Essays in Public Finance and Taxation

by

Ellen Stuart

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2021

Doctoral Committee:

Professor Joel Slemrod, Chair
Professor Charles Brown
Professor Ashley Craig
Professor James R. Hines, Jr.
Professor Toni Whited

Ellen Stuart
stuartem@umich.edu
ORCID ID 0000-0002-0929-9024

To my family and my chosen family

ACKNOWLEDGEMENTS

This dissertation would not exist if not for the support of a large community of people.

I am forever indebted to the guidance, support, patience, and kindness of my committee: Joel Slemrod, Jim Hines, Charlie Brown, Toni Whited, and Ashley Craig. They have, without fail, treated me like a human being first and an academic second. It would be difficult to overstate what a positive difference this distinction made on my experience in graduate school. I hope to carry forward their insightfulness as scholars and thoughtfulness as mentors into all of my future endeavors.

I doubt I would have become a tax researcher if I had not taken Public Finance with Jim and Joel. I am so glad to have ended up here, and am proud to be a member of the Michigan Public Finance family. I am especially thankful to Joel for inviting me to join the project that ultimately became Chapter II. Working with Joel taught me an incredible amount about the research process in economics, and introduced me to the academic community working with administrative tax data. Jim has been a source of inspiration since our first day of class. I've benefitted greatly from his advice and perspective about research and the job market. I'm glad that Ash convinced me to be his GSI for Econ 490, which was an immensely rewarding experience. It is not lost on me that every time I was stressed about my job market paper or the job market and Ash told me it would be fine, he was right.

Over the past six years, I received incredible support from a number of faculty not formally on my committee. I'm particularly grateful to John Bound and Jeff Smith, who have both been amazing sources of guidance and encouragement throughout my graduate school career. Martha Bailey provided excellent counsel over many cups of tea. Andreas Hagemann and Dominick Bartelme both gave me the exact right advice at the right time (Andreas in first year, and Dominick in third year). Mike Mueller-Smith was a great mentor during my time with the Criminal Justice Administrative Records System (CJARS) team, where I learned numerous skills that I continue to use on a daily basis. I'm also grateful for the support of the admin team. In particular, Laura Flak and Lauren Pulay have gone above and beyond to help me navigate everything from maternity leave to continuity of care during the COVID-19 pandemic. I also thank Mary Ceccanese for everything she does to help the Michigan Public Finance family thrive.

I am deeply thankful for the devoted faculty of the Economics department of Williams College, who have been unwavering in their support of me and my professional pursuits for over 10 years. My continued correspondence with Sarah Jacobson, Sara LaLumia, and Lucie Schmidt has been inspiring. Despite the many demands on his time as Provost, Dukes Love spent almost two hours one afternoon introducing me to the world of life cycle modeling. That conversation gave me the confidence to pursue what ultimately became the second half of Chapter I. I also want to mention Ralph Bradburd, who continues to rank for me as one of the best teachers I have ever had.

The research in this dissertation was made possible by the Joint Statistical Research Program run by the Statistics of Income (SOI) division at the IRS. It would not have been possible without the wonderful team at the Research, Applied Analytics, and Statistics (RAAS) Division. In particular, I want to thank Victoria Bryant, John Guyton, Ron Hodge, Alicia Miller, Mark Payne, and Alex Turk. I am especially grateful for Vickie's incredible responsiveness and attention to detail throughout the process of writing Chapter I. For all three chapters: any opinions and conclusions expressed are those of the authors and do not necessarily represent the views of the Internal Revenue Service. All results have been reviewed to ensure that no confidential information is disclosed.

Chapter I was supported by the NBER Pre-Doctoral Fellowship in Retirement and Disability Policy Research. The research was performed pursuant to grant RDR18000003 from the U.S. Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of SSA, any agency of the Federal Government, or NBER. Neither the United States (U.S.) Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this report. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the U.S. Government or any agency thereof.

Chapter I was immeasurably improved by the help and guidance I received when I was learning how to build the structural model. Toni Whited provided me with fantastic code and lecture notes, without which I would not have known where to start. She also answered my questions with insight and clarity, and the model was immeasurably improved by her input. The model would almost surely not exist without the many, many long conversations I had with Dhiren Patki. I cannot fully express how grateful I am for his patience, know-how, and genuine excitement about the project. I also received general programming help from Jonathan Kummerfeld and Aaron Bauer, for which I am very thankful.

I could not have asked for a more supportive group of fellow graduate students. My first year in graduate school was made bearable by the near constant presence of Stephanie Owen and Mattan Alalouf. We spent countless hours together studying and working on problem sets during that first year. Their friendship and encouragement have continued long past prelims, and I am immeasurably grateful to have them in my life.

I spent many hours laughing with Jamie Fogel, Thomas Helgerman, and Andrew Simon. I will miss my office mates Guangye Cao, Wenting Song, and Matt Wilson, long lunches with George Fenton, and eating latkes with Matt Gross and Max Gross. For their support, community, and advice, I thank Will Boning, Ben Glass, Giacomo Brusco, Sam Haltenhof, Steve Hamilton, Molly Hawkins, Yeliz Kaçamak, Steph Karol, Meera Mahadevan, Jenny Mayo, Paul Organ, Dhiren Patki, Mike Ricks, Max Risch, Evelyn Smith, Teju Velayudhan, and Eleanor Wilking.

I was lucky to have friends in Ann Arbor outside of the Economics department. It was a great joy to be able to spend a few years living down the street from my childhood best friend, Kate Satterfield, and her (now) husband, Ian Campbell. I enjoyed many long and delicious meals with Rachel Hildrich and Claire Roberts-Thomson. I could not have asked for better people to swap stories about early parenthood with than Aaron Willis and Rachel Goldstein; it has been a privilege to follow Lila's journey from Day 1. A special shout-out to Aishat Olatunde and Michael Woodall for their generosity and practical medical advice.

I received incredible support from my community outside of Ann Arbor. A big thanks to Tom McGahee for always believing in me and wanting the best for me (and for sending puppy videos when I needed them most: prelim studying.) My 1044 family helped keep me tethered to the world outside of Lorch Hall, and I can't wait until we can all be together again. I am especially grateful for Aaron Bauer and Jake Levinson, whose friendship I have leaned on heavily for over 10 years, and whom I continue to turn to in good times and bad.

My family has been supportive throughout this whole process. I am unbelievably fortunate to like and get along with my parents enough that moving in with them during the COVID-19 pandemic was a viable option when we needed additional child care help. I am not sure what we would have done if we hadn't spent our quarantine in Oshkosh. I'm grateful to my brother, Nick, for always being there for me when I need a sounding board, and to my uncle, Todd, for answering all of my big and small questions about applying to graduate school, being a graduate student, and navigating the job market. I am pretty sure I have the best in-laws in Judy Kay and Bob Kummerfeld, who (among other amazing qualities) always ask how my research is going and are genuinely interested in the response.

Finally, thank you to my husband, Jonathan, and my son, Joseph, who have seen me through some very good days and some very bad days and loved me just as much. Wherever you are is home.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

This dissertation uses administrative tax data to study behavioral responses to taxation and tax enforcement. Chapter I focuses on the two tax penalties associated with Individual Retirement Accounts (IRAs). Chapter II considers the impact of a temporary suspension of tax collection efforts on future tax compliance and income. Chapter III investigates the use of professional tax preparation services by the top 1% of the income distribution.

Chapter I, which is co-authored with Victoria Bryant, focuses on tax-benefited retirement savings accounts. These accounts have features designed to encourage retirement savings, including a penalty for withdrawing before age $59\frac{1}{2}$. Account holders also face a penalty for failing to take required minimum withdrawals after age 72. Using a bunching analysis, we estimate that these penalties cause more than 17% of traditional IRA holders to change their withdrawal timing each year, shifting close to $60 billion of distributions annually. We estimate a dynamic life-cycle model and run counterfactual policy analysis to analyze the effect of changing these penalties. For both penalties, we find alternative combinations of age threshold and penalty rate that lead to increased average welfare and lifetime tax remittances: increasing the age threshold for penalty-free withdrawals while simultaneously lowering the penalty rate, and increasing the age threshold for required withdrawals while leaving the penalty rate unchanged.

In chapter II, which is co-authored with William C. Boning, Joel Slemrod, and Alex Turk, we ask whether a temporary suspension of efforts to collect outstanding tax debt ultimately lead to lower or higher tax compliance and income. When economic hardship prevents a tax debtor from paying basic living expenses, the Internal Revenue Service puts debt collection efforts on hold and designates the debt currently not collectible (CNC). This paper uses the quasi-random assignment of IRS Revenue Officers to tax debtors' cases as an instrumental variable to identify the causal effects of suspending debt collection on tax compliance and future income. In contrast to uninstrumented estimates, we find no evidence that putting off attempts to collect debt reduces compliance with future tax obligations or future reported income. Among marginal hardship cases, pausing collection instead increases future income, specifically wages earned by the taxpayer's spouse.

Chapter III, which is co-authored with Giacomo Brusco, Yeliz Kaçamak, and Mark Payne,

considers a recent addition to the conversation about income inequality and the "top 1%:" the extent to which this population avoids and evades its tax liability. As the vast majority of the top 1% of the income distribution use a paid tax preparer, understanding the role of professional tax preparation services among this population is critically important to understanding their tax outcomes. We find that, among the bottom 99% of the income distribution, self-prepared returns have smaller corrections from a random audit relative to paid-prepared returns, but among the top 1%, paid-prepared returns experience smaller corrections relative to self-prepared returns. We then turn to the question of why someone in the top 1% wouldn't use paid tax preparation. We conclude that individuals in the top 1% who are either an accountant or financial advisor, or who have a higher proportion of their income from third-party reported sources such as wages, are less likely to use paid tax preparation services.

CHAPTER I

The Impact of Withdrawal Penalties on Retirement Savings (with Victoria Bryant)

1.1 Introduction

There is growing concern about the state of retirement security in the United States. In 2015, the Government Accountability Office found that a third of U.S. households with the head of house aged 55 or older had no retirement income other than Social Security, and a quarter had no retirement income other than Social Security and a pension (Government Accountability Office, 2015). For households with additional retirement savings, the median account balance translated into a monthly annuity of just $400.

One way the U.S. government encourages individuals to improve their retirement finances is by offering tax benefits to certain types of retirement savings accounts, including Individual Retirement Accounts (IRAs). IRAs are subject to two tax penalties. To discourage premature spending out of these accounts, withdrawals before age $59\frac{1}{2}$ are penalized with an additional 10% tax (the "early withdrawal penalty"). To minimize the tax cost of IRAs, account holders are required to make annual minimum withdrawals beginning at age $70\frac{1}{2}$.[1] Failure to take these required withdrawals results in an additional 50% tax on the amount not withdrawn (the "excess accumulation penalty"). Although these penalties have existed for decades, their effect on retirement savings, welfare, and tax remittances is not well understood.

