the argument, the DCAD can (and often does) make an offer by mail or phone to reduce the assessed home value. If the homeowner refuses this settlement value or the DCAD does not offer a settlement, the appeal proceeds to a formal hearing with the Appraisal Review Board.[31] Once protests are resolved, the new tax amount becomes payable either immediately or at the billing date if it is later (i.e., on October 1st in 2021). Any unpaid taxes eventually become delinquent (e.g., unpaid 2021 property taxes became delinquent on January 31, 2022).

A key feature of this setting is the difficulty in estimating home market values for homes that have not been sold recently, a process involving significant ambigu-

[29] A sample notification, called the "Notice of Appraised Value", is shown in Online Appendix. This notification is available online for every household, and it is also sent by mail to some households (e.g., households with proposed values that increased since the previous year).

[30] To protest online, homeowners need to look up their account (e.g., searching for their own names or addresses) and then follow some straightforward steps in the uFile system. To protest by mail, households who received a notification from the DCAD can use the protest form included with the notification, and households that did not receive a notification can file by mailing a printed form that can be obtained online on either the DCAD's or the Texas Comptroller's website. In 2020, about 75% of direct protests were filed online while the remaining 25% were filed by mail (Nathan et al., 2020).

[31] Homeowners can contest the Appraisal Review Board's decision in court.

ity and subjectivity. To avoid costly in-person appraisals, the DCAD uses statistical models and large datasets (e.g., recent home sales) to formulate an estimated market value for each property. However, even multibillion-dollar companies like Zillow and Redfin have a hard time estimating market values using statistical models (Parker and Friedman, 2021). This ambiguity in home value is important for the interpretation of our results because it implies that households are not trying to objectively "correct" estimates from the DCAD. Instead, they are presenting a data point (e.g., the sale price of a neighboring home) to support their protest. This distinction is consistent with what was expressed in our conversations with officials from some of the county appraisal districts in Texas. Their prevailing view is that households use the subjective nature of the appraisal process as an excuse to complain about their taxes being too high (for more details, see Nathan et al., 2020) and not necessarily to complain about the county's estimate of their home value.

4.3 Conceptual Framework

To formalize the logic of reciprocal motives, we introduce a simple model of how the provision of government services and redistribution affects the decision to file a protest. Let subscript $j \in \{C, NC\}$ represent the two types of households: those *with* children enrolled in public schools ($j = C$) and those *without* ($j = NC$). Let P_j be the outcome of interest: the probability that the household files a tax protest, which is a proxy for its (un)willingness to pay taxes. Let B_j be how much households in group j benefit from each dollar spent in government services. Consider the following relationship:

$$P_j = \gamma \cdot B_j \tag{4.1}$$

Motivated by the theory of reciprocal motivation, we assume $\gamma < 0$: that is, when households benefit directly from government expenditures, they are less likely to protest their taxes. Let S be the government expenditures in the local public school district and NS be the government expenditures in other local government services (e.g., police, parks, and roads). The two types of households benefit from the two types of government expenditures in the following manner:

$$B_C = \alpha^S \cdot S + \alpha^{NS} \cdot NS \tag{4.2}$$

$$B_{NC} = \alpha^{NS} \cdot NS \tag{4.3}$$

where parameters α^S and α^{NS} capture how households benefit from different types of expenditures. The parameter α^S denotes how much a household *with* children enrolled in public school benefits per dollar spent in public schools. α^{NS} denotes how much households (regardless of whether they have children) benefit per each dollar spent in non-school government expenditures. The key assumption is that households *with* children in public schools benefit more from school expenditures than from non-school expenditures: $\alpha^S > \alpha^{NS}$. This assumption is meant to represent the fact that unlike the benefits from non-school expenditures (e.g., police, roads), which are spread over the entire community, the benefits from school expenditures are concentrated on a subset of the population (households *with* children enrolled in public schools) and thus the members of that subset enjoy them more.

Next, we conduct a simple normalization. Let $G = S + NS$ denote total expenditures and $s = \frac{S}{G}$ denote school expenditures as a fraction of total expenditures, which we previously defined as school share. For the sake of simplicity, we do not incorporate misperceptions into this simple framework. In practice, however, the "s" that matters is the one perceived by the taxpayer when deciding whether to protest. We thus can re-write equations (4.2) and (4.3) as follows:

$$B_C = G \cdot (\alpha^S \cdot s + \alpha^{NS} \cdot (1 - s)) \tag{4.4}$$

$$B_{NC} = G \cdot \alpha^{NS} \cdot (1 - s) \tag{4.5}$$

Combining equations (4.1), (4.4), and (4.5), we obtain the following:

$$P_C = \gamma \cdot G \cdot (\alpha^S \cdot s + \alpha^{NS} \cdot (1 - s)) \tag{4.6}$$

$$P_{NC} = \gamma \cdot G \cdot \alpha^{NS} \cdot (1 - s) \tag{4.7}$$

Using equations (4.6) and (4.7), we can see what happens to protest rates if the school share increases:

$$\frac{\partial P_C}{\partial s} = \gamma \cdot G \cdot (\alpha^S - \alpha^{NS}) < 0 \tag{4.8}$$

$$\frac{\partial P_{NC}}{\partial s} = -\gamma \cdot G \cdot \alpha^{NS} > 0 \tag{4.9}$$

The intuitions are straightforward. Households *with* children benefit most from school expenditures. Thus, an increase in s implies that they benefit more from government services and that their probability of protesting decreases. In contrast, households *without* children do not benefit from school expenditures. Thus, when s increases, their benefits from government services go down and their probability of protesting goes up. Moreover, if we subtract equation (4.9) from (4.8), we obtain the following:

$$\frac{\partial P_C}{\partial s} - \frac{\partial P_{NC}}{\partial s} = \gamma \cdot G \cdot \alpha^S < 0 \tag{4.10}$$

In other words, the difference in the effect of s between households *with* children versus those *without* children can be tracked to a key parameter of interest, α^S, which is how much households *with* children benefit from school expenditures.

Prediction 1. *An increase in the school share should negatively affect the protest probability of households with children in public schools and positively affect the protest probability of households without children in public schools.*

This setup corresponds to the simplest case and is based on two simplifying assumptions. First, it assumes that households are entirely selfish and that households *without* children do not benefit at all from school spending, although in practice these taxpayers may feel good about helping other parents in the community. Second, it assumes that benefits from non-school services are the same for households *with* children as for households *without* children in public schools. We choose this setup due to its simplicity, but in Online Appendix, we show that some of the main predictions still holds under more general assumptions.

It is straightforward to extend this simple model to include redistribution of school taxes. For the sake of brevity, we consider the analysis from the perspective of a household in a wealthy school district whose school taxes are redistributed to disadvantaged school districts.[32] Let $r \in [0, 1]$ represent the fraction of school taxes that are transferred from the household's own school district to other school districts, which we previously defined as the recapture share. For instance, $r = 0.4$ would indicate that 40% of school taxes are redistributed to other school districts. We can extend equations (4.2) and (4.3) to incorporate recapture into the model:

$$B_C = \alpha^S \cdot S \cdot (1 - r) + \alpha^{NS} \cdot NS \tag{4.11}$$

$$B_{NC} = \alpha^{NS} \cdot NS \tag{4.12}$$

We normalize equations (4.11) and (4.12) by total expenditures, combine them with equation (4.1), and then rearrange them as follows:

$$P_C = \gamma \cdot G \cdot (\alpha^S \cdot s \cdot (1 - r) + \alpha^{NS} \cdot (1 - s)) \tag{4.13}$$

$$P_{NC} = \gamma \cdot G \cdot \alpha^{NS} \cdot (1 - s) \tag{4.14}$$

Using these equations, we can see what would happen if we increase the recapture share:

$$\frac{\partial P_C}{\partial r} = -\gamma \cdot G \cdot s \cdot \alpha^S > 0 \tag{4.15}$$

$$\frac{\partial P_{NC}}{\partial r} = 0 \tag{4.16}$$

---

[32] The forces at play are similar from the opposite perspective, wherein a disadvantaged district receives funds from more advantaged districts.

The intuitions are straightforward. Households *without* children in the school district do not benefit from school taxes, regardless of which school district receives the funding, so their willingness to pay taxes is unaffected by recapture. For households *with* children, more recapture means fewer benefits for their local school district and thus less willingness to pay taxes.

We can also subtract (4.16) from (4.15) to show the following:

$$\frac{\partial P_C}{\partial r} - \frac{\partial P_{NC}}{\partial r} = -\gamma \cdot G \cdot s \cdot \alpha^S > 0 \tag{4.17}$$

Again, the difference in effects between households with children and without children is determined by parameter α^S.

Prediction 2. *An increase in the recapture share should increase the protest probability for households with children in public schools, but it should not affect the protest probability for households without children in public schools.*

This framework assumes that households are totally selfish and care only about how they benefit from government services. In practice, households may appreciate that their tax dollars help the community. In Online Appendix, we provide an extension of this framework that incorporates such altruism. We must keep in mind, however, that Prediction 2 no longer holds once we allow for altruistic taxpayers.

4.4 Experimental Design and Implementation

4.4.1 Subject Recruitment

We mailed our letters so that they would be delivered close to the time that homeowners in Dallas County could start filing tax appeals. Online Appendix shows a sample envelope, and Online Appendix shows a sample letter. We included several features to indicate the legitimacy of the letters. For example, the letters were sent on behalf of researchers at The University of Texas at Dallas, a well-known institution in Dallas County. The envelope featured the school's logo, the name of a professor from that university, and non-profit organization postage. The letter itself included a physical address for the researcher and a link to the study's website (see Online Appendix for a screenshot of the website). It also provided contact information for the researchers and Institutional Review Board. The letter salutation included each recipient's name, and recipients' names and addresses were printed at the bottom of the second page so that they appeared through the envelope window. In cases where properties were jointly owned by multiple individuals (typically, husband and wife), we sent one letter to the address but listed all owners on the letter. As previously mentioned, the letter also mentioned the recipient household's proposed value and estimated property tax amount for 2021.

Most importantly, our letters included an invitation to participate in an online survey and included the URL of the survey. Each subject was asked to enter a unique survey code, which was included in the letter right next to the survey URL. This code allowed us to identify survey respondents and link their responses to the administrative records. In addition to the opportunity to contribute to a research study, we included two additional incentives for survey participation. First, the letters indicated that detailed, step-by-step instructions on how to file a protest online or by mail would be provided at the end of the survey.[33] As a second incentive, subjects were informed that survey respondents would be entered into a raffle for 20 prizes worth $100 each.[34]

4.4.2 Survey Design

In this section, we summarize the main features of the survey.[35] We start by asking a critical question, that is, whether the respondent's household has children enrolled in grades K–12 at their local public school district, and if so, how many. This critical information is missing from administrative records of the tax agency and thus the analysis would be impossible without this question, particularly the heterogeneity analysis concerning the framework of reciprocal motivation, which is the main form of heterogeneity that we anticipate in the RCT pre-registration.

The module about school taxes can be summarized as follows:

- **Step 1 (Elicit Prior Belief):** We begin by providing the estimated total property tax amount of the respondent's home in 2020 (based on administrative records). We then explain that this total amount is the sum of different components, such as school, city, and hospital taxes. We ask respondents to

[33]This walkthrough included hyperlinks to relevant websites and screenshots of a sample protest using information for a fictitious household for added clarity. To access these instructions, subjects were provided with a URL and a code on the final screen of the survey. A copy of the web instructions is included in Online Appendix. Nathan et al. (2020) show that these instructions have a significant positive effect on the probability of protesting.

[34]All respondents were entered into the same raffle, but only a random half of respondents were informed about the raffle in the letter (i.e., before deciding whether to participate in the survey). This randomization aimed to assess the effectiveness of raffle prizes in increasing response rates, which can be useful information for future researchers conducting similar field experiments. For more details, see Online Appendix.

[35]A sample of the full survey instrument is attached in Online Appendix.

guess their school share in 2020, using any amount between 0% and 100%.

- **Step 2 (Information-Provision Experiment):** For every subject, we calculate the "correct" answer to the previous question based on administrative records. We then randomize whether the subject sees the correct answer. Each subject faces a 50% probability of being shown this information. To avoid respondents making inferences from the act of receiving information, we make the randomization explicit. On the first screen, we inform respondents that some participants will be randomly chosen to receive the information and that they will find out on the next screen if they are selected. On the following screen, we inform subjects whether they are chosen to receive the feedback.

- **Step 3 (Elicit Posterior Belief):** We give all subjects the opportunity to revise the guess they provided in Step 1. To avoid asking the exact same question about their 2020 taxes (i.e., the year prior to our intervention), we instead ask about their 2021 taxes (i.e., the most recent year). To avoid subjects making inferences based on the opportunity to re-elicit their guesses (e.g., subjects inferring that we ask again only if their answer in step 1 is incorrect), we explicitly inform them that all survey participants have this opportunity, regardless of their initial guesses.

To learn about the causal effects of beliefs, it is critical to leverage information on prior beliefs. When provided with feedback during the information-provision experiment, individuals who underestimate may update their beliefs upward and those who overestimate may adjust their beliefs downward. Some individuals may have accurate priors and thus may not make any updates. Whether an individual's probability of protesting increases, decreases, or remains the same should depend on the individual's beliefs before receiving the information. For this reason, we conduct the information-provision experiment within the survey, as opposed to providing the information directly in the letter, to measure beliefs prior to information provision. To leverage the effect of the information on prior beliefs, we use the same econometric models used in other information-provision experiments (see e.g., Cullen and Perez-Truglia, 2022; Bottan and Perez-Truglia, 2022).

The following module is about the recapture share.[36] Some subjects may not know about or understand recapture. Thus, we start with a couple of short paragraphs summarizing the recapture system. The rest of the module follows the same structure as previously described for steps 1 through 3. We elicit beliefs about the recapture share in two steps. First, we ask respondents to guess if their school district will receive more, the same, or less in taxes than what households in their district paid in school taxes. The second step is quantitative in nature. If the respondent selects "More" (or "Less") in the first question, we ask them to guess how much *more* (or *less*) funding their school district will receive as a share of the district's school tax revenues due to recapture, using any amount between 0% and 100%. We then conduct step 2 (information-provision experiment) and step 3 (elicitation of posterior beliefs).

We cross-randomize subjects to receive the two pieces of information about school taxes and recapture, respectively, with a 50% probability for each. Thus, roughly 25% of the sample receives both pieces of information, 25% receives the first piece of information only, 25% receives the second piece of information only, and 25% receives no information at all.

These questions comprise the core of the survey. We also include a series of additional questions, including one question that serves as a secondary outcome in

[36] Note that the recapture share is ISD-specific, whereas the school share is household-specific.

the analysis of the effects of beliefs. We ask respondents whether they plan to file a protest this year in a 1-4 likelihood scale. This outcome allows us to pick up short-term effects on the *intention* to protest, even if those effects do not materialize into actual protests. For descriptive purposes, we include questions asking respondents' gender, age, ethnicity, education, and political party. To provide supplemental evidence, towards the end of the survey, we include additional questions that are described in more detail in the following sections.

4.4.3 Subject Pool

We sent the previously described letters to 78,128 households representing a subsample of the universe of all households in Dallas County, Texas. We arrived at this subsample by applying several filters (e.g., excluding commercial properties and non-owner-occupied residences.)[37] When selecting this sample, we stratified the randomization at the ISD level to ensure wide representation of the beneficiaries and contributors to the recapture system.[38] We can link each survey respondent to rich sources of administrative data, including whether the subject protested in any year from 2016 to 2020, as well as detailed information on property ownership, address, number of bedrooms and other features, exemption amounts, taxable values, and tax rates.

We timed the intervention so that our letters would arrive early enough before the protest deadline to influence the recipient's decision. We created the letters on April 16th, 2021, as soon as the administrative data, including 2021 proposed values, became available. To accelerate delivery, we used a mailing company in Dallas County (i.e., the same county as all recipients). The mailing company

[37] For the full inclusion criteria, see Online Appendix.

[38] For more details, see Online Appendix.

dropped the letters off at the local post office on April 20, 2021, and estimated that most would be delivered in the next couple of days. Consistent with this projection, we began to receive survey responses and visits to the study's website on April 22, 2021.[39]

Of 78,128 households invited to the survey, 2,966 answered the first two questions and 2,821 completed the two key modules on posterior beliefs about the recapture share.[40] The implied response rate of 3.6% ($=\frac{2{,}821}{78{,}128}$) is comparable to the response rate of 3.7% from a previous study in this same context and using a similar recruitment method (Nathan et al., 2020). Moreover, the response rate of 3.6% is on the same order of magnitude as the response rate of surveys that use this recruitment method (4.7%, as reported in Sinclair et al., 2012).[41] Among respondents, the median time to complete the survey was 11.3 minutes. Towards the end of the survey, we included an attention check similar to the one used in other studies (Bottan and Perez-Truglia, 2020), which 92.1% of respondents successfully passed. This passing rate is relatively high for a survey study, especially given that the attention check was located at the very end of the survey when fatigue was likely at its highest.

Of the 2,821 survey responses, we drop responses that, as explained in the RCT pre-registration, could not be excluded *ex ante* because of data availability. We drop 36 responses from subjects who, according to the DCAD's records, had already filed a protest before starting our survey and 23 additional subjects who responded to the survey after the deadline to file a protest, as the survey infor-

[39]More details about the timing of survey responses are provided in Online Appendix.

[40]See Online Appendix for more details about the sample and about attrition rates and balance tests.

[41]The 4.7% response rate corresponds to a mailing of a personally-addressed postcard inviting a household to complete a web-based survey using a unique alphanumeric code.

mation could not have affected their decisions to protest. We similarly drop 185 subjects who, according the DCAD's records, had already hired a tax agent before starting our survey.[42]

When studying perceptions via survey data, it is important to deal properly with outlier beliefs. Some individuals may provide guesses that are wildly inaccurate not because they truly hold such extreme beliefs but because they misunderstand the question, make a typo, or just do not pay attention to the question. The "information shocks" for these individuals can be large but meaningless, which can create a significant attenuation bias. To reduce sensitivity to outliers, we follow the standard practice in information-provision experiments and drop respondents with the most extreme misperceptions in their prior beliefs (see e.g., Fuster et al., 2022; Cullen and Perez-Truglia, 2022; Bottan and Perez-Truglia, 2020). For the baseline specification, we use a conservative definition of outliers that drops 467 subjects from the bottom 5% and top 5% of the distribution of prior misperceptions.[43] After applying these filters, 2,110 respondents remain, constituting our main subject pool. Since these exclusions are based on pre-treatment variables (e.g., prior beliefs), they should not compromise the validity of the experimental variation. As a robustness check, we reproduce the analysis with more lax definitions of outliers (results presented in Section 4.5.5). Finally, we provide several sharp falsification tests to address any potential concerns about the internal validity of the results, such as event-study analyses.

Our subject pool self-selects to answer the survey for both households *with* and *without* children. The proportion of households *with* and *without* children who answered our survey, 25.5% and 74.5% respectively, approximately matches

[42] For more details, see Online Appendix.

[43] For more details on the distribution of outlier observations, see Online Appendix.

the proportion of families who have or do not have children in Dallas county, 32.3% and 67.4% respectively (Statistical Atlas, 2023).[44] Column (1) of Table 4.1 presents descriptive statistics about the subject pool. Prior to any adjustment resulting from protests, the average subject owns a home with an assessed market value of $349,988 and property taxes of $7,738 (corresponding to an average tax rate of 2.21%). Around 25.5% of respondents have children enrolled in a local public school, 42.9% are women, 44.3% self-identify as White, 38.3% have a college degree, and on average they are 49.6 years old.

In terms of observable characteristics (e.g., home value, number of bedrooms, or tax rate), the subject pool is similar to the universe of households in the county. Differences between survey respondents and non-respondents are statistically significant but small (see Online Appendix). However, one significant difference is that, relative to the universe of households, respondents to the survey are substantially more likely to file a protest in 2021 and in previous years. By design, our study targets individuals who would seriously consider protesting, which increases statistical power by securing more variation in the outcome variable.[45] Moreover, our letter describes tax protests, so subjects considering filing a protest in 2021 are likely to pay attention to the letter and thus also likely to notice the survey link included in the letter.[46]

[44] Moreover, our identification strategy will arise from comparing households *with* and *without* children and we discuss below how "salience" effects are likely to be similar for these two types of household.

[45] Specifically, when selecting households to participate in the survey, we over-sample those most likely to protest, such as households with a history of increased estimated taxes. For more details, see Online Appendix.

[46] Indeed, this higher propensity to protest among survey respondents is consistent with results from Nathan et al. (2020), who use a similar recruiting method to collect survey responses in this same context. Moreover, our letter promises instructions on how to file a protest as a reward for participation, so it is natural that interested respondents would be more likely to participate. Additionally, these instructions likely make it easier for survey respondents to file an appeal, as

Columns (2) through (5) of Table 4.1 break down the average characteristics in each of the four treatment groups. All characteristics shown in Table 4.1 are determined pre-treatment and thus should not be affected by the treatment assignment.[47] Column (6) reports p-values for the null hypothesis that the average characteristics are equal across the four treatment groups. Table 4.1 shows that, consistent with successful random assignment, the observable characteristics are balanced across treatment groups.[48] In Online Appendix, we present alternative versions of the randomization balance tests, such as breaking the sample down by households *with* and *without* children. We also show that participation in the survey and attrition among participants are orthogonal to treatment assignment, which is expected given that subjects randomly receive treatment(s) after they start the survey.

4.4.4 Outcomes of Interest

As stated in the RCT pre-registration, the main outcome of interest is a dummy variable indicating whether the household protested directly in 2021.[49] To get a sense of the baseline protest rate, we consider subjects in the control group (i.e., those who do not receive any information on school taxes nor recapture). Approximately 30.1% of those owners file a tax appeal in 2021. These tax protests

---

documented in Nathan et al. (2020).

[47]Some questions, such as gender of the respondent, are asked after the information-provision stage. However, treatment assignment should not affect these responses. For instance, we do not expect information on school spending to change responses regarding gender or education level.

[48]The difference is statistically significant for one of the variables (owner protest in 2020). Given the large number of tests conducted, a few differences may be statistically significant just by chance. To be safe and to follow best practices in field experiments (Athey and Imbens, 2017), we include this variable in the set of control variables in all regressions.

[49]The protest variable is based on data downloaded from the DCAD website on June 22, 2021.

are consequential: 65.4% lead to a decrease in the assessed home value, among which the average tax savings were $579 in the first year alone.[50]

Owners can file their own protests, which is the main focus of this paper. For the sake of brevity, in the rest of the paper we use the term "protest" as shorthand for direct protests by the homeowner, unless explicitly stated otherwise. Households also have the option to hire an agent to file a protest on their behalf. In addition to the 30.1% of owners who protest directly, 4.8% use an agent.[51] Due to the nature of the setting and the timing of the protest process, and as stated in the RCT pre-registration, we expect our information to primarily affect whether households file their own protests. First, because we provide information to the households and not to their agents, our experiment should not affect the agent's behavior. Second, while it is possible that the information provided in our survey could influence whether a household hires an agent, that's unlikely due to the nature of the setting. According to conversations with households, tax agents and representatives from assessor's offices, households typically sign contracts with agents well in advance of the date when the proposed values are announced. Some households sign long-term contracts to file protests on behalf of the owner over many years. Last, consistent with the above arguments, in prior work we showed that a mail intervention had large effect on direct protests but no effect on protest through agents (Nathan et al., 2020). For all of these reasons, we study protests through tax agents separately, but in the spirit of a falsification test.

[50] These calculations are based on data downloaded from the DCAD website in December 2021. The remaining protests are either unresolved by December 2021 (12.2%) or resolved with no change in the assessed home value (22.4%).

[51] These statistics refer to 2021 protests in the control group.

4.4.5 Expert Prediction Survey

To assess whether the experimental results are surprising, we conduct a forecast survey with a sample of experts. A sample of the full survey instrument is attached in Online Appendix. In this survey, which follows best practices (DellaVigna et al., 2019), we describe the experiment and ask experts to forecast the key results in a way that is comparable to the experimental estimates. More precisely, we elicit their prediction of the effect of a 10 pp shock to the belief about the school share, separately for households *with* and *without* children. We then conduct the corresponding elicitations for beliefs about the recapture share.

We collected responses from experts in two ways. First, we posted the survey on the Social Science Prediction Platform from July 13, 2021, to December 31, 2021. Second, on November 2021, we emailed an invitation to the prediction survey directly to a list of 238 professors with publications related to our experiment. The final sample includes 56 experts' responses. Of these, 21.4% responded to the survey through the Social Science Prediction Platform, and the remaining 78.6% responded through our email invitation.[52] The final sample thus is comprised by 82.1% professors, 12.5% Ph.D. students, 3.6% post-docs, and 1.8% researchers. Most (78.6%) are from the field of economics; 66.1% report having done research on taxation and 25% on preferences for redistribution.

[52]Among the responses from the Social Science Prediction Platform, we exclude respondents who are not academics, who do not have a PhD, or who are not pursuing a PhD.

4.5 Perceptions about School Spending

4.5.1 Accuracy of Prior Beliefs

Transparency and accountability efforts have made information about property taxes publicly available. Each year, the Dallas Central Appraisal District (DCAD) provides homeowners in Dallas County with a Notice of Appraised Value, which contains a detailed break-down of the household's property taxes by tax jurisdiction, including the share of their property taxes that funds public schools.[53] But the ease of access to this information does not mean that everyone searches for it or uses it. Many other contexts show that individuals often misperceive easily accessible information, such as the official inflation rate (Cavallo et al., 2017) or recent trends in national home prices (Bottan and Perez-Truglia, 2020).

Figure 4.1(a) shows a histogram of the degree of misperceptions about the school share.[54] The x-axis corresponds to the difference between the actual school share (i.e., potential feedback) versus that perceived by respondents. For the sake of brevity, we use the term feedback to refer to potential feedback. A minority of subjects have accurate perceptions: more precisely, 32.6% of subjects guess the school share to be within ± 5 pp of the actual school share. Misperceptions are quite large on average: the mean absolute error is 16.57 pp. The large degree of misperceptions implies sufficient scope for the information provision experiment

[53]See Online Appendix for a sample of this notice, with the breakdown by tax jurisdiction shown on the second page. The county uses the prior year's jurisdictional tax rates to estimate taxes due in the Notice of Appraised Value because the tax rates for the current year are set later in the year. In practice, tax rate changes are uncommon, so approximation errors are typically negligible. In our study, we use the same definition of estimated taxes because these are the relevant object of study and they represent the subjects' best approximation at the time of deciding whether to protest.

[54]All results are based on the final survey sample, which excludes the outlier misperceptions (i.e., the bottom and top 5%). Including the extreme observations would increase the degree of misperceptions; for more details, see AOnline Appendix.

to shock beliefs. Another interesting feature of prior beliefs is that the misperceptions show a systematic bias: on average, subjects underestimate the school share by 13.08 pp, as indicated by the mean error. This systematic bias is quite noticeable in Figure 4.1(a), where more observations fall on the the right half of the histogram (corresponding to underestimation) than on the left half (corresponding to overestimation).[55]

4.5.2 Belief Updating

We find that taxpayers update their inaccurate beliefs when provided with accurate feedback. To model belief updating, we use a simple Bayesian model that has been shown to accurately represent belief formation in other information-provision experiments on a wide range of topics, such as inflation expectations (Cavallo et al., 2017), salary expectations (Cullen and Perez-Truglia, 2022), and home price expectations (Fuster et al., 2022).

We use the subscript i to index the subjects. We use the variable s_i^{prior} to represent subject i's belief right before the information-provision experiment. We use the variable s_i^{feed} to represent the value of the feedback that the subject can potentially receive in the experiment. We define the variable T_i^S as a binary variable that equals 1 if subject i is selected to receive the information and 0 if not. We define variable s_i^{post} as the posterior belief. Specifically, s_i^{post} represents the perceived school share after the taxpayer sees, or does not see, the feedback.

An individual shown feedback will form her posterior belief (s_i^{post}) as the average of the prior belief (s_i^{prior}) and the feedback (s_i^{feed}), weighted by a parameter

[55]It might be thought that households *with* children have more accurate perceptions about the school share than households *without* children. We show in Online Appendix that this conjecture is invalid since the distributions of perceptions are similar for these two groups.

α that captures the degree of learning. This parameter can range between 0 (individuals ignore the feedback) and 1 (individuals fully adjust to the feedback), and it is a function of the relative precision of the prior belief versus that of the feedback.[56] This Bayesian updating model can be summarized by the following linear relationship:

$$s_i^{post} - s_i^{prior} = \alpha \cdot \left(s_i^{feed} - s_i^{prior}\right) \tag{4.18}$$

Intuitively, Bayesian learning predicts that, when shown feedback, respondents who overestimate the school share would revise their beliefs downward, whereas respondents who underestimate the school share would revise their beliefs upward. Figure 4.1(b) estimates this Bayesian learning model using a binned scatterplot. The x-axis corresponds to the gaps in prior beliefs ($s_i^{feed} - s_i^{prior}$), and the y-axis corresponds to the belief updating ($s_i^{post} - s_i^{prior}$). Intuitively, the x-axis shows the maximum revision we would expect if the respondent were to fully react to the information, and the y-axis shows the actual revision. In the case of no updating, the observations should form a horizontal line; in the other extreme, under full updating, the observations should form a 45-degree line. The red circles in Figure 4.1(b) correspond with subjects who are shown feedback about the school share. Consistent with significant updating, there is a strong relationship between the updated beliefs and prior gaps: an additional percentage point (pp) in perception gap is associated with an actual revision that is 0.809 pp higher.

The gray squares in Figure 4.1(b) correspond with the subjects who do not receive information about the school share. In the absence of feedback, these

[56]These results assume normal distribution of priors and feedback and assume that the variance of the prior and the variance of the feedback are independent of the mean of the prior. For more details, see Hoff (2009).

subjects should not update their beliefs. However, in practice, individuals might revise their beliefs in the direction of the feedback for spurious reasons even when they receive no feedback. For instance, respondents may reassess their answers or correct typos when asked a question a second time, leading to an answer that is closer to the truth. The gray squares indicate a weak relationship between belief updating and prior gaps in the group that was not shown the feedback: an additional 1 pp in the prior gap is associated with an actual revision that is 0.052 pp higher. This effect is statistically significant (p-value ¡0.001) but economically very small. This result is consistent with other information-provision experiments that show evidence of spurious revisions (e.g., Fuster et al., 2022; Cullen and Perez-Truglia, 2022).

We can exploit the random assignment from the information-provision experiment to control for spurious learning:

$$s_i^{post} - s_i^{prior} = \tau + \alpha \cdot \left(s_i^{feed} - s_i^{prior} \right) \cdot T_i^S + \beta \cdot \left(s_i^{feed} - s_i^{prior} \right) + \epsilon_i \tag{4.19}$$

In this model, parameter α represents true learning arising from the information provision (not spurious learning), whereas parameter β captures spurious learning. Parameter α can be computed from the estimates in Figure 4.1(b). Specifically, the α parameter corresponds to the difference in the regression slopes between the subjects who are and are not shown the feedback. Since α captures the effect of the exogenous shocks induced by the information-provision experiment, it can be used as an excluded instrument in the econometric model explained in Section 4.5.3. The estimated α is large ($0.757 = 0.809 - 0.052$) and highly statistically significant (p-value¡0.001). This difference suggests that a 1 pp information shock induces a

0.757 pp effect in the subject's posterior belief. This shows that, although subjects did not update fully to the feedback, they were close to it. This finding of imperfect updating is consistent with other information-provision experiments and it is likely due to some subjects mistrusting the source of the feedback or simply not paying enough attention to the survey.