This paper makes two contributions. First, we use reduced-form bunching methods to estimate how many people change the timing of withdrawals from traditional IRAs in response to the two withdrawal penalties. Our analysis suggests that, each year, 1.4% of traditional IRA holders change the timing of their withdrawals in response to the early withdrawal penalty, while 16.2% of traditional IRA holders take withdrawals when they otherwise would

[1] In 2019, the age for annual minimum withdrawals was raised to 72.

not have in response to the excess accumulation penalty. These responses shift the timing of nearly \$60 billion worth of withdrawals from traditional IRAs each year.

Second, we estimate a dynamic life-cycle model, which provides a framework for analyzing the effect of changing the early withdrawal and excess accumulation penalties on welfare and tax remittances. The estimation routine yields estimates of four preference parameters: the elasticity of intertemporal substitution (EIS), the discount factor, and two bequest motive parameters. Identification of the preference parameters is driven by exogenous budget set discontinuities generated by the two tax penalties.

For both penalties, we identify combinations of the age threshold and penalty rate that lead to increased average welfare and tax remittances relative to the base policy. For the early withdrawal penalty, these combinations involve increasing the age threshold from $59\frac{1}{2}$ to $62\frac{1}{2}$ through $65\frac{1}{2}$, and lowering the penalty rate from 10% to 5%. These alternative combinations also yield higher IRA balances at age $65\frac{1}{2}$, furthering the purported aim of the early withdrawal penalty: to increase retirement savings. The intuition for this finding is as follows: conditional on having these penalties in place, there are benefits to encouraging taxpayers to keep their money in these accounts as long as possible. However, if individuals need to take early distributions as the result of an unexpected income shock, they can do so with minimal sanction.

For the excess accumulation penalty, we find that increasing the age for required withdrawals from $70\frac{1}{2}$ to $71\frac{1}{2}$ through $73\frac{1}{2}$, while leaving the 50% penalty rate unchanged, increases both welfare and the present discounted value of lifetime income tax remittances. While there is a trade-off between allowing account holders to keep their money in the tax-benefited account longer and the governments receipt of the tax revenue sooner, our results suggest that the delay in receiving the tax revenue is worth the increase in income tax remittances that come from higher account balances.

Understanding the role the two withdrawal penalties play in savings behavior is increasingly urgent. More than half of U.S. households held at least one IRA, 401(k), or other defined-contribution account by the end of 2017, with \$17 trillion of assets saved (Investment Company Institute, 2019). Employers have shifted to offering 401(k)s and other defined-contribution accounts instead of defined-benefit accounts such as pensions, and often auto-enroll their employees. States and localities have begun establishing "Auto-IRA" programs in which workers not eligible for other retirement plans are automatically enrolled in an IRA. Contemporaneous to the growing prevalence of these accounts, policy-makers are actively changing the rules around these penalties in response to mounting unease about retirement security.

Previous work studying behavioral responses to these penalties has generally focused on one of the two penalties in isolation, and has not been able to address welfare considerations. In contrast, we consider both penalties simultaneously. This allows us to study the relative magnitudes of their impacts on the timing of withdrawals, and means that our model more accurately captures the trade-offs considered by individuals when they decide how much to contribute or withdraw from these accounts. In addition, our structural model allows us to explicitly consider how changing either penalty would impact welfare, tax remittances, and IRA balances.

We also add to the nascent literature using bunching moments to identify structural parameters via dynamic models. Using bunching moments to identify structural models is a recent methodological development. Our setting differs from previous work using bunching moments to estimate the EIS (Best et al., 2019; Choukhmane, 2021). Rather than making a one-time refinancing decision (as in Best et al. (2019)), our individuals make the decision to withdraw, or not, each period. This is similar to Einav et al. (2015), in which individuals choose whether or not to fill a prescription each week based on a health event shock. We estimate an EIS of 1.061. Our EIS estimate is consistent with the argument that, in general, wealthier individuals have a higher EIS (Guvenen, 2006).

1.2 Individual Retirement Accounts

Individual Retirement Accounts (IRAs) are tax-benefited personal savings accounts. IRAs are defined-contribution accounts, meaning that contributions are made into individual accounts. A third of U.S. taxpayers hold at least one IRA.

We focus on traditional IRAs, which comprise about three-quarters of all IRAs. Contributions to traditional IRAs are deducted from taxable income, and withdrawals from traditional IRAs are treated as taxable income at the time of withdrawal.[2] Throughout the paper, we will use the terms withdrawal and distributions to discuss "normal" withdrawals (i.e., we do not include withdrawals due to rollovers or Roth conversions, or the death of the account holder) unless otherwise specified.[3]

Traditional IRAs were introduced as part of the Employee Retirement Income Security Act of 1974 to stimulate retirement saving, particularly among individuals who otherwise would reach retirement with little saved. The key tax benefit is that returns to the principal

[2] In contrast, contributions to Roth IRAs are included in taxable income, but withdrawals from Roth accounts are not subject to federal income taxation. See Appendix A.1.2 for more details on Roth IRAs.

[3] Funds can be "rolled over" from a different defined-contribution retirement savings account into an IRA. For example, an individual may rollover funds from a 401(k) into an IRA, or from one IRA into another (essentially combining their separate IRA accounts). The vast majority of IRA assets come from rolled over defined-contribution accounts (Goodman et al., 2019). See Appendix A.1.6 for more details about what happens when an IRA holder dies.

are not separately taxed. There is no annual tax on accrued interest, and no capital gains tax on withdrawals. Instead, returns are treated as regular taxable income upon withdrawal.[4] There are two additional tax benefits. First, because contributions to a traditional IRA are deducted from taxable income, an account holder's taxable income is lower in a year in which they make a contribution. Second, the amount of income tax ultimately due on contributions to a traditional IRA may be lower if the subsequent withdrawal is taken in a year when the account holder faces a lower marginal tax rate than in the year when she made the contribution. These tax benefits provide a large incentive to contribute to traditional IRAs.

To discourage withdrawing from IRAs before retirement, "early" withdrawals are penalized with an additional 10% tax. This penalty, coupled with a contribution limit ($5,500 in 2015 for individuals under age 50), ensures that the tax benefits of IRAs do not extend to all savings (particularly non-retirement savings).[5] Borrowing from IRAs is not permitted.

To limit the benefits accrued by individuals who would have had sufficient retirement savings without IRAs, a second penalty was introduced: the "excess accumulation penalty." At a certain age, IRA holders are required to take a minimum withdrawal each year. If an IRA holder does not take the required withdrawal, she will owe an additional 50% tax on the difference between the required amount and the withdrawn amount. These required withdrawals, known as Required Minimum Distributions (RMDs), were imposed for IRAs as part of the Tax Reform Act of 1986 (TIGTA, 2015-10-042; Mortenson et al., 2019). In addition to ensuring that IRAs do not only amount to a generous tax break (or a tax-benefited bequest), RMDs also minimize the tax revenue cost of IRAs and other defined-contribution retirement savings plans.

We give additional information about the early withdrawal and excess accumulation penalties below. Additional institutional details about traditional IRAs are provided in Appendix A.1.

1.2.1 The early withdrawal penalty

Withdrawals from traditional IRAs made before age $59\frac{1}{2}$ face a penalty of 10% on the full amount withdrawn in addition to income tax. For example, if an individual's marginal tax rate was 15% and she made an early withdrawal of $1,000 from a traditional IRA, she would owe both $150 in income tax as well as a $100 penalty (10% of the amount withdrawn early). In 2016, an estimated 1.2 million individual tax returns reported over $1.5 billion in penalties due because of early withdrawals from retirement accounts (Internal Revenue

[4]See Appendix A.1.1 for a simple algebraic example of this benefit.
[5]See Appendix A.1.4 for more details about IRA contribution limits.

Service, 2016). Goodman et al. (2019) estimate that 20% of individual contributions made to IRAs and 401(k)s by individuals below the age of 48 leave the accounts within 8 years of the contribution.

In some circumstances, early withdrawals from an IRA are exempt from the penalty. IRA holders may receive an exemption to pay for qualified medical or higher education expenses. First-time home-buyers may take a withdrawal of up to $10,000 from an IRA without penalty. Individuals who agree to take "substantially equal" withdrawals for a fixed period of time of at least five years may also be exempt from the penalty. Argento et al. (2015) report that, for individuals under age 55, 21% of withdrawals from retirement accounts were penalized.

Previous work on early withdrawals has focused on the circumstances that lead account holders to incur the penalty. This research has consistently found that penalized withdrawals are substantially more likely for individuals that experience adverse shocks such as job loss and divorce, especially among individuals with lower levels of non-retirement assets (see, e.g., Amromin and Smith (2003), Butrica et al. (2010), and Argento et al. (2015)). One recent attempt to quantify the causal effect of these penalties is Goda et al. (2018), who use exact timing of birth to estimate how withdrawals from IRAs respond to the early withdrawal penalty. They find that average withdrawals increase approximately 80% after account holders cross the $59\frac{1}{2}$ threshold, and that the response is largely due to first-time withdrawals.

1.2.2 The excess accumulation penalty and Required Minimum Distributions

Starting in the tax year that a traditional IRA holder turns a certain age, she is required to take a Required Minimum Distribution (RMD) every year. Because we consider a period before 2019 in our empirical analysis, we use $70\frac{1}{2}$ as the base policy age threshold for RMDs and the excess accumulation penalty throughout the paper.[6] The first RMD tax payment is due by April 1 of the calendar year following the year in which the account holder turns $70\frac{1}{2}$ (i.e., the calendar year in which they turn $71\frac{1}{2}$); subsequent payments must be made by December 31 of each calendar year.[7] More than 15,000 traditional IRA holders reported owing the 50% excess accumulation penalty in 2016, resulting in $7.3 million in additional payments to the IRS (Internal Revenue Service, 2016).

The amount of the RMD is based on the balance of the account on December 31 of the previous tax year and life expectancy tables (see Appendix A.1.5). The penalty for failing to take a required withdrawal is called the "excess accumulation penalty" and is equal to 50% of

[6] The Setting Every Community Up for Retirement Enhancement (SECURE) Act of 2019 raised the age at which traditional IRA holders are subject to RMDs from $70\frac{1}{2}$ to 72.

[7] If a traditional IRA holder waits to submit her first RMD payment until April 1 of the calendar year after the year in which she turns $70\frac{1}{2}$, she will have two RMDs included in her taxable income in the first year she makes an RMD payment.

the required amount not withdrawn. Consider again our individual with a marginal income tax rate of 15%. If she failed to take a required withdrawal of $1,000 from a traditional IRA, she would owe $150 in income tax and a $500 penalty (50% of the money not withdrawn). We discuss how RMDs work in the event of the account holder's death in Appendix A.1.6.

The empirical literature on RMDs has centered around the "RMD Holiday" in 2009, which suspended the RMD rules for the 2009 tax year. Brown et al. (2017) find that a third of traditional IRA holders subject to RMD rules in 2008 did not take a withdrawal in 2009. Mortenson et al. (2019) use year-over-year variation in addition to the RMD Holiday and estimate that over half of traditional IRA holders are constrained by the RMD rule and would take less than their required withdrawal (or none at all) if not for the RMD rules.

The literature studying the role of these penalties has usually considered them separately. The exception is Sabelhaus (2000), who finds that while these policies change withdrawal timing, total projected taxable withdrawals do not change much if the age for penalty-free withdrawals is lowered to $55\frac{1}{2}$, or the age for required withdrawals is raised to $75\frac{1}{2}$. We take a substantively different approach by estimating a structural model, which allows us to focus on the welfare implications of changing the age thresholds in addition to the behavioral responses. We also consider changes to the penalty rates in addition to changing the age threshold.

1.3 Administrative tax data

We use de-identified administrative tax data from the Internal Revenue Service (IRS). We create a panel based on a 5% random sample of individuals with Social Security numbers and aged 18 or older in 1999. We follow these individuals through 2015. The 17-year panel is balanced apart from exit due to death and emigration.

We limit our sample to individuals who have an IRA account. We identify individuals as "IRA holders" if we ever observe them making a contribution to an IRA (including rollovers), taking a normal distribution from an IRA,[8] or having an outstanding IRA balance. Our final sample of IRA holders comprises 3,913,401 unique individuals. We focus on traditional IRA holders in our analysis. We observe that 72.2% of IRA accounts are traditional IRAs.[9]

We focus on IRAs because we can cleanly identify contributions to and withdrawals from

[8] Normal distributions do not include withdrawals due to rollovers or Roth conversions, or the death of the account holder.

[9] In our sample, 27.8% of IRAs are Roth IRAs. We find that 15.7% of IRA holders hold both kinds of accounts. There are two additional types of IRAs: SIMPLE IRAs and SEP IRAs. These plans make up a small fraction of IRA activity (see Appendix A.1.3). As these plans face the same tax benefits and penalties as traditional IRAs, we follow Mortenson et al. (2019) and count contributions toward (and withdrawals from) a SEP or SIMPLE IRA as contributions toward (and withdrawals from) a traditional IRA for our analysis.

IRAs in the administrative tax data, as well as the end-of-year fair market values of IRA accounts. In contrast, we do not separately observe 401(k)s and other employer-sponsored defined-contribution accounts which are subject to the same penalties.[10] Of the $17 trillion dollars saved in IRAs or other defined-contribution accounts in 2017, $9.2 trillion (54%) was in IRAs (Investment Company Institute, 2019). There are two facts that suggest our results likely apply to holders of 401(k)s and other employer-sponsored defined-contribution accounts. First, of households that have an IRA, over 80% also have an employer-sponsored retirement plan.[11] Unless households that only have a 401(k) are systematically different from those that have both an IRA and an employer-sponsored account, we will capture behavior from many 401(k) holders in our analysis. Second, many 401(k)s are ultimately rolled over into IRAs when individuals separate from the employers who offered the 401(k).

We supplement our IRA data with information from individual income tax returns such as adjusted gross income, wage and self-employment income, and other sources of retirement income such as Social Security and employer-sponsored defined benefit plans (such as pensions) and defined-contribution plans (such as 401(k)s). We also include data on additional taxes owed due to early withdrawals, non-qualified withdrawals, or failing to take a minimum required withdrawal. We discuss in detail which IRS forms we use, the relevant sample restrictions, and the construction of key variables in Appendix A.2.

1.4 Bunching evidence and reduced-form estimates

Measuring bunching around kinks and notches in budget sets has become an increasingly popular tool for estimating behavioral responses to various incentives.[12] This approach was developed by Chetty et al. (2011) for budget set kinks and extended to notches by Kleven and Waseem (2013).

In this section, we adapt the standard approach in order to estimate the number of traditional IRA holders who changed the timing of their withdrawals in response to these penalties, and the amount of money shifted. We estimate four outcomes: the number of individuals who shift the timing of their first withdrawal from a traditional IRA (and the amount of money impacted), and the number of individuals who shift the timing of any

[10] While information about 401(k)s is captured in survey data, there are limitations to how that data could be used. IRAs and employer-sponsored retirement savings accounts are vastly underreported in both the Survey of Income and Program Participation (SIPP) and the Current Population Survey (CPS) when compared to IRS records (Bee and Mitchell, 2017). Larrimore et al. (2019) show that, in 2010, the CPS was missing half a billion dollars in income from pensions, annuities, and IRA withdrawals.

[11] Authors' calculations using Figure 8.7 of Investment Company Institute (2019). Note that, in this figure, employer-sponsored retirement plans include both defined-contribution and defined-benefit plans.

[12] Kleven (2016) provides a recent review of this literature.

withdrawal from a traditional IRA (and the amount of money impacted).

1.4.1 Reduced-form estimation strategy

We fit a flexible polynomial to the empirical distributions of our outcomes. This polynomial excludes data in a region around the age thresholds (the "excluded ranges"), as shown in shown in Equation 1.4.1:

$$c_j = \sum_{i=0}^{p} \beta_i (a_j)^i + \sum_{k=a_{59.5,-}}^{a_{59.5,+}} \gamma_k \cdot \mathbb{I}[a_k = a_j] + \sum_{k=a_{70.5,-}}^{a_{70.5,+}} \delta_k \cdot \mathbb{I}[a_k = a_j] + \upsilon_j, \tag{1.4.1}$$

where c_j is the value of the outcome at age j, $[a_{59.5,-}, a_{59.5+}]$ is the excluded region around the early withdrawal penalty, $[a_{70.5,-}, a_{70.5+}]$ is the excluded range around the excess accumulation penalty, and p is the order of the polynomial. The β coefficients capture the distribution without the withdrawal penalties (the "counterfactual" distribution), while the γ and δ coefficients represent the difference from the counterfactual distribution brought on by the withdrawal penalties.

We estimate the counterfactual distribution of the outcome using the predicted values from Equation 1.4.1, leaving out the portion from the middle terms:[13]

$$\widehat{c}_j = \sum_{i=0}^{p} \widehat{\beta}_i (a_j)^i \tag{1.4.2}$$

To determine the excluded range for the early withdrawal penalty, we iterate over all possible combinations of $a_{59.5,-}$ and $a_{59.5+}$ to minimize Equation 1.4.3:[14,15]

$$\{a_{59.5,-}, a_{59.5,+}\} = \arg\min \left| \widehat{B}_{59.5} - \widehat{M}_{59.5} \right| \tag{1.4.3}$$

$$\widehat{B}_{59.5} = \sum_{j=59.5}^{a_{59.5,+}} (c_j - \widehat{c}_j) \quad \widehat{M}_{59.5} = \sum_{j=a_{59.5,-}}^{58.5} (\widehat{c}_j - c_j)$$

Throughout this section, we will refer to $\widehat{B}_{59.5}$ as the "excess mass" or the "bunching mass," and $\widehat{M}_{59.5}$ as the "missing mass." The excess mass is defined as the difference between the

[13] We do not allow the counterfactual levels to drop below 0.

[14] This method differs slightly from Kleven and Waseem (2013), who visually determine the equivalent of $a_{59.5+}$ and then iterate over $a_{59.5,-}$ to minimize Equation 1.4.3. We chose our method because there was not an obvious visual cue for $a_{59.5+}$. Both approaches assume no extensive margin responses.

[15] All possible values of $a_{59.5,-}$ include $25\frac{1}{2}$ to $58\frac{1}{2}$. All possible values of $a_{59.5,+}$ include $59\frac{1}{2}$ to $69\frac{1}{2}$. Because our bin size is fixed to 1, we are not able to exactly obtain $\widehat{B}_{59.5} = \widehat{M}_{59.5}$. Appendix A.4 shows the value of Equation 1.4.3 for all combinations of $a_{59.5,-}$ and $a_{59.5+}$.

empirical and the counterfactual distributions on the low-tax side of the age threshold (in our case, the portion of the excluded range to the right of the age threshold for the early withdrawal penalty). Similarly, the missing mass is the difference between the empirical and the counterfactual distributions on the high-tax side of the age threshold.

The excess accumulation penalty is unusual as a notch because all traditional IRA holders must take withdrawals out at age $70\frac{1}{2}$. This means that, if there are individuals who would rather take their first withdrawal after age $70\frac{1}{2}$, we should see a spike at age $70\frac{1}{2}$ as the age of first withdrawal (which we do), and little to no observations after that (which we also do). This is because we would expect individuals who would rather take the first withdrawal after age $70\frac{1}{2}$ to push the timing of their first withdrawal as late as possible without being penalized. Unlike the case of standard notches, where individuals have an incentive not to cross the threshold, most individuals in our setting will cross the threshold eventually (unless the individual dies before age $70\frac{1}{2}$ or withdraws all of the funds in their IRA before age $70\frac{1}{2}$). The individual has no incentive to deviate from their optimal behavior before age $70\frac{1}{2}$. We can therefore assume that traditional IRA holders who take their first withdrawal in the few years preceding $70\frac{1}{2}$ are behaving optimally. As a result, we simply call age $70\frac{1}{2}$ the excluded range for the excess accumulation penalty.[16] The "excess mass" associated with the excess accumulation penalty, shown in Equation 1.4.4, is calculated similarly to that for the early withdrawal penalty, except that we do not estimate an upper bound for the excluded region:[17]

$$\widehat{B}_{70.5} = c_{70.5} - \widehat{c}_{70.5} \tag{1.4.4}$$

We estimate the magnitude of the bunching responses using Equation 1.4.5, which considers the difference in the cumulative densities above the early withdrawal threshold (below the excess accumulation threshold) compared to the total quantity of the outcome in

[16] In other words, $a_{70.5,-} = a_{70.5,+} = 70\frac{1}{2}$.

[17] The spike in the number of individuals who take their first withdrawal at $70\frac{1}{2}$ (and the corresponding total value of those withdrawals) is so large that we are unable to fully distribute that amount under our counterfactual distribution. For the number of individuals withdrawing, the percentage of traditional IRA holders taking their first withdrawal at or above age $70\frac{1}{2}$ that we are not able to account for with the counterfactual distribution is 38.3%. Similarly, the percentage of total first withdrawals taken at or above age $70\frac{1}{2}$ is 3.6%. We believe the tail of the true counterfactual distribution actually extends well past age $80\frac{1}{2}$. We observe age of first withdrawal for Roth IRA accounts well into the 90s. This is consistent with the idea that traditional IRA holders who are forced to take withdrawals from their IRAs because of the RMD rules would otherwise hold on to those funds, allowing them to enjoy more years of tax-benefited growth, until they would then be included in a bequest. See Appendix A.1.6 for more information about the rules around inheriting an IRA.

our sample, N:

$$mag_j = \frac{\widehat{B}_j}{N}, \quad j \in \{59.5, 70.5\} \tag{1.4.5}$$

mag_j is an estimate for the percentage of the outcome impacted by these thresholds.[18]

We present two pieces of evidence that some traditional IRA holders are responding to these age thresholds. First, we show the distribution of the age at which traditional IRA holders took their first withdrawal, and the total amount withdrawn at each age of first withdrawal. Second, we present the proportion of traditional IRA holders taking withdrawals at each age, and the total amount withdrawn at each age, in a single year. We focus on traditional IRA holders for two reasons: the majority of IRAs are traditional IRAs, and because the tax-benefited in our structural model is designed as a traditional IRA. We present the equivalent figures for Roth accounts in Appendix A.3.2.