Online Appendix provides some additional results and robustness checks. For instance, this appendix shows that learning from the feedback is compartmentalized (i.e., subjects do not use the information about school share to update beliefs about the recapture share). This appendix also shows that the belief updating results are similar for households *with* and *without* children.

4.5.3 Econometric Model

Let P_i^{2021} denote the main outcome of interest: an indicator variable that equals 100 for individuals filing a protest in 2021 (i.e., post-treatment) and 0 otherwise. As discussed in the conceptual model in Section 4.3, and as noted in the RCT pre-registration, the effects of the school share information treatment on protests are expected to have different signs depending on whether the household has children enrolled in public schools. Let $C_i \in 0, 1$ be an indicator variable that equals 1 if the household has a child enrolled in a local public school and 0 otherwise. Hence, we can use the following econometric specification to estimate our parameters of interest:

$$P_i^{2021} = \beta_0 + \beta_C^S \cdot C_i \cdot s_i^{post} + \beta_{NC}^S \cdot (1 - C_i) \cdot s_i^{post} + \beta_1 \cdot C_i + \epsilon_i \tag{4.20}$$

The term ϵ_i represents the error. The two parameters of interest are β_C^S and

β^S_{NC}. The conceptual model from Section 4.3 predicts that $\beta^S_C < 0$ and $\beta^S_{NC} > 0$. Moreover, the difference between these two parameters, $\beta^S_C - \beta^S_{NC}$, is of special interest because it captures a key parameter: how much households *with* children benefit from school expenditures (see equation (4.10) for the case of school share and equation (4.17) for the case of recapture share). As posterior beliefs (s_i^{post}) are endogenous and thus could suffer from a host of omitted variable biases, we estimate equation (4.20) using 2SLS, exploiting the exogenous variation in posterior beliefs induced by the information-provision experiment. More precisely, we estimate the following model:

$$
\begin{aligned}
P_i^{2021} =& \beta_0 + \beta^S_C \cdot C_i \cdot s_i^{post} + \beta^S_{NC} \cdot (1 - C_i) \cdot s_i^{post} + \beta_1 \cdot C_i + \\
&+ \beta_2 \cdot C_i \cdot (s_i^{feed} - s_i^{prior}) + \beta_3 \cdot (1 - C_i) \cdot (s_i^{feed} - s_i^{prior}) + X_i \beta_X + \epsilon_i
\end{aligned}
\tag{4.21}
$$

The endogenous variables are $C_i \cdot s_i^{post}$ and $(1 - C_i) \cdot s_i^{post}$, for which we use the excluded instruments $C_i \cdot T_i^S \cdot (s_i^{feed} - s_i^{prior})$ and $(1 - C_i) \cdot T_i^S \cdot (s_i^{feed} - s_i^{prior})$.[57]

We can illustrate the intuition behind the model using a simple example. Consider a pair of subjects with children enrolled in public schools that share the same bias about the school share: both underestimate the actual school share by 20 pp. Suppose we randomly assign information about the true school share to one of them. We expect that, relative to the subject who does not get the information, the subject who receives the information adjusts their perceived school share upward. For the sake of argument, assume that the subject who does not receive the

[57]Note that equation (4.21) controls for the prior gaps in beliefs ($C_i \cdot (s_i^{feed} - s_i^{prior})$) and $(1 - C_i) \cdot (s_i^{feed} - s_i^{prior})$). The inclusion of these control variables ensure that the excluded instruments isolate the information shocks that are driven purely by the random assignment of the feedback (T_i^S).

information continues to underestimate the actual school share by 20 pp and that the subject who does receive the information reacts to it by underestimating the school share by just 10 pp. The information provision is thus equivalent to a +10 pp shock to the perceived school share. We can then check the behavior of this pair of households in the weeks after they receive the information. For example, the +10 pp shock to the perceived school share could translate to a lower probability of filing a protest. Assume that the +10 pp shock to the belief causes a 2 pp drop in the probability of protesting. Combining these two results, we can estimate that $\beta_C^S = -0.2$, that is, each 1 pp increase in the perceived school share lowers the probability of protesting by 0.2 pp.[58]

The term X_i in equation (4.21) corresponds to a set of additional control variables. In principle, the 2SLS model leverages the experimental variation, so control variables are not needed for causal identification. However, the inclusion of additional control variables can be helpful, for instance, to reduce the variance of the error term and thus improve precision (McKenzie, 2012). The vector of control variables includes basic pre-treatment information, such as the household's prior history of tax appeals.[59]

---

[58]Typically in 2SLS models, if treatment effects are heterogeneous, the estimates identify the local average treatment effects of beliefs (Imbens and Angrist, 1994). More precisely, in our study, our estimates would give a higher weight to subjects whose beliefs are more affected by the information-provision experiment. By construction, this weight will be higher for subjects with larger prior misperceptions and, conditional on the misperceptions, those who react more strongly to feedback.

[59]The full set of additional control variables includes the log of total market value in 2021, the growth in total market value between 2021 and 2020, an indicator for positive growth, an indicator of whether the property value was re-evaluated in 2021, the 2021 estimated property taxes (in logs), a dummy for homestead exemption in 2021, an indicator for homestead binding in 2021, the household's effective tax rate, a dummy variable for multiple owners, a dummy variable for condos, the total living area, the number of bedrooms, the number of full baths, the building age, a set of dummies for school districts, the survey start date, and indicator variables for whether the household protested in each pre-treatment period since 2016 (one set for direct protests and another set for protests through agents).

Following the regression specification we use to study the effects of the school share (equation (4.20)), it is straightforward to define the regression specification to study the effects of recapture. Indeed, as these two information treatments are cross-randomized for the same sample, we estimate all effects simultaneously in a single 2SLS regression. See Section 4.6.3 for a discussion of the recapture share estimates.

4.5.4 2SLS Estimates

The 2SLS estimates for school share are presented in the top half of Table 4.2.[60] In column (1), the dependent variable is the main outcome of interest: an indicator variable that equals 100 if the subject protests directly in 2021 and 0 otherwise. According to the hypothesis of reciprocal motivation, an increase in the perceived school share should decrease the probability of protesting for households *with* children (because they find out that they benefit from government services more than they thought), whereas this information should have the opposite effect on households *without* children. The results are consistent with this hypothesis. The coefficient for households *with* children is negative (-0.367) and statistically significant (p-value=0.096). The coefficient for households *without* children is positive (0.277) and statistically significant (p-value=0.032). Most importantly, the difference between the two coefficients (-0.367 and 0.277) is statistically significant (p-value=0.012).

As a thought experiment, consider what would happen if the perceived school share increases by 10 pp – for reference, this is roughly how much the average

[60]We present the 2SLS estimates directly because they can be interpreted more easily. Nevertheless, due to the strong first stage (i.e., high belief updating), the 2SLS estimates are similar (in terms of magnitude and statistical significance) to the reduced-form estimates. For more details, see Online Appendix.

belief changed due to the information.[61] The estimates from column (1) of Table 4.2 indicate that this change would cause a drop of 3.67 pp ($= 0.367 \cdot 10$) in the probability of filing a protest for households *with* children and an increase of 2.77 pp ($= 0.277 \cdot 10$) in the probability of protesting for households *without* children. These effects would be roughly equivalent to 11% and 10% of the baseline protest rates (33.86 pp and 28.83 pp, reported at the bottom rows of Table 4.2).

Column (2) of Table 4.2 is identical to column (1), except that it uses a different dependent variable: an indicator variable that equals 100 if, at the end of the survey, the subject responds "very likely" to the question on the likelihood to protest in 2021 and 0 otherwise. This outcome measures the intention to protest and allows us to measure if the effects of the information lead to an intention to protest immediately after the information is provided. For reference, at the time of the survey, 45.4% report that they are very likely to protest (this corresponds to the baseline rate, combining subjects *with* and *without* children who do not receive any feedback), which is higher than the actual protest rate in the administrative data, 30.06%. For instance, a respondent may report a high protest likelihood in the survey, but then not protest due to filing frictions (Nathan et al., 2020).[62] It is important to note that the stated intention to protest correlates significantly with whether the individual actually files a protest, but that correlation (0.410) is far from perfect (the correlation coefficient is 0.410 for the no-feedback group,

[61] More precisely, the average posterior belief increased by 10.69 pp due to the feedback (from 38.78 pp in the control group to 49.47 pp in the treatment group).

[62] Spurious reactions (Cavallo et al., 2017) or other "salience" effects might explain part of the difference between the stated intention to protest in the survey and the actual protest. We do not think potential "salience" effects are an important concern in our setting. First, households protests days or weeks after answering the survey. Second, potential "salience" effects would presumably equally affect households *with* and *without* children and our identification strategy arises from comparing these two types of households.

p-value¡0.001).[63] Due to this imperfect correlation, the effects on the intention to protest at the time of answering the survey should not expected to be "mechanically" the same as the effects on actual protests.

The results from column (2) of Table 4.2 are consistent with the results from column (1). In column (2), the coefficient for households with children is negative (-0.408) and similar in magnitude to the corresponding coefficient from column (1) and statistically significant (p-value=0.080). The coefficient for households *without* children is positive (0.269), on the same order of magnitude as the coefficient from column (1), and statistically significant (p-value=0.062). Again, most importantly, the difference between the coefficients for households with children versus those without children (-0.408 and 0.269) is statistically significant (p-value=0.014).

4.5.5 Robustness Checks

To probe the robustness of the school share results, columns (3) and (4) of Table 4.2 provide two falsification tests. For the first and most important falsification test, we exploit the timing of the information intervention in an event-study fashion. In column (3) of Table 4.2, we estimate the same baseline regression from column (1), except that we use as the dependent variable the protest decision in a pre-treatment year (2020), rather than in the post-treatment year (2021). Intuitively, since the information was provided in 2021, it could not possibly have an effect on the decision to protest as of a year earlier (2020). We thus expect the coefficients from this falsification exercise to be close to zero and statistically insignificant. The results from column (4) confirm our expectations: the coefficients from column (4)

[63] Among respondents who report being very likely to protest, 56.8% end up protesting directly or through an agent. On the other hand, among the individuals who do not report being very likely to protest, 16.8% end up protesting.

are close to zero (0.110 and -0.065, for households *with* and *without* children, respectively), precisely estimated, and statistically insignificant (p-values of 0.545 and 0.504); most importantly, the difference between households *with* children and *without* children is also close to zero (0.175) and statistically insignificant (p-value=0.398). Indeed, we can extend this same falsification test to other pre-treatment years for which we have readily available data. For ease of exposition, the results are presented in graphical form in Figure 4.2(a). The x-axis denotes the year of the dependent variable (i.e., whether the owner protests directly in years 2016 through 2021). This figure focuses on the main result, corresponding to the difference in coefficients between households *with* children versus *without* children. Thus, the 2020 coefficient from Figure 4.2(a), which takes the value 0.175, is by construction equal to the corresponding coefficient from column (3) of Table 4.2. As expected, for each pre-treatment year (2016–2020) the coefficients are close to zero and statistically insignificant; by contrast, the coefficient is negative and statistically significant in the post-treatment year (2021).

The second falsification test uses a dependent variable that indicates whether the household protests through an agent. As explained in Section 4.4.4, it is highly unlikely that the information provided in our survey would affect protests through an agent. The results are reported in column (3), which estimates the same regression from column (1) but using protests conducted by agents as the dependent variable. As expected, the coefficients from column (3) are close to zero (-0.028 and -0.033) for both households *with* and *without* children, precisely estimated, and statistically insignificant (p-values of 0.816 and 0.518). The difference between the coefficients for households *with* and *without* children is close to zero (0.006), precisely estimated, and statistically insignificant (p-value=0.966).

One usual concern with 2SLS estimation concerns weak instruments (Stock et al., 2002). Given the strong belief updating documented in Section 4.5.2, weak instruments should not be a concern in our setting. Nevertheless, for a more rigorous assessment, Table 4.2 reports the Cragg-Donald F-statistic, which is commonly used to diagnose weak instruments. The value of this statistic in each regression is well above the rule of thumb of F ¿ 10 proposed by Stock et al. (2002): it equals 30.10, 30.22, 30.10, and 30.02, respectively, in columns (1)–(4) of Table 4.2.[64]

The 2SLS model used for the results in equation (4.21) assumes a linear relationship between school share and the probability of protesting. This is a natural starting point due to its simplicity and because it is common practice in the literature on information-provision experiments. To probe that assumption, Figure 4.2(b) presents a binned scatterplot representation of the reduced-form effects of the information-provision experiment (i.e., not accounting for how the information provision affects prior beliefs). The x-axis corresponds to the interaction between the potential information disclosure and the prior gap (i.e., the excluded instrument). The y-axis corresponds to the probability of protesting in 2021. This binned scatterplot includes all the same control variables used in the 2SLS model. Figure 4.2(b) seeks to assess whether the relationship between the interaction term on the horizontal axis and the protest probability on the vertical axis is linear, and the figure shows that a linear fit is a reasonable functional form assumption for this context. Additionally, this figure shows that the previously discussed regression results are not driven by outliers. In a similar spirit, Online Appendix shows that the results are robust to an alternative approach that does not use 2SLS, which is simpler although it makes a less efficient use of the data.[65]

[64]For the detailed coefficients from the first-stage of the regression, see Online Appendix.

[65]The appendix also shows the difference in sizes between the OLS regression not using the prior to posterior belief update and 2SLS regression using the belief update.

Table 4.3 presents additional robustness checks. Columns (1) and (2) of Table 4.3 reproduce the baseline specification given by columns (1) and (2) of Table 4.2 for reference. Columns (3) through (10) of Table 4.3 present the results under alternative specifications. The specification from columns (3) and (4) is identical to the specification from columns (1) and (2), except that we include some additional control variables: the respondent's age, a dummy for individuals that self-identify as White, a dummy for gender, a dummy for college degree, and a dummy for political party (which equals 1 for individuals who self-identify as Democrat). Note that these variables are measured at the end of the survey, and some respondents did not finish the full survey. Thus, the inclusion of these additional controls reduces the number of observations, which is the main reason why we exclude these variables from the baseline controls. The results from columns (3) and (4) are similar to the baseline results from columns (1) and (2). If anything, the inclusion of the additional controls yields effects that are are slightly stronger (-0.714 vs. -0.644 and -0.744 vs. -0.678).

In columns (5) through (8) of Table 4.3, we try alternative definitions of outliers in prior misperceptions. The baseline specification is already conservative in that it excludes the extreme top and bottom 5% of the distributions. In columns (5) and (6), we use a less stringent definition of outliers based on the upper and bottom 2.5% instead of 5%. The results from columns (5) and (6) are similar to those from the baseline specification of columns (1) and (2), although the coefficients are slightly smaller in magnitude. In columns (7) and (8), we consider an even more lax definition of outliers, excluding only the upper and bottom 1% of misperceptions. The coefficients from columns (7) and (8) remain consistent with those from the baseline specification of columns (1) and (2), although again the magnitudes are somewhat smaller. These results are consistent with the arguments in Section

4.4.3 that we should be cautious when including extreme misperceptions because they probably reflect a lack of attention or mistakes, rather than legitimate misperceptions. To explore this further, columns (9) and (10) are identical to the baseline specification from columns (1) and (2), except they exclude respondents who do not pass the attention check included at the end of the survey. Consistent with the attention argument, when we focus on subjects who pass the attention check, the coefficients increase somewhat.

4.5.6 Comparison to Expert Predictions

Next, we compare our experimental results to expert predictions, as shown in Figure 4.3. Panel (a) presents the predictions of experts for households *with* children, and panel (b) presents the predictions for households *without* children. The histograms correspond with the distribution of expert predictions estimating the effect of a 1 pp increase in the school share.[66] The solid vertical red line in each panel represents the corresponding estimate from the baseline 2SLS model (column (1) of Table 4.2), and the red shading denotes the corresponding confidence intervals.

Figure 4.3 shows that our experimental findings are not obvious to the sample of experts. Our experimental results are consistent with experts who predicted that the school share belief would have a negative effect on the protest rate for households *with* children (panel (a)) and a positive effect for households *without* children (panel (b)). They also are consistent with the mean of the experimental estimates in these two panels. However, the forecasts of most experts are incon-

[66]To make the elicitation easier, in the prediction survey, we ask subjects to predict the effects of a 10 pp increase in the school share. In Figure 4.3, we divide those predictions by 10 to obtain the effect per 1 pp, so that it can be compared directly to the 2SLS estimates.

sistent with the experimental results: the majority of the forecasts predict either zero effect or an effect of the opposite sign compared to the experimental findings. Only a few expert predictions are close to the experimental estimates, even if we account for the sampling variation in the experimental estimates. More precisely, for households *with* children, only 41.1% of predictions are within the 90% confidence interval of the experimental estimate. For households *without* children, just 17.9% of predictions are within the 90% confidence interval of the experimental estimate. That the majority of experts' predictions do not coincide with the experimental findings may not be surprising since their predictions are consistent with the general takeaway from the extant literature on how nudges affect tax compliance which suggests that deterrence nudges are effective whereas tax morale messages are less effective or have no effects whatsoever (see Bergolo et al. (2021)).

At the end of the survey, we ask the experts to express how confident they feel about their forecasts. One notable finding is that experts do not feel confident about their predictions: on a scale of 1 to 5, where 1 is "not confident at all" and 5 is "extremely confident", the average confidence is 2.07.[67] In any case, we find that the comparison between the forecasts and experimental estimates is similar if we weight the forecasts by the confidence of the experts (results reported in Online Appendix).

[67] More precisely, 25.0% of experts selected "not confident at all," 51.8% selected "slightly confident," 19.6% selected "somewhat confident," 3.57% selected "very confident", and 0% selected "extremely confident."

4.5.7 Non-Experimental Evidence

We complement this experimental evidence with some non-experimental evidence by including a survey question asking individuals to choose between hypothetical tax policies, in the spirit of Weinzierl (2014) and Saez and Stantcheva (2016). More specifically, we include a question about public school taxes. We present the respondent with a hypothetical situation in which two households (A and B) own homes worth \$200,000 each. Both households are identical except that household A has two children enrolled in the public school district and household B has no children enrolled in the public school district. The respondent has to levy a total tax of \$8,000, which can be spread across the two households in any way (e.g., assign all the burden to household A, all the burden to household B, or anything in the middle). According to the hypothesis of reciprocal motivation, respondents will want the household *with* children to pay more in taxes than the household *without* children, because the former benefits more from that government service. We find that most (58.8%) of the respondents behave according to the reciprocal mechanism, that is, they assign a higher tax burden to the household *with* children even though both homes are worth the same.[68] This evidence suggests that the mechanism of reciprocal motivation resonates with most taxpayers.

A feature of property tax policy in the state of Texas is suggestive of reciprocal motives. Texas homeowners who are older than 65, most of whom do not have school-aged children, qualify for an exemption that limits their school taxes to the amount paid in the year that the owner turned 65, regardless of future increases in the home's proposed value (Texas Comptroller, 2021).[69] This exemption policy

[68]For detailed results, see Online Appendix.

[69]The tax amount paid can increase if property improvements are made beyond maintenance and repairs. Homeowners also must apply to receive this benefit.

for households unlikely to have children is consistent with benefit-based reasoning.

4.6 Perceptions about Recapture

4.6.1 Accuracy of Prior Beliefs

Unlike the information on the school share, the information on recapture is not readily available in the Notice of Appraised Value from the DCAD. However, households may be informed about the recapture system through its media coverage. Also, it is probably widely knwon that the recapture system redistributes from more to less advantages districts. As a result, if a homeowner knows whether he or she lives in a more or less advantaged district, that information on its own may be enough to form a decent guess about the recapture share.

Figure 4.4(a) shows a histogram of the degree of misperceptions about the recapture share. The x-axis corresponds to the difference between the actual recapture share versus that perceived by respondents. A minority of subjects have accurate perceptions: around 20% of subjects guess the recapture share to be within ± 5 pp of the actual share. Misperceptions are significant in magnitude: the mean absolute error is 11.36 pp. However, the mean absolute error for the recapture share (11.36 pp) is substantially less pronounced than that of the school share (16.57). The fact that misperceptions for the recapture share are smaller than those for the school share implies that there is less scope for the information provision experiment to update beliefs and thus less statistical power for the 2SLS estimates.

Unlike misperceptions about the school share, misperceptions about the recapture share have no systematic bias: on average, subjects overestimate the recapture share by just 0.28 pp. This can be seen directly from Figure 4.4(a), which shows that households are roughly equaly likely to be in the left half of the histogram (corresponding to overestimation) as in the right half (corresponding to underestimation).

4.6.2 Belief Updating

Next, we summarize how subjects update their beliefs in reaction to the information provision about the recapture share. Figure 4.4(b) shows the results as a binned scatterplot. The x-axis corresponds to the gaps in prior beliefs, and the y-axis denotes the belief updating. The x-axis in Figure 4.4(b) shows the maximum revision we would expect if the respondents were to fully react to the provided information, and the y-axis shows the revision observed in practice. The red circles from Figure 4.4(b) correspond to subjects who are shown the feedback about the recapture share. Consistent with significant learning, there is a strong relationship between the belief revisions and prior gaps: an additional percentage point (pp) in perception gap is associated with a revision that is 0.632 pp higher. The gray squares from Figure 4.4(b) correspond with the subjects who do not receive information about the school share. In turn, the gray squares indicate a statistically significant (p-value¡0.001) but economically small (0.099) degree of spurious revision. Most importantly, the degree of true learning corresponds to the difference in slopes between subjects who are shown the feedback and subjects who are not shown the feedback. This difference is large ($0.533 = 0.632 - 0.099$) and highly statistically significant (p-value¡0.001). This difference suggests that a 1 pp information shock induces a 0.533 effect in posterior beliefs. Though large, this rate of information pass-through (0.533) is quite smaller than the corresponding rate for the school share (0.757).

Many reasons help explain the weakly updated beliefs about recapture. For example, respondents may feel more confident in their prior beliefs about recapture, or they may trust the feedback on recapture less. Indeed, the recapture estimates that we use for the feedback are based on a number of assumptions, so subjects may naturally find the recapture feedback less persuasive. Last, subjects may pay less attention to the recapture feedback due to survey fatigue, as this information appears later in the survey. The most important implication of the weaker belief updating for recapture share (relative to school share) is that it will result in less exogenous variation in posterior beliefs and thus less precisely estimated 2SLS coefficients.

4.6.3 2SLS Estimates

Let r_i^{post} be the posterior belief about the funds recaptured from individual i's own school district, in percentage points. Positive values indicate that individual i's district is a net contributor to the recapture system; in other words, $r_i^{post} = 40$ means that 40% of school taxes from household i's district are redistributed to disadvantaged school districts. Negative values, on the contrary, represent situations where individual i's school district benefits from recapture: $r_i^{post} = -30$ means that the school district receives an additional 30% over the amount of its own school taxes from taxpayers in other school districts.[70] We use the following econometric specification:

$$P_i^{2021} = \beta_0 + \beta_C^R \cdot C_i \cdot r_i^{post} + \beta_{NC}^R \cdot (1 - C_i) \cdot r_i^{post} + \beta_1 \cdot C_i + \epsilon_i \tag{4.22}$$

[70] The negative values can be lower than -100 because an ISD can receive more than 100% of the amount of its own school taxes in redistributed tax.

The two parameters of interest are β_C^R and β_{NC}^R for households *with* and *without* children, respectively. The framework of reciprocal motivation from Section 4.3 predicts that $\beta_C^R > 0$ and $\beta_{NC}^S = 0$. Again, the difference between these two parameters, $\beta_C^R - \beta_{NC}^R$, captures the reciprocal motivation behind public schools. As in the estimation of the change in perceptions about the school share, we estimate equation (4.22) using 2SLS to exploit the variation in r_i^{post} induced exogenously by the information provision experiment. As mentioned in Section 4.5.3, we estimate the effects of school share and recapture share jointly in the same 2SLS regression.

The 2SLS estimates for the recapture share are presented at the bottom panel of Table 4.2. In column (1) of Table 4.2, the dependent variable indicates if the subject protests directly in 2021. The causal effects of the beliefs about the recapture share are very imprecisely estimated, so the results for this treatment arm are largely inconclusive. Consistent with the hypothesis of reciprocal motivation, the belief about recapture share does not have significant effects on the decision to file a tax appeal among households *without* children: the coefficient is positive (0.498) statistically insignificant (p-value=0.101). This finding must be taken with a grain of salt, however: since the coefficient is imprecisely estimated, we cannot rule out large effects, positive or negative.

To illustrate how imprecisely estimate this coefficient is, note that the standard error for recapture share is 135% larger than the corresponding standard error for school share (0.303 vs. 0.129). In other words, the effects for recapture share should be more than twice as high as the effects of school share to have enough power to detect statistically significant effects. The less precise estimation for the coefficients for recapture share occurs for two reasons, both of which are difficult to anticipate ex-ante in the experimental design. First, as explained in Section 4.6.1, the misperceptions about recapture share were smaller (mean absolute difference of 11.36 pp) than those about school share (mean absolute difference of 16.57 pp). Second, as documented in Section 4.6.2, conditional on a level of misperceptions, subjects updated their beliefs more strongly in response to the feedback about school share than in response to the feedback about recapture share.

We do not find evidence of significant positive effects for households *with* children. The coefficient for households *with* children is positive (0.076) but statistically insignificant (p-value=0.875). Again, this coefficient is so imprecisely estimated that it does not really constitute evidence against the hypothesis of reciprocal motivation, because we cannot rule out very large positive effects. More precisely, the 95% confidence interval cannot rule out a positive coefficient of up to 1.02, which is several times the magnitude of the effects documented for the first treatment arm. Likewise, the difference between the coefficients for households with versus without children is statistically insignificant (p-value=0.454), but it is very imprecisely estimated so we cannot rule out large differences.

The coefficients from column (2) of Table 4.2 show that the results for recapture share are similar if we look at the intention to protest instead of the actual protest decision.[71]

The coefficients from column (3) and (4) of Table 4.2 show that, as expected, the results for recapture share do not show any effects on the falsification outcomes.[72] And Table 4.3 shows that the results for recapture share are similar

[71] More precisely, column (2) of Table 4.2 is identical to column (1), except that the dependent variable is the intention to protest instead of whether the household actually files a protest. As in column (1), the estimates from column (2) are all statistically insignificant. The only coefficient from column (1) that is borderline significant, for households *without* children (p-value=0.101), is not even close to being statistically significant in column (2), and furthermore it has the opposite sign.

[72] More precisely, column (3) of Table 4.2 uses the protest decision in a pre-treatment year (2020) as a dependent variable. As expected, the coefficients from column (4) (0.164 and -0.039 for households *with* and *without* children, respectively) are both statistically insignificant

under alternative specifications.[73] Some additional results are reported in the Appendix. Using binned scatterplots, we show that the results for the recapture share are not driven by non-linearities or outliers (results presented in Online Appendix). We also compare our 2SLS estimates to the expert predictions.

4.7 Conclusions

Compared to abundant causal evidence on the importance of institutions for tax compliance, little causal evidence shows that tax morale is important. In this paper, we attempt to fill this gap by providing evidence from a natural field experiment. Our novel research design studies tax morale by linking data from a survey experiment to administrative tax records at the individual level. Our subjects are homeowners who pay property taxes and have the opportunity to appeal their property tax assessment. We find that even though accurate information is publicly available and easily accessible, households have large misperceptions about how tax dollars are spent. Through an information-provision experiment, we corrected misperceptions about where their tax dollars go. The effects of the information provision experiment are consistent with our hypothesis of reciprocal motivation. After learning that a higher share of property taxes funds public

(p-values of 0.694 and 0.867); the difference between the two (0.203) is also statistically insignificant (p-value=0.664). We find similar results if we expand this falsification test to other pre-treatment years (results presented in Online Appendix). For the second falsification test, column (4) of Table 4.2 uses the dependent variable that indicates whether the household ever protested through an agent. As expected, the coefficients are statistically insignificant (p-values of 0.249 and 0.359 for households *with* and *without* children, respectively) and the difference between the two coefficients (-0.207) is also statistically insignificant (p-value=0.486).

[73]More precisely, columns (1) and (2) of Table 4.3 reproduce the baseline specification given by columns (1) and (2) of Table 4.2 for reference, and columns (3) through (10) of Table 4.3 present the results under alternative specifications: including additional control variables, using alternative definitions of outliers, etc. The baseline results from columns (1) and (2) are consistent with the results from all the alternative specifications from columns (3) through (10).

schools, households *with* children enrolled in public schools become less likely to appeal their property taxes and households *without* children become more likely to appeal their property taxes.

In this paper, we limited our exploration to the role of specific beliefs like the share of school taxes. However, we seek to make a more general methodological contribution: our research design can be used to study other mechanisms under the umbrella of tax morale. For instance, this approach could be used to assess the willingness to pay taxes in response to changes in the perceived quality of government spending or perceived corruption.

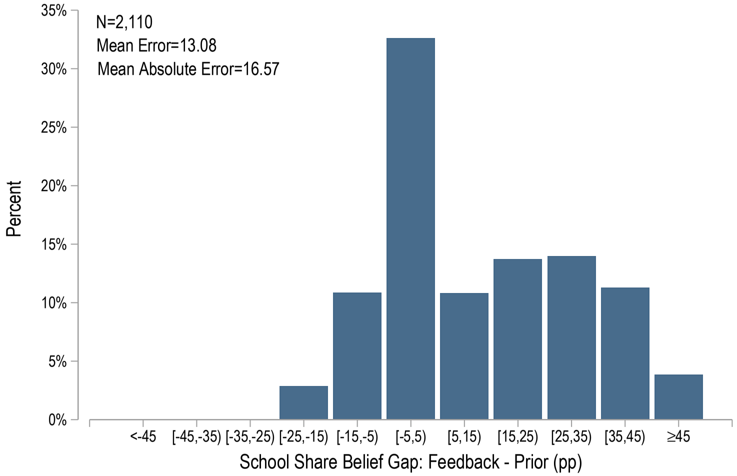
Our results stress the challenges of public communication policies. First, we document evidence of large misperceptions about government spending, even when such information is publicly available. For governments interested in educating their citizens on how tax dollars are spent, they should do more than post information on a website. Additionally, governments may want to simplify the connection between the taxes they collect and the government services those taxes support. Indeed, local governments tend to do this well in that they typically break down property taxes into a school tax, a hospital tax, and so on. Even in the simple context of property taxes, however, we still find that taxpayers have large misperceptions about how tax dollars are spent. In the case of state and federal governments, for which tax dollars follow a complicated path on their way to becoming public services, there is likely much room for improvement if the goal is to educate citizens.

Our experimental intervention was designed to disentangle causal mechanisms, not to increase average tax compliance. Nevertheless, our findings provide some hints for policy-makers looking to boost tax compliance. Our results underscore the challenges and limitations of transparency policies and information campaigns.