1.4.2 Bunching by age of first withdrawal

We define the age of first withdrawal as the age at which we first observe an individual making a withdrawal, excluding early withdrawals (before age $59\frac{1}{2}$) with known, qualifying exceptions. This definition is an upper bound on the actual first year of withdrawals, because some individuals may have begun taking withdrawals before our sample period. For this analysis, we exclude individuals who are older than 65 in the first year of our sample so that we do not overestimate the number of individuals taking their "first" withdrawal over age $70\frac{1}{2}$. We also exclude data from 2009. Traditional IRA holders were not subject to the same age thresholds in 2009 as in all other years in our sample as a result of the one-year suspension of the RMD rules included in the Worker, Retiree, and Employer Recovery Act of 2008.

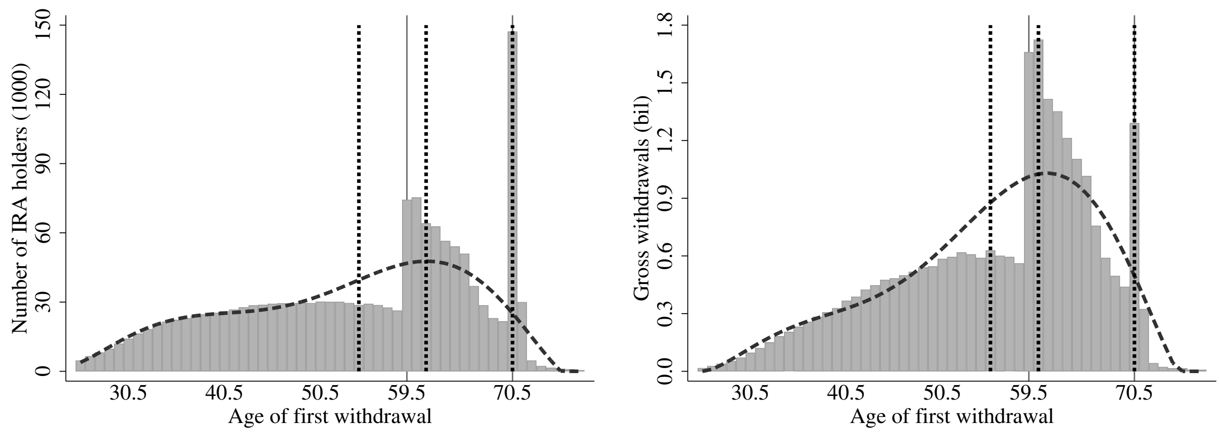
We consider two outcomes by age of first withdrawal: the number of individuals who took their first observed withdrawal at each age, and the amount of money distributed at first withdrawal at each age. Figure 1.1 shows the empirical distribution (as bars) and our estimated counterfactuals from Equation 1.4.1 (as dotted lines) for both outcomes. Figure 1.1a shows the number of individuals taking their first withdrawal at each age in our sample, while Figure 1.1b shows the total amount withdrawn by those individuals. In each panel of Figure 1.1, the two dotted vertical lines show our estimated values for $a_{59.5,-}$ and $a_{59.5,+}$. For these outcomes, we use $p = 6$ to fit the counterfactual.

It is clear in Figure 1.1 that traditional IRA holders change their behavior in response

[18] For example, consider the question of "what percentage of IRA holders changing the timing of their withdrawals?" If we estimated that $\widehat{B}_j = 100$ individuals, and there were $N = 1{,}000$ individuals in the relevant analysis, we would conclude that 10% of individuals were altering their withdrawal timing.

Figure 1.1: Bunching by age of first withdrawal

(a) Number of IRA holders (b) Gross withdrawal amounts ($bil)



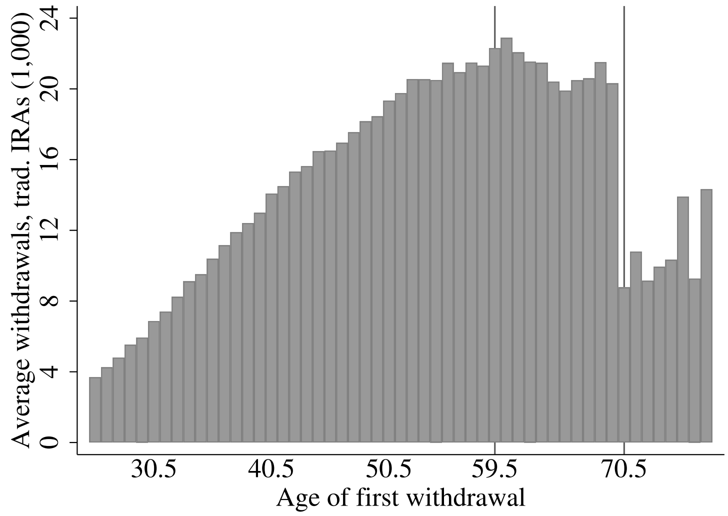
Notes: Figure 1.1a: $N = 1{,}536{,}392$ unique individuals. Figure 1.1b based on $25,967,606,263 total withdrawals, inflated to 2015 values. Excludes individuals older than age 65 at the beginning of our sample period, and all data from 2009. We define the age of first withdrawal as the age at which we first observe an individual making a withdrawal, excluding early withdrawals with known, qualifying exceptions. This definition is an upper bound on the actual first year of withdrawals, because some individuals may have begun taking withdrawals before our sample period. Figure 1.1a shows the number of individuals in our sample we observe taking their first withdrawal from a traditional IRA at each age and the estimated counterfactual distribution. Figure 1.1b shows the total amount distributed in the first withdrawal by traditional IRA holders who took their first withdrawal at each age and the estimated counterfactual distribution. Withdrawal amounts are inflated to 2015 values. For both figures, the bars show the empirical distribution, and the the dotted lines represent our estimated counterfactual distribution.

to these penalties. There is a dramatic increase in the number of individuals taking their first withdrawal at both age thresholds, and in the amount of money withdrawn in first distributions. The number taking their first withdrawal at age $70\frac{1}{2}$ is significantly higher than the number taking their first withdrawal at age $59\frac{1}{2}$, but the total amount withdrawn at age $59\frac{1}{2}$ is greater than the total amount withdrawn at age $70\frac{1}{2}$. In other words, the average first withdrawal amount for those taking their first withdrawal at age $70\frac{1}{2}$ is considerably less than for those taking their first withdrawal at age $59\frac{1}{2}$, as shown in Figure 1.2.

There are a small number of individuals whom we observe taking their first withdrawal after age $70\frac{1}{2}$. The vast majority of these cases show the individual taking the withdrawal at age $71\frac{1}{2}$. This is not surprising, given that the first RMD payment is actually due to the IRS by April 1 of the calendar year after the year in which the individual turns $70\frac{1}{2}$ (that is, the calendar year in which they turn $71\frac{1}{2}$).[19] There are two likely explanations for the small number of traditional IRA holders we observe taking their first withdrawal after age $71\frac{1}{2}$. These individuals may have rolled a 401(k) into an IRA after age $70\frac{1}{2}$. These individuals would not necessarily have held an IRA before the rollover. Another possibility is

[19]See Section 1.2.2 for more details.

Figure 1.2: Average value of first withdrawals from traditional IRAs ($1,000)



Notes: Based on 1,536,392 unique individuals and $25,967,606,263 total withdrawals, inflated to 2015 values. Excludes individuals older than age 65 at the beginning of our sample period, and all data from 2009. We define the age of first withdrawal as the age at which we first observe an individual making a withdrawal, excluding early withdrawals (before age $59\frac{1}{2}$) with known, qualifying exceptions. This definition is an upper bound on the actual first year of withdrawals, because some individuals may have begun taking withdrawals before our sample period. Figure shows the average amount withdrawn in the first withdrawal from a traditional IRA by c who took their first withdrawal at each age. Withdrawal amounts are inflated to 2015 values.

that these traditional IRA holders may not have fully understood the RMD rules. While these individuals should in theory remit the excess accumulation penalty, the IRS may waive the penalty if the individual can prove that they did not make the required payment as a result of reasonable error and undertook steps to correct their mistake. Mortenson et al. (2019) report that only 2 to 3% of individuals who fail to make an RMD payment violate the RMD rules the following year.

Our estimates for $a_{59.5,-}$, $a_{59.5,+}$, $\widehat{B}_{59.5}$, and $\widehat{B}_{70.5}$, as well as bootstrapped standard errors, are given in Table 1.1.[20] Table 1.1 also includes estimates of $mag_{59.5}$, $mag_{70.5}$, and the implied

[20]Standard errors for $a_{59.5,-}$, $a_{59.5,+}$, $\widehat{B}_{59.5}$, and $\widehat{B}_{70.5}$ are calculated using a bootstrap procedure (see Chetty et al. (2011) and Kleven and Waseem (2013)). We generate alternative data by randomly sampling (with replacement) the residuals produced by Equation 1.4.1 and adding those on to the predicted values for each age. Although we force our counterfactual values to be at least 0, we keep the original residual terms from Equation 1.4.1. Unlike when calculating the counterfactual distribution, the predicted values in this case are $\widehat{c}_{j,bs} = \sum_{i=0}^{p} \widehat{\beta}_i (a_j)^i + \sum_{i=a_{59.5,-}}^{a_{59.5,+}} \widehat{\gamma}_i \cdot \mathbb{I}[a = i] + \sum_{i=a_{70.5,-}}^{a_{70.5,+}} \widehat{\delta}_i \cdot \mathbb{I}[a = i]$. We then re-calculate $a_{59.5,-}$, $a_{59.5,+}$, $\widehat{B}_{59.5}$, and $\widehat{B}_{70.5}$ using the simulated data as if it were our original data. We repeat this procedure 1,000 times to obtain a distribution for each estimated variable. We define the standard error for each variable as the standard deviation of the bootstrapped values of that variable. While one benefit of our setting is that we do not have to estimate bin size, the drawback is that we have very few bins (and even fewer after removing the estimated excluded region) and therefore few standard errors to sample for our bootstrap procedure. This means that there are a small number of residuals being used for the bootstrap procedure and, ultimately, to determine our standard errors.