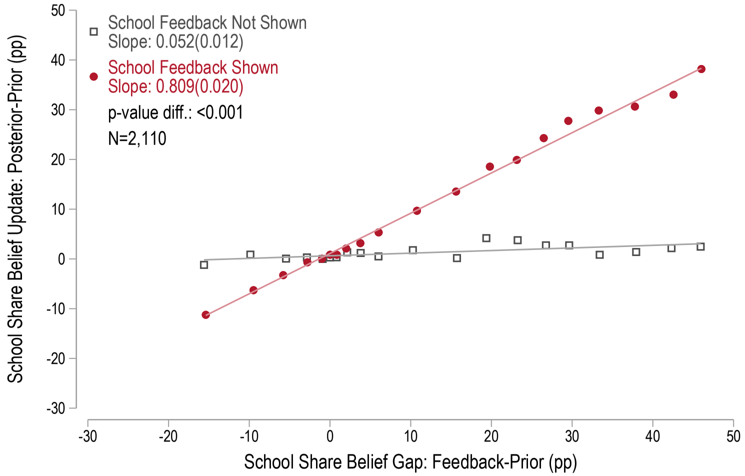
For example, a message highlighting a government service (e.g., public schools) can boost tax compliance among individuals who benefit most from that service (e.g., households *with* children), but it can reduce compliance from taxpayers who do not benefit from that service (e.g., households *without* children). As a result, these effects may cancel each other out, resulting in a null average effect on tax compliance. In some cases, this approach may even backfire. Our findings suggest that governments may be able to use reciprocal motives to boost average tax compliance, but only if they are willing to target the information (e.g., informing households *with* children about public school spending). Also, governments could try to persuade taxpayers that their tax dollars are spent efficiently or that their tax payments are not captured by corrupt politicians or wasted by bureaucrats. To the extent that these messages raise the average taxpayers' perception that their tax dollars are well spent, they also may increase the average tax compliance.

Figure 4.1: Perceptions about the Share of Property Taxes Going to Public Schools

(a) Gap in Prior Beliefs



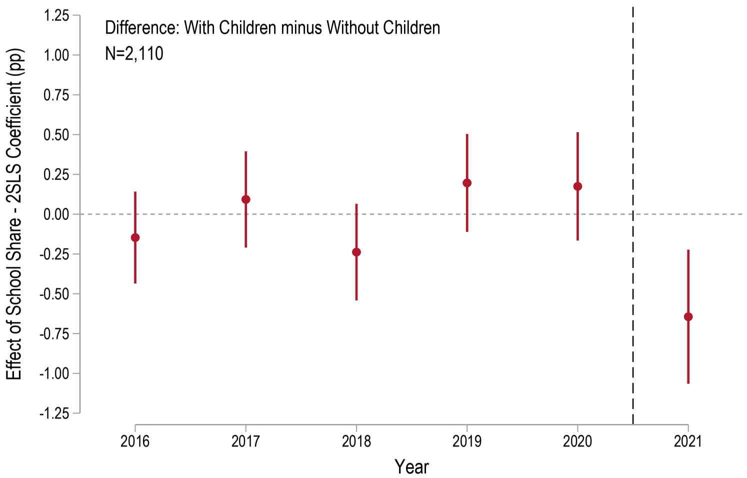
(b) Belief Updating



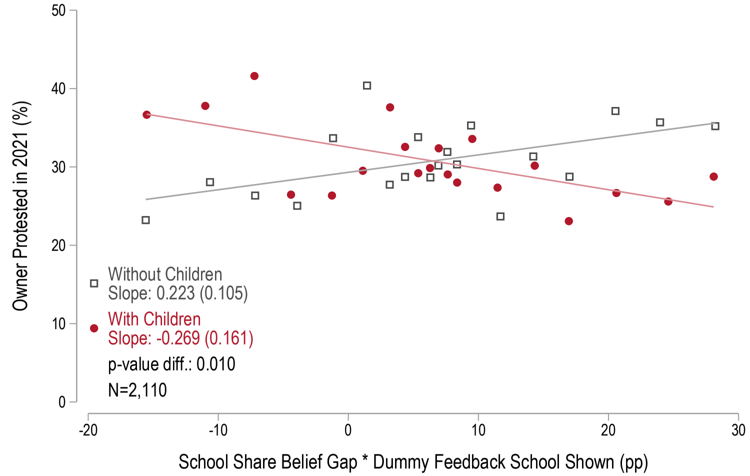
Notes: Panel (a) shows the gap in prior beliefs about the school share. The x-axis reports the difference between the actual school share and respondents' prior beliefs about the school share in 10 pp-width bins. The y-axis reports the percentage of survey respondents in each bin. The upper left corner reports the total number of observations, the mean error, and the mean absolute error. Panel (b) shows how respondents update their beliefs using a binned scatterplot (using 20 equally-sized bins). The x-axis reports the difference between the actual school share and respondents' prior beliefs about the school share. The y-axis reports the difference between posterior and prior beliefs (i.e., belief updating). Red circles (gray squares) represent the average update within each bin for the group of homeowners that were selected (were not selected) into the school share treatment. Each line corresponds to the fitted values from separate OLS regressions where the dependent variable is the actual update and the independent variable is the school share belief gap. The coefficients associated with the gap variable are reported in the upper left corner, as well as their robust standard errors (in parentheses), the p-value of the difference in the slopes, and the number of observations included in the analysis.

Figure 4.2: The Effects of School Share Perceptions on Protests: Additional Robustness Checks

(a) Event-Study Analysis



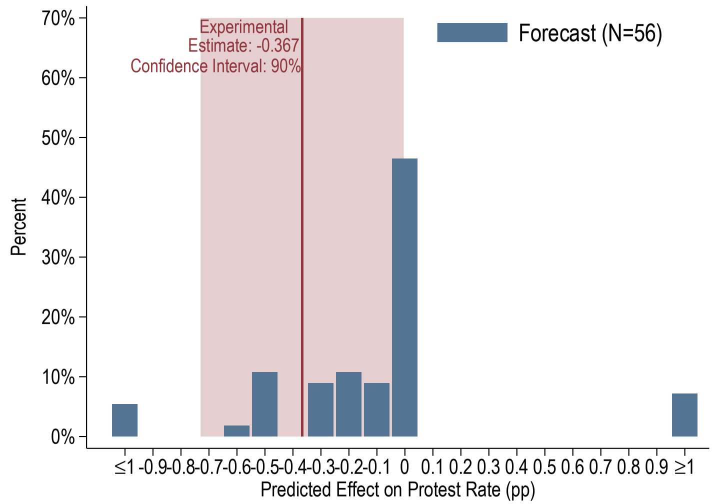
(b) Binned Scatterplot (Reduced Form)



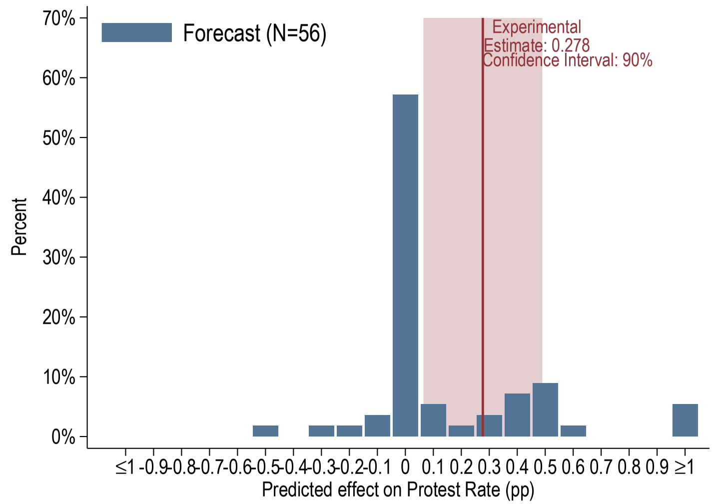
Notes: Panel (a) reports an event-study analysis of the differential effect of school share belief on the protest probability for households *with* children versus *without* children. The estimates plotted in this figure correspond with the 2SLS point estimate based on equation (4.21), with 90% confidence intervals based on robust standard errors. The coefficient plotted for 2021 is the coefficient reported in the "difference" row of panel (a), column (1) of Table 4.2. The remaining coefficients come from similar regressions but using the outcomes in pre-treatment years as falsification tests and restricting the pre-treatment controls to the corresponding years. The vertical dashed line separates the post-treatment year (2021) from the pre-treatment years (2016-2020). Panel (b) depicts a scatterplot representation of the reduced-form effect for households *with* and *without* children separately, using red circles and gray squares respectively and 20 equally-sized bins. The x-axis corresponds to the interaction between the prior school share belief gap (defined as the difference between the actual school share and the prior belief about the school share) and a dummy variable that indicates if the homeowner was selected into the school share treatment group. The y-axis corresponds to the probability of a direct protest in 2021. Each line corresponds to a separate OLS binned scatterplot regression, including the same control variables used in the 2SLS specification. The coefficients reported in the lower left corner and their (robust) standard errors are based on a unique regression that interacts the key variables with a dummy for having children at school (for the results in table form, see Online Appendix). In addition we report the p-value of the difference in the effect for the two groups and the number of observations used in the estimation.

Figure 4.3: The Effects of School Share on Protests: Comparison to Expert Predictions

(a) With Children in Public School



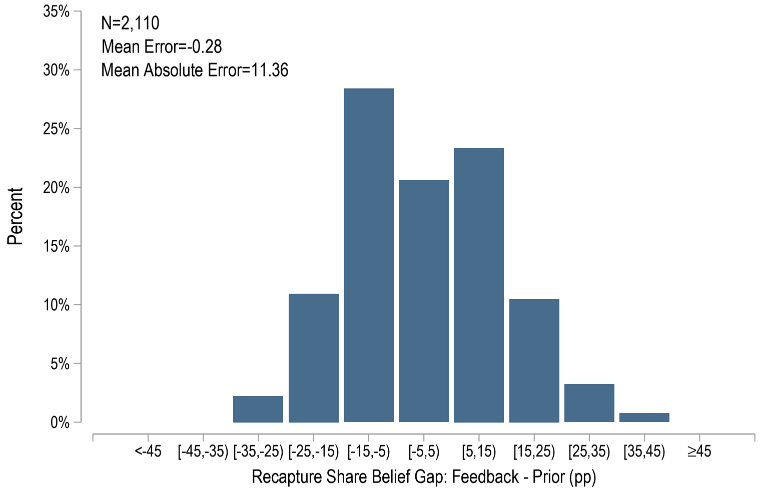
(b) Without Children in Public School



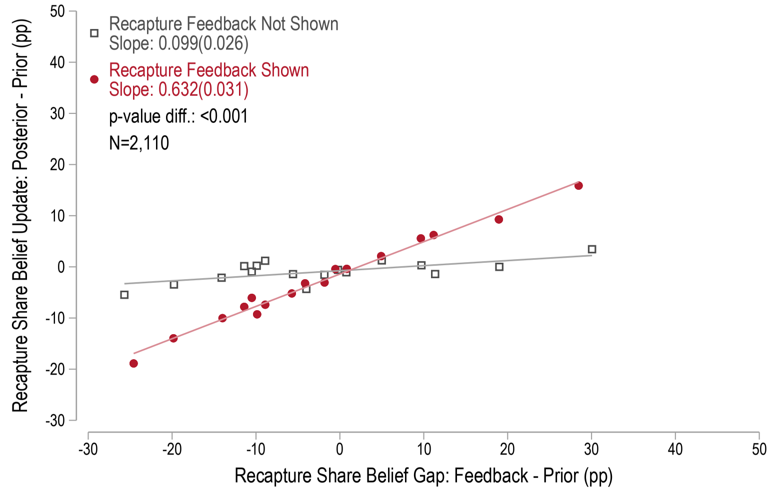
Notes: This figure shows the distribution of expert predictions about the effects of a 1 pp increase in school share beliefs on the probability that a homeowner files a protest directly for households *with* children enrolled in the public school district (panel (a)) and households *without* children enrolled in the public school district (panel (b)), based on the data collected in the forecast survey. To make the elicitation easier, in the prediction survey we asked subjects to predict the effects of a 10 pp increase in beliefs about school share. For this figure, we divide those predictions by 10 and we obtained the effect per 1 pp so these coefficients can be compared directly to the 2SLS estimates. In both panels, we pooled responses that were greater than 1 in absolute value into the corresponding extreme bins. The vertical red solid line corresponds to the experimental estimate based on the 2SLS specification reported in Table 4.2. The shaded area (in pink) corresponds to the 90% confidence interval. The full questionnaire for the prediction survey can be found in Online Appendix.

Figure 4.4: Perceptions about the Share of School Taxes Affected by Recapture

(a) Gap in Prior Beliefs



(b) Belief Updating



Notes: Panel (a) shows the gap in prior beliefs about the recapture share. The x-axis reports the difference between the actual recapture share and respondents' prior beliefs about the recapture share in 10 pp width bins. The y-axis reports the percentage of survey respondents in each bin. The upper left corner reports the total number of observations, the mean error, and the mean absolute error. Panel (b) shows how respondents update their beliefs using a binned scatterplot (with 20 equally-sized bins). The x-axis reports the difference between the actual recapture share and respondents' prior beliefs about the recapture share. The y-axis reports the difference between posterior and prior beliefs (i.e., belief updating). Red circles (gray squares) represent the average update within each bin for the group of homeowners that were selected (were not selected) into the recapture share treatment. Each line corresponds to the fitted values from separate OLS regressions where the dependent variable is the actual update and the independent variable is the recapture share belief gap. The coefficients associated with the gap variable are reported in the upper left corner, as well as their robust standard errors (in parentheses), the p-value of the difference in the slopes, and the number of observations included in the analysis.

Table 4.1: Balance of Households' Characteristics across Treatment Groups

		Treatment Arm				
	All (1)	No Feedback (2)	Recapture Feedback (3)	School Feedback (4)	Both Feedback (5)	p-value test (6)
a. Admin. Records Variables:						
2021 Home Value ($1,000)	349.988	365.355	330.631	365.198	340.088	0.163
	(6.774)	(14.907)	(10.302)	(16.461)	(12.037)	
2021 Property Tax Amount ($1,000s)	7.738	8.018	7.448	7.960	7.546	0.292
	(0.129)	(0.296)	(0.218)	(0.287)	(0.228)	
School Share (%)	50.726	50.603	50.566	50.701	51.029	0.140
	(0.079)	(0.155)	(0.160)	(0.155)	(0.158)	
Recapture Share (%)	1.622	1.852	1.054	2.505	1.130	0.351
	(0.325)	(0.678)	(0.633)	(0.672)	(0.622)	
2020 Owner Protested (%)	18.057	23.121	14.815	19.883	14.684	0.000
	(0.838)	(1.852)	(1.530)	(1.764)	(1.527)	
2020 Agent Protested (%)	1.659	1.156	2.407	1.754	1.301	0.375
	(0.278)	(0.470)	(0.660)	(0.580)	(0.489)	
2019 Owner Protested (%)	13.365	15.029	10.926	14.035	13.569	0.238
	(0.741)	(1.570)	(1.344)	(1.535)	(1.478)	
2018 Owner Protested (%)	13.460	13.680	12.407	14.815	13.011	0.697
	(0.743)	(1.510)	(1.420)	(1.570)	(1.452)	
2017 Owner Protested (%)	10.853	11.561	11.111	11.891	8.922	0.400
	(0.677)	(1.405)	(1.354)	(1.430)	(1.230)	
2016 Owner Protested (%)	7.773	8.478	6.667	8.187	7.807	0.705
	(0.583)	(1.224)	(1.074)	(1.212)	(1.158)	
Multiple Owners (%)	24.645	22.929	24.444	25.146	26.022	0.693
	(0.938)	(1.847)	(1.851)	(1.917)	(1.893)	
Living Area (1,000s Sq. Feet)	2.313	2.317	2.302	2.331	2.302	0.959
	(0.022)	(0.046)	(0.042)	(0.046)	(0.040)	
Number of Bedrooms	3.428	3.432	3.398	3.423	3.459	0.609
	(0.016)	(0.032)	(0.033)	(0.034)	(0.031)	
Number of Baths	2.273	2.274	2.272	2.292	2.253	0.883
	(0.017)	(0.034)	(0.033)	(0.039)	(0.032)	
b. Survey Variables:						
With Children (%)	25.498	24.470	25.370	26.316	25.836	0.918
	(0.949)	(1.889)	(1.874)	(1.946)	(1.889)	
Female (%)	42.898	44.922	43.774	40.990	41.887	0.574
	(1.086)	(2.200)	(2.157)	(2.191)	(2.145)	
Age	49.608	49.711	49.381	50.438	48.945	0.146
	(0.234)	(0.470)	(0.481)	(0.461)	(0.460)	
Race: White (%)	44.300	44.727	47.818	44.422	40.265	0.103
	(1.092)	(2.200)	(2.178)	(2.220)	(2.134)	
Education: Grad. Degree (%)	38.309	39.844	37.761	38.446	37.240	0.841
	(1.069)	(2.166)	(2.114)	(2.173)	(2.104)	
Prior Belief: School Share (%)	37.642	37.741	37.186	37.935	37.726	0.918
	(0.394)	(0.804)	(0.760)	(0.790)	(0.800)	
Prior Belief: Recapture Share (%)	1.910	1.799	1.372	2.945	1.570	0.216
	(0.287)	(0.632)	(0.505)	(0.593)	(0.564)	
Observations	2,110	519	540	513	538	

Notes: This table lists pre-treatment characteristics' averages. Statistics are based on the 2,110 homeowners that comprise the subject pool. Standard errors are reported in parentheses. The statistics in panel (a) are based on administrative records available on the DCAD's website. The statistics in panel (b) are based on survey responses. Column (1) is based on the entire subject pool. Column (2) is based on homeowners not selected to receive any information (control group). Column (3) is based on homeowners selected to receive information on the recapture share only. Column (4) is based on homeowners selected to receive information on the school share only. Column (5) is based on homeowners selected to receive information on both the school share and the recapture share. Column (6) reports the p-value of a test of equal means across the four treatment groups.