Table 1.1: Changes in first withdrawals from traditional IRAs

	Number of taxpayers	Gross withdrawals
Parameter estimates		
N	1,536,392	\$25,967,606,263
$a_{59.5,-}$	54.5	55.5
	(6.2)	(6.1)
$a_{59.5,+}$	61.5	60.5
	(2.8)	(2.8)
$\widehat{B}_{59.5}$	71,962	\$1,345,804,160
	(20,904)	(467,059,441)
$\widehat{B}_{70.5}$	122,088	\$794,694,208
	(5,674)	(139,377,942)
Magnitude estimates		
U.S. total in 2015	201 million	\$277 billion
Proportion holding a traditional IRA	23.6%	n/a
First withdrawal proportion	4.5%	34.7%
Relevant population for scaling	2.1 million	\$94.5 billion
In response to early withdrawal penalty:		
$mag_{59.5}$	4.7%	5.2%
Scaled to U.S., annual	99,566	\$4.9 billion
In response to excess accumulation penalty:		
$mag_{70.5}$	7.9%	3.1%
Scaled to U.S., annual	168,920	\$2.9 billion

Notes: Bootstrapped standard errors given in parentheses. Our preferred estimate of $a_{59.5,+}$ minimize the difference between $\widehat{B}_{59.5}$ and $\widehat{M}_{59.5}$ given $a_{59.5,+} = 49\frac{1}{2}$ (see Appendix A.4). Scaled amounts are calculated by multiplying the relevant U.S. population in 2015 by the appropriate magnitude estimate. Estimate of number of taxpayers in the U.S. in 2015 based on authors' internal calculations. We multiply this value by the percentage of our sample that hold traditional IRAs to estimate the number of individuals with traditional IRAs, and then by the average proportion of traditional IRA holders that take their first withdrawal each year. Total amount withdrawn from traditional IRAs in 2015 taken from Statistics of Incomes Tax Stats: Accumulation and Distribution of Individual Retirement Arrangements (IRA) Table 1 (2015), available at https://www.irs.gov/pub/irs-soi/15in01ira.xls. We multiply this value by the average amount of withdrawals from traditional IRAs that are part of first withdrawals.

number of individuals and amount of money impacted by these age thresholds when scaled to the full population of taxpayers of the United States in 2015.

When considering the number of individuals that change when they take their first withdrawal from an IRA, we find $a_{59.5,-} = 54.5$ and $a_{59.5,+} = 61.5$.[21] We estimate that 4.7% of the individuals included in the analysis changed when they took their first withdrawal in response to the early withdrawal penalty, and 7.9% changed when they took their first

[21]The difference between $\widehat{B}_{59.5}$ and $\widehat{M}_{59.5}$ equals 1,412 individuals at these values of $a_{59.5,-}$ and $a_{59.5,-}$.

withdrawal in response to the excess accumulation penalty. These estimates translate into approximately 99,600 and 168,900 individuals changing when they take their first withdrawal due to these penalties each year.

We find similar values for $a_{59.5,-}$ and $a_{59.5,+}$ when we look at the total amount of withdrawn by age of first withdrawal (55.5 and 61.5, respectively). We have more faith in our value for $a_{59.5,-}$ based on the number of individuals rather than the amount of money, because we are not able to identify those exact individuals changing their behavior around first withdrawal (and therefore cannot know precisely which withdrawals were shifted). As this excluded range is smaller than that estimated for the number of individuals who shifted their behavior, the results from this exercise will be conservative relative to the exercise where we used the excluded range estimated for the number of individuals shifting the timing of their first withdrawal.

We estimate that 5.2% of gross first withdrawals are moved as a result of the early withdrawal penalty, and 3.1% of gross first withdrawals are moved as a result of the excess accumulation penalty. If we convert these percentages to annual dollar amounts, we find that \$4.9 billion worth of withdrawals is shifted up to seven years as a result of the early withdrawal penalty, and \$2.9 billion is withdrawn earlier than would have been without the excess accumulation penalty.

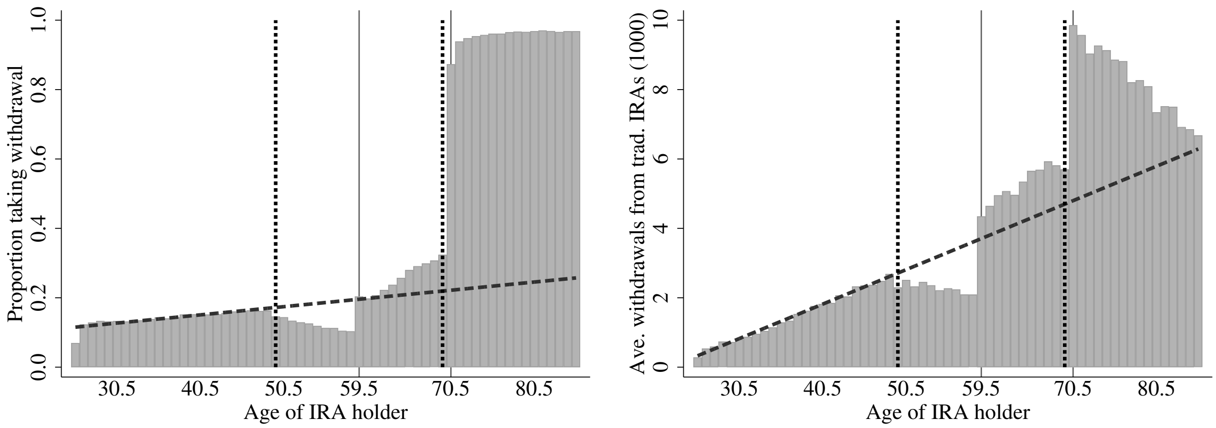
These magnitude estimates should be considered lower bounds on the number of individuals who change their behavior in response to these penalties, and the dollar amount of withdrawals shifted, for four reasons. First, we are focused on traditional IRAs. Roth IRA holders are also subject to the early withdrawal penalty for withdrawals larger than the size of the principal investment. Second, our analysis is focused on individuals who hold IRAs, but the same penalties exist for 401(ks) and other traditional defined-contribution retirement savings accounts. The early withdrawal penalty also applies to a particular type of defined-benefit account: cash balance plans. Third, we focus on the first withdrawal, whereas traditional IRA holders face these thresholds (and corresponding penalties) every age before (after) age $59\frac{1}{2}$ $(70\frac{1}{2})$. The timing of individuals' second, third, fourth, etc. withdrawals could also be impacted by these penalties. Finally, for both populations (IRA and 401(k) holders), these estimates do not take into account the extent to which they affect the extensive margin. In other words, there may be individuals who do not contribute to a retirement savings plan because of these age thresholds.

1.4.3 Bunching within a single year

In this section, we apply a similar methodology to estimate the impact these penalties have on the withdrawal behavior of traditional IRA holders in the cross-section. This allows

Figure 1.3: IRA withdrawal behavior in 2005

(a) Proportion of trad. IRA holders (b) Average withdrawal amount ($1,000)



Notes: Figure 1.3a: $N = 1{,}449{,}868$ unique individuals. Figure 1.3b: based on $6,483,125,796 total withdrawals, inflated to 2015 values. Excludes early withdrawals (before age $59\frac{1}{2}$) with a known, qualifying exception. Limited to individuals with a positive traditional IRA balance in 2004 or 2005. Figure 1.3a shows the proportion of account holders taking a withdrawal in 2005 and the estimated counterfactual distribution. Figure 1.3b shows the average amount withdrawn by age in 2005 and the estimated counterfactual distribution. Average withdrawal amounts are inflated to 2015 values. For both figures, the bars show the empirical distribution, and the the dotted lines represent our estimated counterfactual distribution.

us to estimate the magnitude of the effect on all traditional IRA holders, not just those considering their first withdrawal. We focus on a single year to ensure that, as in our previous analysis, each individual appears only once. We picked 2005 because the youngest individuals in our sample were 18 in 1999. We anticipate that the majority who attended college would have graduated within 6 years. We limit this analysis to individuals in our sample who had a non-zero balance in a traditional IRA at the end of 2004 or 2005.

We consider both the proportion of IRA holders who take a withdrawal as well as the average size of withdrawals. Figure 1.3 shows the empirical distribution and our estimated counterfactuals for both outcomes. The response at the age thresholds is striking. The percentage of account holders who take withdrawals jumps from 10.4% at age $58\frac{1}{2}$ to 20.4% at age $59\frac{1}{2}$ (an increase of 10.0 percentage points). Even more dramatic is the jump at age $70\frac{1}{2}$: the percentage of account holders who take withdrawals increases from from 32.4% at age $69\frac{1}{2}$ to 87.3% at age $70\frac{1}{2}$ (an increase of 54.9 percentage points).[22]

Our cross-sectional analysis differs from that in Section 1.4.2 in several respects. For

[22] There are two likely explanations for why the proportion of account holders taking withdrawals jump to exactly 100% at age $70\frac{1}{2}$. First, our data are pre-audit, and therefore some initial noncompliance could be embedded. Second, it's possible our definition of "who is an IRA holder in 2005" is over-inclusive, and therefore we are including more individuals in the denominator than we should. The withdrawal pattern we observe after age $70\frac{1}{2}$ is very similar to Figure 6 in Mortenson et al. (2019), including the fact that we don't observe 100% of traditional IRA holders taking a withdrawal after age $70\frac{1}{2}$.

these outcomes, we use a fitted first-degree polynomial ($p = 1$) rather than the sixth-order polynomial we used in Section 1.4.2. This is based on the observation that both trends in Figure 1.3 appear to be linear before they drop ahead of the early withdrawal penalty. We also exclude the region above $70\frac{1}{2}$ for these estimates. Finally, we visually determine the value of $a_{59.5,-}$ to be $49\frac{1}{2}$ and iterate over possible values of $a_{59.5,+}$ to minimize the difference between $\widehat{B}_{59.5}$ and $\widehat{M}_{59.5}$, rather than iterating over all possible combinations of $a_{59.5,-}$ and $a_{59.5,+}$ as we did in Section 1.4.2. We make this change because it is visually clear where the proportion of traditional IRA holders taking a withdrawal in the cross section begins to fall ahead of the early withdrawal penalty. The dotted vertical lines in Figure 1.3 show $a_{59.5,-}$ ($49\frac{1}{2}$ for this analysis) and our estimated values for $a_{59.5,+}$.

Our estimates for $a_{59.5,+}$ and scaled values for $\widehat{B}_{59.5}$ and $\widehat{B}_{70.5}$, as well as bootstrapped standard errors, are given in Table 1.2. We find that $a_{59.5,+} = 69\frac{1}{2}$ for both outcomes.[23] Table 1.2 also includes estimates of $mag_{59.5}$, $mag_{70.5}$, and the implied number of individuals and amount of money impacted by these age thresholds each year when scaled to the full population of taxpayers of the United States in 2015.

We estimate that, in a single year, 1.4% of our sample of traditional IRA holders changes the timing of their withdrawals in response to the early withdrawal penalty. This translates into approximately 648,400 traditional IRA holders in the U.S. changing their withdrawal behavior each year. This is considerably larger than our estimate of the number of individuals who change the timing of their first withdrawal each year (about 99,600), which underscores our point that our estimates based on the timing of first withdrawal are likely lower bounds. We estimate that approximately \$12.6 billion is not withdrawn from traditional IRAs each year as a result of the early withdrawal penalty.

Our estimates are even larger for the excess accumulation penalty. We find that 16.2% of our sample of traditional IRA holders change the timing of their withdrawals because of RMDs (about 7.7 million individuals, compared with the 168,900 estimated to change the timing of their first withdrawal each year). We estimate that about \$45.3 billion is withdrawn from traditional IRAs earlier than it would have been without the excess accumulation penalty.

The true counterfactual distribution over age $70\frac{1}{2}$ could be higher or lower than what we estimate. Our results for the excess accumulation penalty rely on a strong assumption: that the linear counterfactual distribution is correct past age $70\frac{1}{2}$. If traditional IRA holders become increasingly likely to take withdrawals as they age, our estimates would be lower than the true distribution (i.e., we would overstate the number of individuals impacted).

[23] Tables A.7 and A.8 in Appendix A.4 show the difference between $\widehat{B}_{59.5}$ and $\widehat{M}_{59.5}$ for every possible combination of $a_{59.5,-}$ and $a_{59.5,+}$.

Table 1.2: Changes in withdrawals from traditional IRAs in a single year

	Number of taxpayers	Gross withdrawals
Parameter estimates		
N	1,449,868	\$6,483,125,796
$a_{59.5,-}$	49.5	49.5
$a_{59.5,+}$	69.5	69.5
	(0.1)	(0.4)
$\widehat{B}_{59.5}$ (scaled by N)	19,746	\$295,103,584
	(2,035)	(29,284,054)
$\widehat{B}_{70.5}$ (scaled by N)	234,677	\$1,060,085,440
	(4,080)	(66,206,502)
Magnitude estimates		
U.S. total in 2015	201 million	\$277 billion
Proportion holding a traditional IRA	23.6%	n/a
Relevant population for scaling	47.6 million	\$277 billion
In response to early withdrawal penalty:		
$mag_{59.5}$	1.4%	4.6%
Scaled to U.S., annual	648,385	\$12.6 billion
In response to excess accumulation penalty:		
$mag_{70.5}$	16.2%	16.4%
Scaled to U.S., annual	7,706,011	\$45.3 billion

Notes: Bootstrapped standard errors given in parentheses. Our preferred estimate of $a_{59.5,+}$ minimize the difference between $\widehat{B}_{59.5}$ and $\widehat{M}_{59.5}$ given $a_{59.5,+} = 49\frac{1}{2}$ (see Appendix A.4). Scaled amounts are calculated by multiplying the relevant U.S. population in 2015 by the appropriate magnitude estimate. Estimate of number of taxpayers in the U.S. in 2015 based on authors' internal calculations. We multiply this value by the percentage of our sample that hold traditional IRAs to estimate the number of taxpayers with traditional IRAs. Total amount withdrawn from traditional IRAs in 2015 taken from Statistics of Incomes Tax Stats: Accumulation and Distribution of Individual Retirement Arrangements (IRA) Table 1 (2015), available at https://www.irs.gov/pub/irs-soi/15in01ira.xls.

There is evidence that this is true: the apparently linear trend from age $60\frac{1}{2}$ to $69\frac{1}{2}$ has a steeper slope than the linear trend below age $59\frac{1}{2}$. If some individuals who haven't taken a withdrawal by age, say, 75 would never take a withdrawal, the counterfactual distribution would flatten at some point and our counterfactual estimates would be higher than the true distribution (i.e., we would understate the number of individuals impacted).

There are other reasons to believe these estimates are lower bounds. We do not include Roth IRAs, which are also subject to the early withdrawal penalty,[24] or 401(k)s and other

[24]Figure A.5 in Appendix A.3.2 shows withdrawal behavior in 2005 for Roth IRAs. The proportion of Roth IRA holders taking a withdrawal at any age is always under 0.015, and drops over time. While we observe a small spike at age $59\frac{1}{2}$, the proportion of Roth IRA holders taking a withdrawal continues to fall after that.

traditional defined-contribution retirement savings accounts, which are subject to both of these penalties. As with our estimates based on the age of first observed withdrawal, these estimates do not take into account the extent to which they affect the extensive margin. In other words, there may be individuals who do not contribute to a retirement savings plan as a result of these age thresholds.[25]

1.4.4 Threats to reduced-form identification

There are two primary threats to identification in our setting. The first threat to identification is the size of the excluded region. The reduced-form bunching approach outlined in this section is less reliable as an estimation strategy when the distortions created by the kink(s) or notch(es) are not very local. While our excluded region is relatively narrow when considering changes to the age of first withdrawal, the excluded range is quite large when we consider the proportion of IRA holders in a single year who are changing the timing of their withdrawals.

The second threat to identification concerns the shape of the estimated counterfactual distribution. A key assumption of our reduced-form analysis is that the counterfactual distribution is smooth through the excluded range. We would worry that the counterfactual distribution might not be smooth if there are other related policies with the same age thresholds. We conduct two diagnostic tests to test the assumption that our observed bunching is specifically due to the age thresholds for penalties related to IRA withdrawals and not other changes in the traditional IRA holders' financial environments. First, we plot the proportion of IRA holders receiving Social Security and receiving a wage, by age. We do not observe bunching at the age threshold for either penalty. Second, we compare the proportion of account holders taking a withdrawal at each age by the half of the year in which their birthday falls. We observe bunching at ages 59 and 70 for individuals whose half birthday is in the same calendar year as their birthday (e.g., individuals who turn $59\frac{1}{2}$ in the same calendar of the year as 59), but at 60 and 71 for individuals whose half birthday is in the calendar year after their birthday. These diagnostics support our assumption that the bunching we observe is in fact due to the age thresholds related to withdrawal penalties (see Appendix A.3.1 for more details).

The assumption that the counterfactual distribution is smooth is also problematic if individuals use the age thresholds as reference points. In our setting, this would be true if traditional IRA holders consider the age thresholds as suggestions that they should start taking out withdrawals at those ages. There is a hint of this occurring for Roth accounts.

[25] It is worth noting that if these penalties resulted in fewer contributions to traditional IRAs, this would (likely) reduce the amount of consumption that is deferred to age $59\frac{1}{2}$ or later.

Figure A.5a in Appendix A.3.2 shows a small increase in the proportion of individuals taking withdrawals from Roth IRAs at age $70\frac{1}{2}$ even though the RMD rules do not apply for Roth accounts. In Figure A.6, we show that IRA holders with both types of accounts may be more likely to take normal withdrawals from a Roth account at age $70\frac{1}{2}$ than IRA holders who only have a Roth account. That is, we believe individuals who hold both types of accounts may not understand that the RMD rules do not apply to their Roth account as well (or choose to impose a similar heuristic to their Roth account), and therefore begin taking withdrawals at age $70\frac{1}{2}$ from both accounts. The primary concern with reference points in bunching analyses is that it is impossible to disentangle the reference point response from the response to the financial incentive. This means any estimated elasticities will overstate the true structural elasticities. Even if these individuals do use the age thresholds for these penalties as reference points, however, it should not be a problem for our reduced-form results because we estimate the magnitude of the behavioral response rather than underlying behavioral elasticities.

1.5 The dynamic life-cycle model

Section 1.4 presented strong evidence of a behavioral reaction to the penalties for early withdrawal and excess accumulation. While our reduced-form results help us answer the question of how many people are shifting the timing of their IRA withdrawals in response to these penalties, we are not able to use our reduced-form estimates to understand the potential welfare and tax revenue impact of changes to these penalties. In order to evaluate counterfactual policies, we develop and estimate a dynamic life-cycle model. The estimated model gives us a framework in which we are able to change these penalties and analyze the impact on savings behavior, welfare, and tax remittances.

The key features of the model are as follows. Every period, individuals choose consumption and how much to save (or dissave) in two different assets: a standard savings account and a tax-benefited account. Dissaving from the tax-benefited account is penalized before period $t < t_e$, and required after period $t > t_{rmd}$. Individuals receive exogenous labor income and, at period t_P, begin receiving an annual pension. Individuals receive utility from consumption and from a bequest motive.

Section 1.5.1 provides details about the model, and Section 1.5.2 sets up the individual's problem. A complete list of the parameters used in the model is given in Appendix A.6.

1.5.1 Model set-up

Lifespan and survival probabilities Individuals live for no more than T periods. The conditional probability of living to period t conditional on surviving to period $t-1$ is π_t. We adopt the norm that the t subscript refers to the beginning of the period.

Exogenous income The log of labor income y_t is determined by two components: a deterministic component that is a function of the period, and a stochastic component:

$$\ln y_t = g(t, X) + \varepsilon_t^y. \tag{1.5.1}$$

The stochastic component, ε_t^y, follows an AR(1) process with normally distributed errors with mean 0 and variance σ_ε^2:

$$\begin{gathered} \varepsilon_t^y = \eta \varepsilon_{t-1}^y + \zeta_t \\ \zeta \sim \mathcal{N}\big(0, \sigma_\zeta^2\big) \end{gathered} \tag{1.5.2}$$

Formally modelling labor supply decisions is outside the scope of the model. Instead, individuals receive exogenous labor income in each period, and decreasing labor supply over the lifespan is captured in decreased labor income. Individuals receive a Social Security-style pension at an exogenously given age, t_p, which is known in advance. The value of the pension income is a function of the period at which the individual begins to receive it. Total income, z_t, is the sum of labor income and the annual pension:

$$z_t = y_t + P \cdot \mathbb{I}[t \geq t_P] \tag{1.5.3}$$

Savings Individuals have access to two types of savings accounts: a standard savings account (S) and a tax-benefited account (A). At the beginning of period t, the stock of assets in the standard savings account is denoted as S_t and in the tax-benefited account as A_t. The individual chooses how much to (dis)save in both accounts: s_t in the standard savings account and a_t in the tax-benefited account. $s_t > 0$ indicates savings; $s_t < 0$ indicates withdrawals. The same is true for a_t.

The two accounts follow similar laws of motion, which are summarized in Equations 1.5.4 and 1.5.5. After the individual has chosen her (dis)saving amounts s_t and a_t for period t, the amount is added from the account balance at the beginning of the period. The stock of assets in each type of savings account at the start of the next period is equal to the balance at the end of the previous period after interest. The two savings accounts face different pre-tax rates of return, with $r_S < r_A$. There is no borrowing.

$$S_{t+1} = (1 + r_S)(S_t + s_t) \tag{1.5.4}$$

$$A_{t+1} = (1 + r_A)(A_t + a_t) \tag{1.5.5}$$

Contributions to the tax-benefited account are deducted from taxable income, while withdrawals from the tax-benefited account are added to taxable income. Contributions are capped at a period-dependent level, $\overline{a}_t$. Withdrawals from the tax-benefited account before period t_e are penalized at rate p_e. Minimum withdrawals, $a_{t,rmd}$, are required after period t_{rmd}, with failure to withdraw penalized at rate p_{rmd}. The cost of saving in the tax-benefited account is reduced liquidity before period t_e. This is relevant given uncertainty in the earnings process.