Table 4.2: Main Regression

	P_D^{2021} (1)	I^{2021} (2)	Falsification Tests P_D^{2020} (3)	 P_A^{2021} (4)
a. Effects of School Share:				
With Children	-0.367*	-0.408*	0.110	-0.028
	(0.221)	(0.234)	(0.181)	(0.118)
Without Children	0.277**	0.269*	-0.065	-0.033
	(0.129)	(0.144)	(0.097)	(0.051)
Difference (Children - No Children)	-0.644**	-0.678**	0.175	0.006
	(0.256)	(0.275)	(0.207)	(0.129)
b. Effects of Recapture Share:				
With Children	0.076	-0.313	0.164	-0.321
	(0.485)	(0.541)	(0.417)	(0.278)
Without Children	0.498	-0.101	-0.039	-0.114
	(0.303)	(0.325)	(0.234)	(0.124)
Difference (Children - No Children)	-0.422	-0.212	0.203	-0.207
	(0.563)	(0.620)	(0.468)	(0.297)
Cragg-Donald F-Statistic	30.10	30.22	30.02	30.10
Mean Outcome (Baseline):				
With Children	33.86	47.20	25.98	7.09
Without Children	28.83	44.87	22.19	4.08
Observations	2,110	2,090	2,110	2,110

Notes: Significant at *10%, **5%, ***1%. Robust standard errors in parentheses. This table reports 2SLS estimates of equation (4.21) discussed in Section 4.5.3. Panel (a) reports the estimates corresponding to the school share treatment effect. We present the coefficients for households *with* children and households *without* children, as well as the difference between these two types of households. Panel (b) reports analogous results but for the recapture share treatment effects. The dependent variable in column (1) is an indicator variable that takes the value 100 if the subject protested directly in 2021. The dependent variable in column (2) is an indicator variable that takes the value 100 if the subject answered "very likely" to the question on the subject's protest likelihood in 2021. Columns (3) and (4) report the results of falsification tests. The dependent variable in column (3) is an indicator variable that takes the value 100 if the subject protested directly in 2020. The dependent variable in column (4) is an indicator variable that takes the value 100 if the subject used an agent to protest in 2021. Mean outcomes at baseline correspond with the mean of the dependent variables computed using the group of subjects who did not receive feedback about the school share nor recapture share (the control group).

Table 4.3: Main Regression: Robustness Checks

	P_D^{2021} (1)	I^{2021} (2)	P_D^{2021} (3)	I^{2021} (4)	P_D^{2021} (5)	I^{2021} (6)	P_D^{2021} (7)	I^{2021} (8)	P_D^{2021} (9)	I^{2021} (10)
a. Effects of School Share:										
With Children	-0.367*	-0.408*	-0.429*	-0.457*	-0.330*	-0.250	-0.226	-0.088	-0.369	-0.418*
	(0.221)	(0.234)	(0.225)	(0.235)	(0.190)	(0.205)	(0.168)	(0.191)	(0.237)	(0.247)
Without Children	0.277**	0.269*	0.285**	0.286**	0.196*	0.321**	0.197*	0.256**	0.301**	0.324**
	(0.129)	(0.144)	(0.133)	(0.146)	(0.119)	(0.132)	(0.116)	(0.130)	(0.139)	(0.153)
Difference (Children - No Children)	-0.644**	-0.678**	-0.714***	-0.744***	-0.525**	-0.571**	-0.423**	-0.344	-0.671**	-0.743**
	(0.256)	(0.275)	(0.262)	(0.278)	(0.224)	(0.244)	(0.203)	(0.231)	(0.274)	(0.290)
b. Effects of Recapture Share:										
With Children	0.076	-0.313	0.141	-0.222	0.166	0.135	0.065	0.013	0.231	-0.059
	(0.485)	(0.541)	(0.478)	(0.536)	(0.417)	(0.451)	(0.330)	(0.373)	(0.442)	(0.492)
Without Children	0.498	-0.101	0.436	-0.125	0.414	-0.129	0.247	-0.051	0.473	-0.051
	(0.303)	(0.325)	(0.307)	(0.325)	(0.273)	(0.291)	(0.243)	(0.265)	(0.318)	(0.338)
Difference (Children - No Children)	-0.422	-0.212	-0.295	-0.098	-0.248	0.264	-0.182	0.063	-0.242	-0.009
	(0.563)	(0.620)	(0.559)	(0.616)	(0.500)	(0.536)	(0.394)	(0.438)	(0.527)	(0.579)
Cragg-Donald F-Statistic	30.10	30.22	29.68	29.68	35.26	35.34	47.35	47.55	34.35	34.35
Mean Outcome (Baseline):										
With Children	33.86	47.20	34.68	47.58	35.00	47.10	33.11	47.95	36.27	50.00
Without Children	28.83	44.87	29.12	45.10	29.77	45.64	29.53	46.09	29.33	44.28
Observations	2,110	2,090	2,070	2,070	2,335	2,309	2,482	2,454	1,807	1,807
Additional Controls			√	√						
2.5% Outliers					√	√				
1% Outliers							√	√		
Attention Check									√	√

Notes: Significant at *10%, **5%, ***1%. Robust standard errors in parentheses. This table reports 2SLS estimates of equation (4.21) discussed in Section 4.5.3. Panel (a) reports the estimates corresponding to the school share treatment effect. We present the coefficients for households *with* children and households *without* children separately, as well as the difference between these two types of households. Panel (b) reports analogous results but for the recapture share treatment effects. Columns (1) and (2) are identical to those in Table 4.2 (for reference). The rest of the columns in this table use the same dependent variables from columns (1) and (2). Columns (3) and (4) add additional control variables collected in the survey: age, gender, college degree, and political party. Columns (5) and (6) drop 2.5% of the outliers at each tail of the distribution (instead of the 5% used in the baseline specification). Columns (7) and (8) drop 1% of the outliers at each tail. Columns (9) and (10) restrict the samples to subjects who passed the attention check included in the questionnaire (see Online Appendix for the survey). Each mean outcome corresponds with the mean of the dependent variable for subjects who did not receive feedback about the school share nor the recapture share (the control group).

CHAPTER 5

Concluding Remarks

In this dissertation, I have presented evidence of the effects of different types of policy reforms or information shocks on economic agents' decisions. I have done this by covering three different subjects within the field of public economics. In the first paper, I focused on the cash transfer side of the public sector. The empirical evidence reported here shows that cash transfers can be a powerful tool to reduce poverty and inequality in the long run by affecting individuals' transition to adulthood, especially for women. In the second paper, I focused on how top-income earners respond to changes in the tax schedule. I showed that top-income earners usually respond to changes in tax rates, but the overall efficiency costs are small, given the role of fiscal externalities. Finally, in the third paper, I focused on the determinants of tax compliance. In particular, I provided evidence of the key role of the tax morale mechanism in understanding individuals' tax compliance decisions, which has been elusive in the literature.

Altogether, these papers illustrate the potential of public policy and the information environment surrounding it to affect individuals' decisions with sizable and persistent consequences. A second takeaway refers to the importance for researchers to have access to reliable and comprehensive information, either survey or administrative records, in order to understand the nuances of individuals' behavior. Empirical evidence obtained through the highest scientific standards, e.g., comprehensive and detailed cost-benefit evaluations and welfare assessments based on high-quality information, should be a priority to improve the debate on the optimal design of public policy.

REFERENCES

Aarbu, K. O. and T. O. Thoresen (2001). Income responses to tax changes—evidence from the norwegian tax reform. *National Tax Journal 54*(2), 319–335.

Agrawal, D. R. and D. Foremny (2019). Relocation of the rich: Migration in response to top tax rate changes from spanish reforms. *Review of Economics and Statistics 101*(2), 214–232.

Aizer, A., S. Eli, J. Ferrie, and A. Lleras-Muney (2016, April). The Long-Run Impact of Cash Transfers to Poor Families. *American Economic Review 106*(4), 935–971.

Aizer, A., H. Hoynes, and A. Lleras-Muney (2022, May). Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children. *Journal of Economic Perspectives 36*(2), 149–174.

Ajzenman, N. and R. Durante (2022). Salience and Accountability: School Infrastructure And Last-Minute Electoral Punishment. *Economic Journal, forthcoming*.

Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva (2022). Taxation and Innovation in the Twentieth Century. *The Quarterly Journal of Economics 137*(1), 329–385.

Akee, R. K. Q., W. E. Copeland, G. Keeler, A. Angold, and E. J. Costello (2010, January). Parents' Incomes and Children's Outcomes: A Quasi-experiment Using Transfer Payments from Casino Profits. *American Economic Journal: Applied Economics 2*(1), 86–115.

Almond, D., J. Currie, and V. Duque (2018, December). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature 56*(4), 1360–1446.

Alstadsæter, A. and M. Jacob (2016). Dividend taxes and income shifting. *The Scandinavian Journal of Economics 118*(4), 693–717.

Alstadsæter, A. and M. Jacob (2016). Dividend taxes and income shifting. *The Scandinavian Journal of Economics 118*(4), 693–717. Publisher: Wiley Online Library.

Altonji, J. G. and R. M. Blank (1999, January). Chapter 48 Race and gender in the labor market. In *Handbook of Labor Economics*, Volume 3, pp. 3143–3259. Elsevier.

Alvaredo, F. (2010). The rich in Argentina over the twentieth century, 1932-2004. In P. T. Atkinson A. and S. E. (Eds.), *Top incomes: A Global Perspective*, pp. 253–298. Oxford: Oxford University Press.

Alvaredo, F., L. Chancel, T. Piketty, E. Saez, and G. Zucman (2018). *World inequality report 2018*. Belknap Press.

Alvaredo, F. and J. Londoño Velez (2014). High income and income tax in Colombia, 1993-2010. *Revista de Economía Institucional 16*(31), 157–194.

Amarante, V., M. Manacorda, E. Miguel, and A. Vigorito (2016, May). Do Cash Transfers Improve Birth Outcomes? Evidence from Matched Vital Statistics, and Program and Social Security Data. *American Economic Journal: Economic Policy 8*(2), 1–43.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Araujo, M. C. and K. Macours (2021). Education, Income and Mobility: Experimental Impacts of Childhood Exposure to Progresa after 20 Years. Technical Report IDB-WP-01288.

Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist 55*, 469–480. Place: US Publisher: American Psychological Association.

Athey, S. and G. W. Imbens (2017). The econometrics of randomized experiments. In *Handbook of economic field experiments, Vol. 1*, pp. 73–140.

Atkinson, A. B. (2007). Measuring top incomes: methodological issues. *Top incomes over the twentieth century: A contrast between continental European and English-speaking countries 1*, 18–42.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top incomes in the long run of history. *Journal of economic literature 49*(1), 3–71.

Attanasio, O., L. C. Sosa, C. Medina, C. Meghir, and C. M. Posso-Suárez (2021). Long term effects of cash transfer programs in Colombia. Technical Report w29056, National Bureau of Economic Research.

Attanasio, O. P. and K. M. Kaufmann (2014, July). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics 109*, 203–216.

Attanasio, O. P., C. Meghir, and A. Santiago (2012). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *The Review of Economic Studies 79*(1), 37–66.

Auten, G. and R. Carroll (1999). The Effect of Income Taxes on Household Income. *The Review of Economics and Statistics 81*(4), 681–693.

Auten, G. and L. Kawano (2014). How the rich respond to anticipated tax increases: Evidence from the 1993 tax act. Available at ssrn.

Auten, G., D. Splinter, and S. Nelson (2016). Reactions of high-income taxpayers to major tax legislation. *National Tax Journal, 96*(4), 935–964.

Bagchi, S. and L. Dušek (2021). The effects of introducing withholding and third-party reporting on tax collections: Evidence from the U.S. state personal income tax. *Journal of Public Economics 204*, 104537.

Bailey, M. J. (2006, February). More Power to the Pill: The Impact of Contraceptive Freedom on Women's Life Cycle Labor Supply*. *The Quarterly Journal of Economics 121*(1), 289–320.

Bailey, M. J., H. W. Hoynes, M. Rossin-Slater, and R. Walker (2020). Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program. Technical report, NBER.

Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014). Conditional, Unconditional, and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes. *Journal of Development Effectiveness 6*(1), 1–43. Publisher: Taylor & Francis.

Baird, S., C. McIntosh, and B. Ozler (2011, November). Cash or Condition? Evidence from a Cash Transfer Experiment. *The Quarterly Journal of Economics 126*(4), 1709–1753.

Baird, S. and B. Özler (2016). Transactional Sex in Malawi. In S. Cunningham and M. Shah (Eds.), *The Oxford Handbook of the Economics of Prostitution.* Oxford University Press.

Barham, T., K. Macours, and J. A. Maluccio (2018). Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes.pdf. Technical report.

Barr, A., J. Eggleston, and A. A. Smith (2022). Investing in Infants: the Lasting Effects of Cash Transfers to New Families*. *The Quarterly Journal of Economics Forthcoming*, qjac023.

Barrett, G. F. and D. S. Hamermesh (2019). Labor supply elasticities: Overcoming nonclassical measurement error using more accurate hours data. *Journal of Human Resources 54*(1), 255–265.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, and T. Schmidt (2019). The Impact of Cash Transfers: A Review of the Evidence from Low- and Middle-Income Countries. *Journal of Social Policy 48*(3), 569–594. Publisher: Cambridge University Press.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, T. Schmidt, and L. Pellerano (2016). Cash Transfers: What Does the Evidence Say? A Rigourus Review of Programme Impact and the Role of Design and Implementation Features. Tech. Rep., Overseas Dev. Inst., London.

Bastani, S., Y. Moberg, and H. Selin (2021). The anatomy of the extensive margin labor supply response. *The Scandinavian Journal of Economics 123*(1), 33–59.

Bastian, J., L. Bian, and J. Grogger (2022, August). How Did Safety-Net Reform Affect the Education of Adolescents from Low-Income Families? *Labour Economics 77*, 102031.

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

Becker, G. S. and H. G. Lewis (1973). On the Interaction between the Quantity and Quality of Children. *Journal of political Economy 81*(2, Part 2), 279–288.