The individual faces two potential penalties for withdrawals from the tax-benefited account. The early withdrawal penalty is given by τ_e in Equation 1.5.6. If $a_t < 0$ (indicating a withdrawal) and $t < t_e$, the individual owes an additional $p_e \cdot |a_t|$ in taxes.

$$\tau_e = p_e \cdot |a_t| \cdot \mathbb{I}[a_t < 0, t < t_e] \tag{1.5.6}$$

Starting in period t_{rmd}, individuals are required to take Required Minimum Distributions from their tax-benefited savings account. Required Minimum Distributions are a function of the beginning-of-period account balance and the period:

$$a_{t,rmd} = f(A_t, t)$$

Failure to take the minimum amount required triggers the excess accumulation penalty, represented by τ_{rmd} in Equation 1.5.7. If $|a_t| < a_{t,rmd}$ and $t \geq t_{rmd}$, the individual owes an additional $p_{rmd} \cdot (a_{t,rmd} - |a_t|)$ in taxes.

$$\tau_{rmd} = p_{rmd} \cdot (a_{t,rmd} - |a_t|) \cdot \mathbb{I}[a_t < a_{t,rmd}, t \geq t_{rmd}] \tag{1.5.7}$$

Taxation The individual owes income tax in each period. Income tax owed is determined by a function $\tau_y(\cdot)$ with its argument equal to the sum of total income z_t and withdrawals from the tax-benefited savings account. Total taxes owed are equal to income tax plus any penalties due to early withdrawals or failure to take a Required Minimum Distribution:

$$T(z_t, a_t, \tau_e, \tau_{rmd}) = \tau_y\big(z_t - a_t\big) + \tau_e + \tau_{rmd}. \tag{1.5.8}$$

Utility from consumption Utility from consumption is given by a constant EIS function:

$$u(c_t) = \frac{\sigma}{\sigma - 1} c_t^{\frac{\sigma-1}{\sigma}}, \tag{1.5.9}$$

where σ is the elasticity of intertemporal substitution.

Bequest motive Individuals value end-of-life wealth W via the warm-glow bequest motive given in Equation 1.5.10:

$$B(W_{t+1}) = \theta \frac{\alpha}{\alpha - 1} \left(\frac{W_{t+1}}{\theta} \right)^{\frac{\alpha-1}{\alpha}}, \tag{1.5.10}$$

where $W_{t+1} = (1 + r_S)S_{t+1} + (1 + r_A)A_{t+1}$. That is, if an individual dies between periods t and $t+1$, the value of her bequest is equal to what would have been her total starting wealth in period $t+1$. The specification of the bequest motive follows that used in Jakobsen et al. (2019). The individual does not owe any taxes on $B(W_{t+1})$.

1.5.2 The individual's maximization problem

An individual starts period t knowing the following state variables: the period (t), her exogenous labor income shock (ε_t^y), the level of assets in both savings accounts (S_t and A_t), the period when she will receive her annual pension (t_P), and what will be the value of her annual pension (P). We collectively describe these state variables as $\Omega_t = \{t, \varepsilon_t^y, S_t, A_t, t_P, P\}$.

Knowing the state variables, individuals make two choices each period to maximize expected lifetime utility: how much to (dis)save in both savings accounts (s_t and a_t). This decision implies post-tax consumption (c_t) and what would be left in a bequest.

The individual's problem in recursive form is to pick $\{c_t, s_t, a_t\}$ to maximize the Bellman equation given by Equation 1.5.11:

$$V_t(\Omega_t) = u(c_t) + \beta\Big(\pi_{t+1}\mathbb{E}\big[V_{t+1}(\Omega_{t+1})\big] + (1 - \pi_{t+1})B(W_{t+1})\Big) \tag{1.5.11}$$

subject to the following constraints:

$$\text{Budget constraint: } z_t + S_t + A_t = c_t + T(z_t, a_t, \tau_e, \tau_{rmd}) + \frac{S_{t+1}}{1 + r_S} + \frac{A_{t+1}}{1 + r_A} \tag{1.5.12}$$

$$\text{No borrowing constraint: } S_{t+1} \geq 0, \quad A_{t+1} \geq 0 \tag{1.5.13}$$

as well as the savings laws of motion (Equations 1.5.4 and 1.5.5).[26]

V_t is the present value of expected lifetime utility at period t. The model is governed by four preference parameters: β is the discount factor, σ is the elasticity of intertemporal

[26]In general, individuals are not allowed to borrow against IRAs. In 2020, the Coronavirus Aid, Relief, and Economic Security (CARES) Act allowed individuals to take early withdrawals without penalty to pay for specific COVID-19-related expenses, and to repay the withdrawal if the repayment is within three years of the withdrawal.

substitution, α is the bequest elasticity, and θ is the weight put on the bequest motive. Individuals save both to protect against bad labor income draws and to finance additional consumption later in life when labor income is low. The exogenous discontinuities in the budget constraint caused by the two penalties mean there is not a closed form solution to the model. We provide details on how we solved the problem numerically in Appendix A.7.

1.6 Model results

We estimate the model in two steps. In the first step, we set the value of parameters that can be cleanly estimated outside of the model (e.g., survival probabilities) or are institutional in nature (e.g., the income tax schedule). In the second step, we use the Simulated Method of Moments (SMM) to estimate the preference parameters of the model: the elasticity of intertemporal substitution (EIS), the discount factor, and the two parameters governing the bequest motive. This two-step process is standard in the estimation of life-cycle models (see, e.g., French (2005)).

We present the first-step parameter estimates in Section 1.6.1 and the second-step parameter estimates in Section 1.6.2. We describe the estimation procedure in Section 1.6.2.1, the variation that identifies the four preference parameters in Section 1.6.2.2, the preference parameter estimates in 1.6.2.3, and model fit in Section 1.6.2.4.

1.6.1 First step estimates

1.6.1.1 Survival probabilities and tax policy parameters

Lifespan and survival probabilities We model individuals starting at age $t = 40\frac{1}{2}$ and set $T = 85\frac{1}{2}$.[27] Survival probabilities for each age are calculated using the U.S. Social Security Actuarial Life Tables for 2010. We normalize the survival parameters so that the probability of death at age $85\frac{1}{2}$ is 1.[28] The probabilities are given separately for men and women; we use the average of the two.

Tax-benefited savings account We set the parameters specifying the age and penalty levels for the early withdrawal and excess accumulation penalties equal to the statutory levels for traditional IRAs: withdrawals before age $59\frac{1}{2}$ are subject to an additional $p_e = 10\%$ tax due on the amount withdrawn, while minimum withdrawals must be taken after age $70\frac{1}{2}$

[27] In the 2010 U.S. Social Security Actuarial Life Tables, about 10% of individuals survive past age 85.

[28] See Table 6 of Life Tables for the United States Social Security Area 1900-2100, available at https://www.ssa.gov/OACT/NOTES/as120/LifeTables_Tbl_6_2010.html. We use the survival probability to age 85 for age $85\frac{1}{2}$, the survival probability to age 84 for age $84\frac{1}{2}$, etc.

or be subject to an additional penalty of $p_{rmd} = 50\%$ on the amount not withdrawn. The required distribution schedule is modeled after the true schedule and discussed in Appendix A.1.5.

Contributions to the tax-benefited savings account are capped. Because all dollar values in the model are inflated to their 2015 equivalents, we use the contribution limits from 2015: the contribution limit was \$5,500 for individuals under age 50 and \$6,500 for individuals aged 50 or older. For more on contribution limits, see Appendix A.1.

Income tax schedule Income is taxed according to the U.S. 2015 single-filer income tax brackets, given in Table 1.3. We assign all individuals the standard deduction value for single-filers in 2015 (\$6,300).

Table 1.3: U.S. 2015 single-filer income tax schedule

For income over...	Marginal tax rate	For income over...	Marginal tax rate
\$0	10.0%	\$189,300	33.0%
\$9,225	15.0%	\$411,500	35.0%
\$37,450	25.0%	\$413,200	39.6%
\$90,750	28.0%		

1.6.1.2 First step parameters estimated from the data

We describe below how we determine initial asset holdings, effective rates of return to the two savings accounts, the labor income process parameters, and pension receipt in the structural model. We make three sample restrictions to our main panel to generate the sample used to estimate the first step parameters (hereafter referred to as the "model input sample"). These restrictions ensure that the model input sample and our model set-up are internally consistent. First, we only include individuals who were never married during our sample period. This means that all of the first step parameters are estimated on individuals rather than households. Second, we limit the sample to individuals who had non-missing income in all periods. Third, we keep individuals who were born between 1920 and 1968. This restricts the sample to individuals who were between the ages of 30 and 79 in 1999. This yields a balanced panel with 60,615 unique individuals.

Initial asset holdings To determine initial balances in both the regular savings account and the tax-benefited savings account, we fit lognormal distributions to the empirical distributions at age 40 using maximum likelihood estimation. We estimate the initial empirical distribution of the regular savings account balance using information from IRS Form 1099-INT. We use the fair market value of end-of-year traditional IRA balances reported on Form 5498, winsorized at the 1^{st} and 99^{th} percentiles, to obtain the empirical distribution of the tax-benefited savings account. We find that the mean of the lognormal distribution of initial tax-benefited assets is lower than that of regular savings (5.433 versus 9.277), but that the standard deviation is larger (3.530 versus 1.765). Our simulated individuals receive one draw from the distribution for regular savings and one from the distribution for tax-benefited savings.[29]

Rates of return We calibrate the values of $1 + r_S$ and $1 + r_A$ to be 1.02 and 1.05, respectively. Having a sufficient wedge in the rate of return for the standard savings account and tax-benefited saving account is critical for the model simulations to fit the data. Investments in tax-benefited accounts enjoy abnormally high real after-tax rates of return for multiple reasons. The predominant reason is that returns to tax-benefited accounts are not subject to annual taxation. Instead, returns accumulate tax-free and are subject to tax upon withdrawal, a treatment that effectively permits investors to obtain compounded returns on deferred tax liabilities. Any positive inflation rate increases nominal returns and affects the effective tax rate on investments outside of tax-preferred accounts, further increasing the difference between real after-tax rates of return in taxable and tax-benefited accounts (Feldstein, 1976; Feldstein et al., 1978). In addition, IRAs tend to be heavily invested in equities, whereas assets in non-IRAs are often more diversified. The equity premium is a long-standing puzzle in economics (see, e.g., Mehra and Prescott (1985); Benartzi and Thaler (1995)).

We estimate $1 + r_A$ directly from our data. We limit the model input sample to account-years when the account holder did not not take a withdrawal or make a contribution or rollover, and when the observed growth was between 0% and 15%. After accounting for inflation, the average observed growth rate is 5.1%. The value of $1 + r_S = 1.02$ corresponds to the long-term real interest rate assumed by the 2020 OASDI Trustees Report.[30]

Labor income We estimate the deterministic component of exogenous labor income $g(t, X)$ using a fixed effects model that is cubic in age with year-of-birth fixed effects, as shown in Equation 1.6.1. Labor income is defined as the sum of wage income and self-employment

[29]See Appendix A.2.1 for additional details about the initial distribution of the regular savings account.