Behrman, J. R., S. W. Parker, and P. E. Todd (2011). Do Conditional Cash Transfers for Schooling Generate Lasting Benefits?: A Five-Year Followup of PROGRESA/Oportunidades. *Journal of Human Resources 46*(1), 203–236.

Bergolo, M., G. Burdin, M. De Rosa, M. Giaccobasso, and M. Leites (2021). Digging Into the Channels of Bunching: Evidence from the Uruguayan Income Tax. *The Economic Journal 131*(639), 2726–2762.

Bergolo, M. and G. Cruces (2021). The anatomy of behavioral responses to social assistance when informal employment is high. *Journal of Public Economics 193*, 104313.

Bergolo, M. and E. Galván (2018, March). Intra-household Behavioral Responses to Cash Transfer Programs. Evidence from a Regression Discontinuity Design. *World Development 103*, 100–118.

Bergolo, M., M. Leites, R. Perez-Truglia, and M. Strehl (2020a). What makes a tax evader? Working Paper 28235, National Bureau of Economic Research.

Bergolo, M., M. Leites, R. Perez-Truglia, and M. Strehl (2020b). What Makes a Tax Evader? *NBER Working Paper No. 28235*.

Berthelon, M. E. and D. I. Kruger (2011, February). Risky behavior among youth: Incapacitation effects of school on adolescent motherhood and crime in Chile. *Journal of Public Economics 95*(1), 41–53.

Best, M. and H. Kleven (2012). Optimal Income Taxation with Career Effects of Work Effort. Available at ssrn.

Bitler, M. P. and T. Figinski (2019). Long-run effects of food assistance: Evidence from the Food Stamp Program. *Economic Self-Sufficiency Policy Research Institute*.

Bjørneby, M., A. Alstadsæter, and K. Telle (2021). Limits to third-party reporting: Evidence from a randomized field experiment in norway. *Journal of Public Economics 203*, 104512.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2008). Staying in the Classroom and Out of the Maternity Ward? The Effect of Compulsory Schooling Laws on Teenage Births. *The Economic Journal 118*(530), 1025–1054.

Blau, F. D. and L. M. Kahn (2017, September). The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature 55*(3), 789–865.

Blumenthal, M., C. Christian, and J. Slemrod (2001). Do Normative Appeals Affect Tax Compliance? Evidence From a Controlled Experiment in Minnesota. *National Tax Journal 54*(1), 125–138.

Blundell, R. and M. C. Dias (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources 44*(3), 565–640.

Bobonis, G. J. and F. Finan (2009, November). Neighborhood Peer Effects in Secondary School Enrollment Decisions. *The Review of Economics and Statistics 91*(4), 695–716.

Bosch, M. and M. Manacorda (2012). Social Policies and Labor Market Outcomes in Latin America and the Caribbean - A Review of the Existing Evidence. Technical Report CEPOP32, The London School of Economics and Political Science, Center of Economic Performance, London, UK.

Bott, K. M., A. W. Cappelen, E. Sørensen, and B. Tungodden (2020). You've Got Mail: A Randomized Field Experiment on Tax Evasion. *Management Science 66*(7), 2801–2819.

Bottan, N. and R. Perez-Truglia (2020). Betting on the House: Subjective Expectations and Market Choices. *NBER Working Paper No. 27412*.

Bottan, N. L. and R. Perez-Truglia (2022). Choosing Your Pond: Location Choices and Relative Income. *The Review of Economics and Statistics 104*(5), 1010–1027.

Bratti, M. (2015). Fertility Postponement and Labor Market Outcomes. *IZA World of Labor*.

Bratti, M. and L. Cavalli (2014, February). Delayed First Birth and New Mothers' Labor Market Outcomes: Evidence from Biological Fertility Shocks. *European Journal of Population 30*(1), 35–63.

Browning, M. and P. A. Chiappori (1998). Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica 66*(6), 1241–1278. Publisher: [Wiley, Econometric Society].

Bulman, G., R. Fairlie, S. Goodman, and A. Isen (2021, April). Parental Resources and College Attendance: Evidence from Lottery Wins. *American Economic Review 111*(4), 1201–1240.

Burdín, G., M. De Rosa, A. Vigorito, and J. Vilá (2022). Falling inequality and the growing capital income share: Reconciling divergent trends in survey and tax data. *World Development 152*, 105783.

Cabella, W. and C. Velázquez (2022). Abortion Legalization in Uruguay: Effects on Adolescent Fertility. *Studies in Family Planning 53*(3), 491–514. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12204.

Cahyadi, N., R. Hanna, B. A. Olken, R. A. Prima, E. Satriawan, and E. Syamsulhakim (2020, November). Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia. *American Economic Journal: Economic Policy 12*(4), 88–110.

Cait, L., J.-E. De Neve, and M. I. Norton (2018). The Power of Voice in Stimulating Morality: Eliciting Taxpayer Preferences Increases Tax Compliance. *Special Issue: Marketplace Morality*, 310–328.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics 101*(3), 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica 82*(6), 2295–2326. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA11757.

Carrillo, P. E., E. Castro, and C. Scartascini (2021). Public good provision and property tax compliance: Evidence from a natural experiment. *Journal of Public Economics 198*, 104422.

Castro, L. and C. Scartascini (2015). Tax Compliance and Enforcement in the Pampas Evidence From a Field Experiment. *Journal of Economic Behavior & Organization 116*, 65–82.

Cattaneo, M. D., N. Idrobo, and R. Titiunik (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Cattaneo, M. D., M. Jansson, and X. Ma (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal 18*(1), 234–261. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

Cavallo, A., G. Cruces, and R. Perez-Truglia (2017). Inflation expectations, learning, and supermarket prices: Evidence from survey experiments. *American Economic Journal: Macroeconomics 9*(3), 1–35.

Cellini, S. R., F. Ferreira, and J. Rothstein (2010, February). The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design*. *The Quarterly Journal of Economics 125*(1), 215–261.

Ceni, R., C. Parada, I. Perazzo, and E. Sena (2021). Birth Collapse and a Large-Scale Access Intervention with Subdermal Contraceptive Implants. *Studies in Family Planning 52*(3), 321–342. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12171.

Cesarini, D., E. Lindqvist, R. Östling, and B. Wallace (2016, May). Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players *. *The Quarterly Journal of Economics 131*(2), 687–738.

Chen, G. (2021). An Overview of the Funding of Public Schools. *https://www.publicschoolreview.com/blog/an-overview-of-the-funding-of-public-schools*.

Chetty, R. (2009). Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance. *American Economic Journal: Economic Policy 1*(2), 31–52.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star *. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Chiapa, C., J. L. Garrido, and S. Prina (2012, October). The effect of social programs and exposure to professionals on the educational aspirations of the poor. *Economics of Education Review 31*(5), 778–798.

Chiappori, P.-A. (1988). Rational Household Labor Supply. *Econometrica 56*(1), 63–90. Publisher: [Wiley, Econometric Society].

Chiappori, P.-A. (1992, June). Collective Labor Supply and Welfare. *Journal of Political Economy 100*(3), 437–467. Publisher: The University of Chicago Press.

Clark, D. and P. Martorell (2014). The Signaling Value of a High-School Diploma. *Journal of Political Economy 122*(2), 282–318.

Conger, R. D., K. J. Conger, G. H. Elder, F. O. Lorenz, R. L. Simons, and L. B. Whitbeck (1993). Family Economic Stress and Adjustment of Early Adolescent Girls. *Developmental Psychology 29*(2), 206–219. Place: US Publisher: American Psychological Association.

Cullen, J., N. Turner, and E. Washington (2020). Political alignment, attitudes toward government and tax evasion. *American Economic Journal: Economic Policy, forthcoming*.

Cullen, Z. and R. Perez-Truglia (2022). How Much Does Your Boss Make? The Effects of Salary Comparisons. *Journal of Political Economy 130*(3), 766–822.

Cunha, F. and J. Heckman (2007). The Technology of Skill Formation. *American Economic Review 97*(2), 31–47.

Currie, J. (2009). Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development. *Journal of Economic Literature 47*(1), 87–122.

Dahl, G. B. and A. C. Gielen (2021, April). Intergenerational Spillovers in Disability Insurance. *American Economic Journal: Applied Economics 13*(2), 116–150.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014, November). Family Welfare Cultures *. *The Quarterly Journal of Economics 129*(4), 1711–1752.

Dallas Morning News (2018). Do your schools get your property tax dollars? *July 4, 2018*.

Davis, R. D. (2021, April). More than 250 advocate groups urge White House to fight child poverty | Campaign For Children.

De Chaisemartin, C. and X. D'Haultfoeuille (2022). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. Technical report, National Bureau of Economic Research.

De Neve, J.-E., C. Imbert, T. Tsankova, and M. Luts (2021). How to Improve Tax Compliance? Evidence from Population-wide Experiments in Belgium. *Journal of Political Economy, forthcoming*.

De Rosa, M., I. Flores, and M. Morgan (2020). Inequality in Latin America Revisited: Insights from Distributional National Accounts. WIL Technical Notes.

DellaVigna, S. and E. Linos (2022). Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica 90*(1), 81–116.

DellaVigna, S., D. Pope, and E. Vivalt (2019). Predict science to improve science. *Science 366*(6464), 428–429.

Deshpande, M. (2016, November). Does Welfare Inhibit Success? The Long-Term Effects of Removing Low-Income Youth from the Disability Rolls. *American Economic Review 106*(11), 3300–3330.

Dobay, N., F. Nicely, A. Sanderson, and P. Sanderson (2019). The best (and worst) of international property tax administration. Technical report, Council On State Taxation.

Doerrenberg, P., A. Peichl, and S. Siegloch (2017). The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics 151*, 41–55.

Duflo, E. (2003, June). Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa. *The World Bank Economic Review 17*(1), 1–25.

Duncan, G. J. and S. D. Hoffman (1990, November). Welfare Benefits, Economic Opportunities, and Out-of-Wedlock Births Among Black Teenage Girls. *Demography 27*(4), 519.

Dustan, A. (2020, March). Can Large, Untargeted Conditional Cash Transfers Increase Urban High School Graduation Rates? Evidence from Mexico City's Prepa Sí. *Journal of Development Economics 143*, 102392.

Dynarski, S. M. (2003, March). Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion. *American Economic Review 93*(1), 279–288.

ECLAC (2013). *Fiscal Panorama of Latin America and the Caribbean. Tax Reform and Renewal oft he Fiscal Covenant* (United Nations ed.). Santiago, Chile: Economic Commission for Latin America and the Caribbean (ECLAC).

Eissa, N. (1995). Taxation and labor suuply of married women: The tax reform act of 1986 as a natural experiment. Working Paper 5023, National Bureau of Economic Research.

Feldstein, M. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy 103*(3), 551–572.

Feldstein, M. and D. Feenberg (1996). The effect of increased tax rates on taxable income and economic efficiency: A preliminary analysis of the 1993 tax rate increases. *Tax policy and the economy 10*, 89–117.

Fiszbein, A., N. R. Schady, F. H. G. Ferreira, M. Grosh, N. Keleher, P. Olinto, and E. Skoufias (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington D.C: World Bank.

Flores, I., C. Sanhueza, J. Atria, and R. Mayer (2020). Top incomes in chile: A historical perspective on income inequality, 1964–2017. *Review of Income and Wealth 66*(4), 850–874.

Foremny, D., L. Muinelo-Gallo, and J. Vázquez-Grenno (2018). Intertemporal Income Shifting and Tax Evasion: Evidence From an Uruguayan Tax Reform. Available at SSRN.

Fuster, A., R. Perez-Truglia, M. Wiederholt, and B. Zafar (2022). Expectations with Endogenous Information Acquisition: An Experimental Investigation. *Review of Economics and Statistics 104*(5), 1059–1078.

Gershoff, E. T., J. L. Aber, C. C. Raver, and M. C. Lennon (2007). Income Is Not Enough: Incorporating Material Hardship Into Models of Income Associations With Parenting and Child Development. *Child Development 78*(1), 70–95. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8624.2007.00986.x.

Giertz, S. H. (2010). The Elasticity of Taxable Income during the 1990s: New Estimates and Sensitivity Analyses. *Southern Economic Journal 77*(2), 406–433.

Goolsbee, A. (2000). What happens when you tax the rich? evidence from executive compensation. *Journal of Political Economy 108*(2), 352–378.

Gordon, R. and W. Li (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics 93*(7-8), 855–866.

Gordon, R. and J. Slemrod (2000). Are Real Responses to Taxes Simply Income Shifting between Corporate and Personal Tax Bases? In *Does Atlas Shrug? The Economic Consequences of Taxing the Rich* (J. Slemrod ed.). Cambridge, MA: Harvard University Press and Russell Sage.

Gruber, J. and E. Saez (2002). The elasticity of taxable income: evidence and implications. *Journal of public Economics 84*(1), 1–32.

Gustafsson, S. (2001, June). Optimal age at motherhood. Theoretical and empirical considerations on postponement of maternity in Europe. *Journal of Population Economics 14*(2), 225–247.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica 69*(1), 201–209. Publisher: [Wiley, Econometric Society].

Harju, J. and T. Matikka (2016). The elasticity of taxable income and income-shifting: what is "real" and what is not? *International Tax and Public Finance 23*(4), 640–669.

Hartley, R. P., C. Lamarche, and J. P. Ziliak (2022, March). Welfare Reform and the Intergenerational Transmission of Dependence. *Journal of Political Economy 130*(3), 523–565. Publisher: The University of Chicago Press.

Hermle, J. and A. Peichl (2018). Jointly optimal taxes for different types of income. *Available at SSRN 3275422*.

Hoff, P. D. (2009). *A first course in Bayesian statistical methods.* Springer Science & Business Media.

Holmlund, B. and M. Söderström (2011). Estimating dynamic income responses to tax reform. *The B.E. Journal of Economic Analysis & Policy 11*(1).

Hoynes, H., D. W. Schanzenbach, and D. Almond (2016). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review 106*(4), 903–934.

Hoynes, H. W. and D. W. Schanzenbach (2018). Safety Net Investments in Children.

Huet-Vaughn, E. (2019, 2). Stimulating the Vote: ARRA Road Spending and Vote Share. *American Economic Journal: Economic Policy 11*(1), 292–316.

Huet-Vaughn, E., A. Robbett, and M. Spitzer (2019). A taste for taxes: Minimizing distortions using political preferences. *Journal of Public Economics 180*, 104055.

Iacus, S., G. King, and G. Porro (2012). Causal inference without balance checking: Coarsened exact matching. *Policy Analysis 20*, 1–24.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–475.

Imbens, G. W. and T. Lemieux (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics 142*(2), 615–635.

Institute, A. E. (2021). The Conservative Case Against Child Allowances.

Jensen, A. (2022). Employment structure and the rise of the modern tax system. *American Economic Review 112*(1), 213–34.

Jensen, R. (2010, May). The (Perceived) Returns to Education and the Demand for Schooling*. *The Quarterly Journal of Economics 125*(2), 515–548.

Jepsen, C., P. Mueser, and K. Troske (2016). Labor Market Returns to the GED Using Regression Discontinuity Analysis. *Journal of Political Economy 124*(3), 621–649. Publisher: University of Chicago Press Chicago, IL.

Jones, P. (2019). Loss Aversion and Property Tax Avoidance. *Working Paper*.

Jouste, M., T. Kaidu, O. Ayo, and J. P. R. Pirttila (2021). The effects of personal income tax reform on employees' taxable income in uganda. Available at wider.

Kawano, L., C. Weber, and A. Whitten (2016). Estimating the elasticity of broad income for high-income taxpayers. Available at ssrn.

Keane, M. P. and K. I. Wolpin (2010). The Role of Labor and Marriage Markets, Preference Heterogeneity, and the Welfare System in the Life Cycle Decisions of Black, Hispanic, and White Women*. *International Economic Review 51*(3), 851–892. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-2354.2010.00604.x.

Kearney, M. S. and P. B. Levine (2009, February). Subsidized Contraception, Fertility, and Sexual Behavior. *The Review of Economics and Statistics 91*(1), 137–151.

Kearney, M. S. and P. B. Levine (2014). Income Inequality and Early Nonmarital Childbearing. *Journal of Human Resources 49*(1), 1–31.

Klepinger, D., S. Lundberg, and R. Plotnick (1999). How Does Adolescent Fertility Affect the Human Capital and Wages of Young Women? *The Journal of Human Resources 34*(3), 421.

Kleven, H., C. Landais, M. Muñoz, and S. Stantcheva (2020). Taxation and migration: Evidence and policy implications. *Journal of Economic Perspectives 34*(2), 119–42.

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller (2019). Child Penalties across Countries: Evidence and Explanations. *AEA Papers and Proceedings 109*, 122–126.

Kleven, H., C. Landais, and J. E. Søgaard (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics 11*(4), 181–209.

Kleven, H. J., C. Landais, and E. Saez (2013). Taxation and international migration of superstars: Evidence from the european football market. *American Economic Review 103*(5), 1892–1924.

Kleven, H. J. and E. A. Schultz (2014). Estimating Taxable Income Responses Using Danish Tax Reforms. *American Economic Journal: Economic Policy 6*(4), 271–301.

Kohler, H.-P., F. C. Billari, and J. A. Ortega (2002). The Emergence of Lowest-Low Fertility in Europe During the 1990s. *Population and Development Review 28*(4), 641–680. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4457.2002.00641.x.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics 89*(11), 2093–2119.

Kopczuk, W. (2012). The polish business "flat" tax and its effect on reported incomes: a pareto improving tax reform. In *Columbia University Working Paper.*

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016). Tax reforms and intertemporal shifting of wage income: Evidence from danish monthly payroll records. *American Economic Journal: Economic Policy 8*(3), 233–57.

Kresch, E. P., M. Walker, M. C. Best, F. Gerard, and J. Naritomi (2023). Sanitation and property tax compliance: Analyzing the social contract in Brazil. *Journal of Development Economics 160*, 102954.

Lalive, R. and M. A. Cattaneo (2009, August). Social Interactions and Schooling Decisions. *The Review of Economics and Statistics 91*(3), 457–477.

Londoño-Vélez, J. and J. A. Mahecha (2021). Behavioral responses to wealth taxation: Evidence from colombia. Working paper.

LoPiccalo, K., J. Robinson, and E. Yeh (2016). Income, Income Shocks, and Transactional Sex. In S. Cunningham and M. Shah (Eds.), *The Oxford Handbook of the Economics of Prostitution.* Oxford University Press.

Lundberg, S. and R. D. Plotnick (1995). Adolescent Premarital Childbearing: Do Economic Incentives Matter? *Journal of Labor Economics 13*(2), 177–200.

Luttmer, E. F. P. and M. Singhal (2014). Tax Morale. *Journal of Economic Perspectives 28*(4), 149–168.

MacLeod, W. B. and M. Urquiola (2019). Is Education Consumption or Investment? Implications for School Competition. *Annual Review of Economics 11*(1), 563–589. _eprint: https://doi.org/10.1146/annurev-economics-080218-030402.

Manacorda, M., E. Miguel, and A. Vigorito (2011, July). Government Transfers and Political Support. *American Economic Journal: Applied Economics 3*(3), 1–28.

Mankiw, N. G. and M. Weinzierl (2010). The optimal taxation of height: A case study of utilitarian income redistribution. *American Economic Journal: Economic Policy 2*(1), 155–176.

Manoli, D. and N. Turner (2018, May). Cash-on-Hand and College Enrollment: Evidence from Population Tax Data and the Earned Income Tax Credit. *American Economic Journal: Economic Policy 10*(2), 242–271.

Martinelli, C. and S. W. Parker (2003). should Transfers To Poor Families Be Conditional On School Attendance? A Household Bargaining Perspective*. *International Economic Review 44*(2), 523–544. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-2354.t01-1-00079.

Martinelli, C. and S. W. Parker (2008). Do School Subsidies Promote Human Capital Investment among the Poor? *The Scandinavian Journal of Economics 110*(2), 261–276. Publisher: [Wiley, The Scandinavian Journal of Economics].

Martorano, B. (2018). Taxation and Inequality in Developing Countries: Lessons from the Recent Experience of Latin America. *Journal of International Development 30*(2), 256–273.

Martorell, P. and I. McFarlin, Jr (2011). Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes. *The Review of Economics and Statistics 93*(2), 436–454.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics 99*(2), 210–221.

Miao, D., H. Selin, and M. Soderstrom (2020). Earnings responses to even higher taxes. Working paper.

Michelmore, K. and L. M. Lopoo (2021, December). The Effect of EITC Exposure in Childhood on Marriage and Early Childbearing. *Demography 58*(6), 2365–2394.

Miller, A. R. (2011, July). The Effects of Motherhood Timing on Career Path. *Journal of Population Economics 24*(3), 1071–1100.

Milligan, K. and M. Smart (2015). Taxation and top incomes in Canada. *Canadian Journal of Economics/Revue canadienne d'économique 48*(2), 655–681. Publisher: Wiley Online Library.

Mills, M., R. R. Rindfuss, P. McDonald, E. te Velde, and on behalf of the ESHRE Reproduction and Society Task Force (2011, November). Why do people postpone parenthood? Reasons and social policy incentives. *Human Reproduction Update 17*(6), 848–860.

Moffitt, R. and M. Wilhelm (2000). Taxation and the Labor Supply Decisions of the Affluent. In Slemrod, Joel (Ed.), *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, pp. 193–234. Cambridge: Russell Sage Foundation Books at Harvard University Press.

Molina Millán, T., T. Barham, K. Macours, J. A. Maluccio, and M. Stampini (2019). Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence. *The World Bank Research Observer 34*(1), 119–159.

Molina Millán, T., K. Macours, J. A. Maluccio, and L. Tejerina (2020, March). Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program. *Journal of Development Economics 143*, 102385.

Morgan, M. (2017). Extreme and persistent inequality: New evidence for Brazil combining national accounts, surveys and fiscal data, 2001-2015. *World Inequality Database (WID. org) Working Paper Series 12*, 1–50.

Muñoz, M. (2019). Do european top earners react to labour taxation through migration? Available at world inequality lab.

Musgrave, R. (1959). The Theory of Public Finance. *McGraw-Hill*.

Nathan, B., R. Perez-Truglia, and A. Zentner (2020). My Taxes are Too Darn High: Why Do Households Protest their Taxes? *NBER Working Paper No. 27816*.

Neisser, C. (2021). The elasticity of taxable income: A meta-regression analysis. *The Economic Journal 131*(640), 3365–3391.

Olivetti, C. and B. Petrongolo (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics 8*(1), 405–434. _eprint: https://doi.org/10.1146/annurev-economics-080614-115329.

Oreopoulos, P. (2011). Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes. *American Economic Journal: Economic Policy 3*(4), 148–171.

Parker, S. and T. Vogl (2018, February). Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico. Technical Report w24303, National Bureau of Economic Research, Cambridge, MA.

Parker, S. W. and P. E. Todd (2017). Conditional Cash Transfers: The Case of *Progresa/Oportunidades*. *Journal of Economic Literature 55*(3), 866–915.

Parker, W. and N. Friedman (2021). Zillow Quits Home-Flipping Business, Cities Inability to Forecast Prices. *The Wall Street Journal, November 2 2021*.

Piketty, T. (2003). Income inequality in France, 1901–1998. *Journal of political economy 111*(5), 1004–1042.

Piketty, T., E. Saez, and S. Stantcheva (2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities. *American Economic Journal: Economic Policy 6*(1), 230–271.

Piketty, T., E. Saez, and G. Zucman (2018). Distributional National Accounts: Methods and Estimates for the United States. *The Quarterly Journal of Economics 133*(2), 553–609.

Pirttilä, J. and H. Selin (2011). Income shifting within a dual income tax system: Evidence from the finnish tax reform of 1993. *Scandinavian Journal of Economics 113*(1), 120–144.

Pomeranz, D. and J. Vila-Belda (2019). Taking state-capacity research to the field: Insights from collaborations with tax authorities. *Annual Review of Economics 11*(1), 755–781.

Price, D. J. and J. Song (2018). The Long-Term Effects of Cash Assistance. pp. 87.

Querejeta, M. and M. Bucheli (2022, October). The Effect of Childbirth on Women's Formal Labour Market Trajectories: Evidence from Uruguayan Administrative Data. *The Journal of Development Studies 0*(0), 1–15. Publisher: Routledge _eprint: https://doi.org/10.1080/00220388.2022.2128777.

Ramírez Leira, L. (2021). Segregación escolar público-privado por nivel socioeconómico en Uruguay: Un análisis en base a microdescomposiciones. Working Paper 275, Documento de Trabajo.

Romanov, D. (2006). The corporation as a tax shelter: Evidence from recent israeli tax changes. *Journal of Public Economics 90*(10), 1939–1954.

Rosero-Bixby, L., T. Castro-Martín, and T. Martín-García (2009). Is Latin America starting to retreat from early and universal childbearing? *Demographic Research 20*, 169–194. Publisher: Max-Planck-Gesellschaft zur Foerderung der Wissenschaften.

Rothschild, C. and F. Scheuer (2016). Optimal Taxation with Rent-Seeking. *The Review of Economic Studies 83*(3), 1225–1262.

Saez, E. (2003). The effect of marginal tax rates on income: a panel study of 'bracket creep'. *Journal of Public Economics 87*(5-6), 1231–1258.

Saez, E. (2004). Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications. NBER Working Paper 10273, National Bureau of Economic Research.

Saez, E. (2017). Taxing the rich more: Preliminary evidence from the 2013 tax increase. *Tax Policy and the Economy 31*(1), 71–120.

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

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45.

Scherf, R. and M. Weinzierl (2020). Understanding Different Approaches to Benefit-Based Taxation. *Fiscal Studies 41*(2), 385–410.

Schmidt, L., T. Sobotka, J. Bentzen, A. Nyboe Andersen, and on behalf of the ESHRE Reproduction and Society Task Force (2012, January). Demographic and Medical Consequence of the Postponement of Parenthood. *Human Reproduction Update 18*(1), 29–43.

Seligman, E. R. A. (1908). Progressive Taxation in Theory and Practice. Technical Report 4.

Settersten Jr, R. A., F. F. Furstenberg, and R. G. Rumbaut (2008, September). *On the Frontier of Adulthood: Theory, Research, and Public Policy*. University of Chicago Press. Google-Books-ID: zAEV2pTN16EC.

Sinclair, M., J. O'Toole, M. Malawaraarachchi, and K. Leder (2012). Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. *BMC Medical Research Methodology 12*(1), 132.

Slemrod, J. (1995). Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986. *The American Economic Review 85*(2), 175–180.

Slemrod, J. (2001). A general model of the behavioral response to taxation. *International Tax and Public Finance 8*, 119–128.

Slemrod, J. (2019a). Tax Compliance and Enforcement. *Journal of Economic Literature 57*(4), 904–954.

Slemrod, J. (2019b). Tax Compliance and Enforcement. *Journal of Economic Literature 57*(4), 904–954.

Slemrod, J. and W. Kopczuk (2002). The optimal elasticity of taxable income. *Journal of Public Economics 84*(1), 91–112.

Sobotka, T. (2004). Is Lowest-Low Fertility in Europe Explained by the Postponement of Childbearing? *Population and Development Review 30*(2), 195–220. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4457.2004.010_1.x.

Sobotka, T. (2010). Shifting Parenthood to Advanced Reproductive Ages: Trends, Causes and Consequences. In *A Young Generation Under Pressure*, pp. 129–154. Berlin, Heidelberg: Springer.

Statistical Atlas (2023). The Demographic Statistical Atlas of the United States: Household Types in Dallas County, Texas. https://statisticalatlas.com/county/Texas/Dallas-County/Household-Types. Accessed: 2023-03-27.

Stock, J. H., J. H. Wright, and M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics 20*(4), 518–529.

Tax Policy Center (2021a). Amount of Revenue by Source. *https://www.taxpolicycenter.org/statistics/property-tax-revenue*.

Tax Policy Center (2021b). Amount of Revenue by Source. *https://www.taxpolicycenter.org/statistics/amount-revenue-source*.

Tazhitdinova, A. (2020a). Are changes of organizational form costly? income shifting and business entry responses to taxes. *Journal of Public Economics 186*, 104187.

Tazhitdinova, A. (2020b). Do only tax incentives matter? labor supply and demand responses to an unusually large and salient tax break. *Journal of Public Economics 184*, 104–162.

Tazhitdinova, A. (2022). Increasing hours worked: Moonlighting responses to a large tax reform. *American Economic Journal: Economic Policy 14*(1), 473–500.

Texas Comptroller (2021). Frequently Asked Questions. *https://comptroller.texas.gov/taxes/property-tax/exemptions/age65older-disabled-faq.php*.

Texas Education Agency (2021). What is House Bill 3? *https://tea.texas.gov/about-tea/government-relations-and-legal/government-relations/house-bill-3*.

Thistlethwaite, D. L. and D. T. Campbell (1960). Regression-Discontinuity Analysis: An Alternative to the Ex-post Facto Experiment. *Journal of Educational Psychology 51*(6), 309–317.

Thomas, D. (1990). Intra-Household Resource Allocation: An Inferential Approach. *The Journal of Human Resources 25*(4), 635–664. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

Todd, P. E. and K. I. Wolpin (2006). Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *American Economic Review 96*(5), 1384–1417.

Todd, P. E. and K. I. Wolpin (2008). Ex Ante Evaluation of Social Programs. *Annales d'Économie et de Statistique* (91/92), 263–291. Publisher: [GENES, ADRES].

Tortarolo, D., G. Cruces, and V. Castillo (2020). It takes two to tango: labor responses to an income tax holiday in argentina. Discussion Paper 2020 07, Nottingham Interdisciplinary Centre for Economic and Political Research (NICEP).

U.S. Census Bureau (2021). Population, Dallas County, Texas. *https://www.census.gov/quickfacts/fact/table/dallascountytexas/POP010220*.

Van Bavel, J. (2010, May). Choice of study discipline and the postponement of motherhood in Europe: The impact of expected earnings, gender composition, and family attitudes. *Demography 47*(2), 439–458.

Villanueva, C. (2018). What is Recapture? *Center for Public Policy Priorities Report, August 30, 2018*.

Waldfogel, J. (1998). Understanding the "Family Gap" in Pay for Women with Children. *The Journal of Economic Perspectives 12*(1), 137–156.

Waseem, M. (2018). Taxes, informality and income shifting: Evidence from a recent Pakistani tax reform. *Journal of Public Economics 157*, 41–77.

Weber, C. E. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics 117*, 90–103.

Weinzierl, M. (2014). The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. *Journal of Public Economics 118*, 128–142.

Weinzierl, M. (2017). Popular acceptance of inequality due to innate brute luck and support for classical benefit-based taxation. *Journal of Public Economics 155*, 54–63.

Weinzierl, M. (2018). Revisiting the Classical View of Benefit-based Taxation. *The Economic Journal 128*(612), F37–F64.

Wolfe, B., K. Wilson, and R. Haveman (2001). The role of economic incentives in teenage nonmarital childbearing choicesq. *Journal of Public Economics 81*, 39.

World Bank (2019). The Administrative Review Process for Tax Disputes: Tax Objections and Appeals in Latin America and the Caribbean.

Yeung, W. J., M. R. Linver, and J. Brooks–Gunn (2002). How Money Matters for Young Children's Development: Parental Investment and Family Processes. *Child Development 73*(6), 1861–1879. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8624.t01-1-00511.

Youngman, J. (2016). *A Good Tax.* New York: Columbia University Press.

Zawisza, T. (2019). Optimal taxation of employment and self-employment: evidence from poland and implications. EUI MWP 2019/08, European University InstituteEUI MWP.