[30]The long-range annual real interest rate assumed by the 2020 Old Age, Survivors, and Disability Insurance (OASDI) report was 1.8% in the high-cost scenario, 2.3% in the intermediate-cost scenario, and 2.8% in the low-cost scenario. As the assumed real interest rates used in the annual OASDI Trustees report have been slowly falling over the past decade, we err on the low side and use 2% (OASDI, 2020).

income.

$$\ln y_{it} = \rho_0 + \rho_1 t + \rho_2 t^2 + \rho_3 t^3 + YOB_i + \varepsilon_{it}^y \tag{1.6.1}$$

The constant is estimated as the average of the year-of-birth fixed effects. We set the fixed effect equal to 0 in our simulations as there is only one cohort.

The residuals from Equation 1.6.1 follow an AR(1) process, as given in Equation 1.6.2:

$$\begin{aligned} \varepsilon_{it}^y &= \eta \varepsilon_{it-1}^y + \zeta_{it} \\ \zeta_{it} &\sim \mathcal{N}\big(0, \sigma_\zeta^2\big) \end{aligned} \tag{1.6.2}$$

We estimate the labor income process using the sum of wage income and self-employment income. The results of this exercise are given in Table 1.4:

Table 1.4: Exogenous labor income process parameter estimates

Deterministic component				AR(1) Parameters	
$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2 * 100$	$\hat{\rho}_3 * 1000$	$\hat{\eta}$	$\hat{\sigma}_\zeta^2$
7.682	0.143	-0.185	0.0036	0.829	0.267

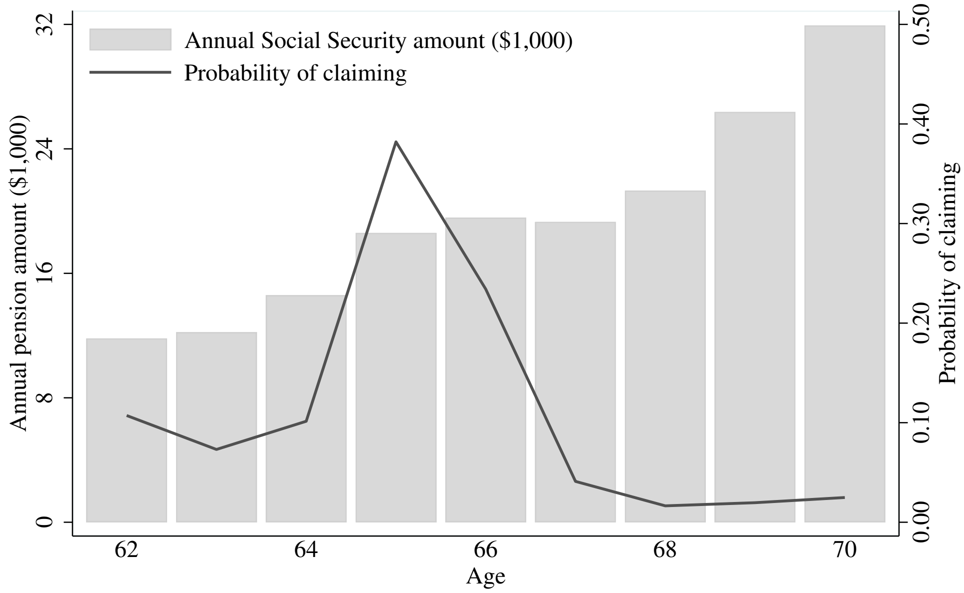
Notes: N = 60,615.

Our estimate of the persistence parameter η is within the range of previous estimates. For example, Laibson et al. (1998) finds a persistence parameter of 0.511 for high school drop-outs and 0.688 and 0.686 for high-school graduates and college graduates, respectively, while Laibson et al. (2018) estimates a persistence parameter of 0.782 and Choukhmane (2021) reports a persistence parameter of 0.974.

Annual pension Individuals are exogenously assigned the age at which they begin receiving the Social Security-style pension. The probability of claiming the pension at a given age is determined by the proportion of individuals in our data that we observe first receiving Social Security at each age between 62 and 70. Simulated individuals receive the average value of (non-zero) Social Security received by the individuals in the model input sample whom we first observe claiming Social Security at that age.[31] Figure 1.4 summarizes the probability that an individual claims the pension at each age (shown by the solid lines) and the value of the pension when claimed at each age (given by the bars).

[31]We calculate these averages using the amount received one year older than the claiming age to ensure that we capture a full year of Social Security receipt.

Figure 1.4: Value of annual pension and probability of claiming



Notes: N = 3,460 individuals. Includes individuals aged 55-60 in 1999 whom we observe first receiving non-zero Social Security between the ages of 62 and 70, inclusive. The probability that an individual claimed at each age is given by the solid lines. The average amount of Social Security given at each age is shown by the bars. Social Security amounts are inflated to 2015 values.

1.6.2 Second step preference parameter estimates

1.6.2.1 Estimation procedure: Simulated Method of Moments

As our dynamic model does not have a closed-form solution, we solve the model computationally using the Simulated Method of Moments (SMM). The estimation routine is as follows. We set an initial guess for the four preference parameters, Φ_0: the elasticity of intertemporal substitution (σ_0), the discount factor (β_0), the bequest elasticity (α_0), and the weight on the bequest motive (σ_0). Using backwards iteration, we solve the model using Φ_0 and the output from the first step estimates described in Section 1.6.1.

We use the model solution to generate S sets of simulated data. For each simulated dataset $m_s(\Phi)$ ($s = \{1, 2, ..., S\}$), we calculate a series of moments denoted $h(m_s(\Phi))$. Note that the simulated data, and therefore the value of $h(m_s(\Phi))$, are a function of the structural parameters Φ used to solve the model.

We compare the average of these simulated moment vectors to the values of the same moments calculated from the administrative tax data, $h(w)$ (where w represents the true empirical data). Specifically, we calculate Equation 1.6.3:

$$g(w,\Phi) = h(w) - \frac{1}{S}\sum_{s=1}^{S} h(m_s(\Phi)). \tag{1.6.3}$$

Our SMM estimator $\hat{\Phi}_{SMM}$ is then defined as the solution to the following minimization problem:

$$\hat{\Phi}_{SMM} = \arg\min_{\Phi} Z(\Phi,n) \equiv g(w,\Phi)'\hat{W}_n g(w,\Phi), \tag{1.6.4}$$

where $\hat{W}_n$ is a positive definite weighting matrix. We set $\hat{W}_n$ to be the identity matrix. Using the identity matrix places more weight on the moments that are largest in absolute value, whereas the optimal weighting matrix (i.e., the inverse of the variance-covariance matrix of the empirically estimated moments) places more weight on the moments that are most precisely estimated. As is shown in Table 1.6, some of our moments are estimated with a considerably larger sample size than others, which mechanically makes them more precisely estimated. This results in a model fit that quantitatively distorts our counterfactual exercises by providing a better fit for some age groups than for others. As our moments are generally of the same magnitude, we find that using the identity matrix yields better model fit overall relative to the optimal weighting matrix.

We minimize $Z(\Phi,n)$ using Nelder-Mead optimization. We generate 10,000 individuals for each simulation. More details about the numerical solution method are given in Appendix A.7.

1.6.2.2 Estimation moments and identification

We jointly estimate the four preference parameters of the model. Because all four parameters govern savings decisions, we cannot rely on the shape of the savings profile alone to separately identify these parameters. Instead, we take advantage of the bunching moments generated by the early withdrawal and excess accumulation penalties. Matching on bunching and other quasi-experimental moments provides more credible variation for identifying preference parameters in structural models.

We use 25 moments in our estimation procedure: the average value of IRA withdrawals at each age between $50\frac{1}{2}$ and $74\frac{1}{2}$. We show in a simple, highly-stylized, model that bunching in our setting can be used to identify the EIS and the discount factor in Appendix A.5. In our model, individuals face uncertain survival and receive utility from the ability to leave a bequest. The expected utility from leaving a bequest impacts consumption and, subsequently, withdrawal decisions. As a result, average withdrawal sizes enable us to identify the two bequest parameters of the model.

1.6.2.3 Preference parameter estimates

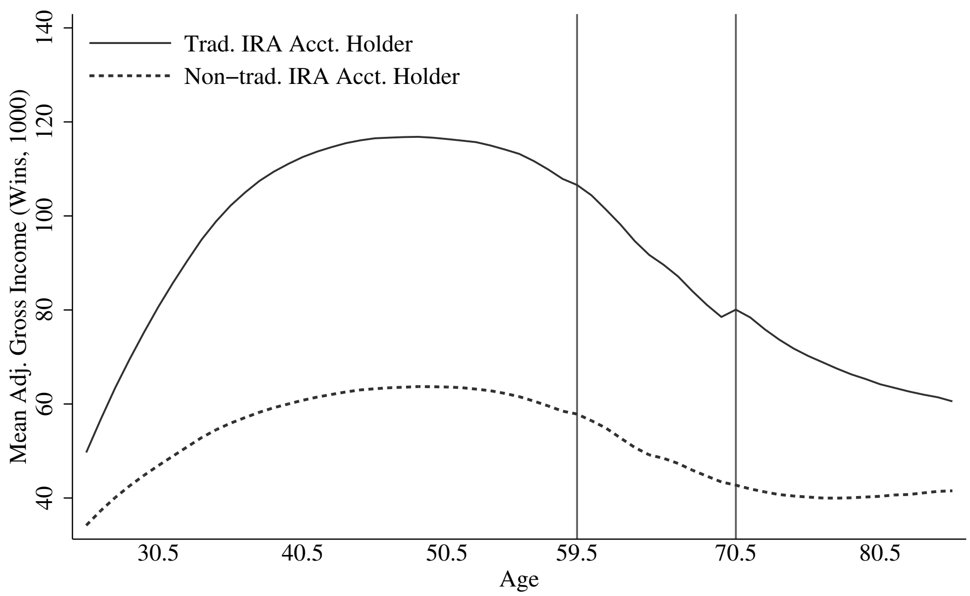
The preference parameter estimates are given in Table 1.5:

Table 1.5: Preference parameter estimates

σ: EIS	β: Discount factor	A: Bequest weight	α: Bequest elasticity	χ^2
1.061	0.920	2.660	1.069	307
(0.157)	(0.00857)	(0.461)	(0.157)	d.f. = 307

Notes: This table show the estimated preference parameters of the structural model described in Section 1.5. Standard errors are shown in parentheses. The optimal weight matrix used to calculate the standard errors is generated via bootstrap.

Figure 1.5: Average adjusted gross income for IRA and non-IRA holders, by age



Notes: N = 60,615 unique individuals. Adjusted gross income winsorized at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles. Adjusted gross income amounts are inflated to 2015 values.

We estimate an EIS of 1.061. This is considerably higher than that estimated by both Best et al. (2019) and Choukhmane (2021), who estimate the EIS equal to about 0.1 and 0.4, respectively, but not inconsistent with the upper-end of the range of estimates found in the literature.[32] Chetty (2006) notes that, in models where the EIS is equal to the inverse of the

[32]For example, Gourinchas and Parker (2002) report a value for the coefficient of relative risk aversion

CRRA, it is unreasonable to find an EIS less than 1, as it would require an uncompensated wage elasticity lower than any value previously estimated. Guvenen (2006) argues that, in general, wealthier individuals have a higher EIS. Figure 1.5 compares average adjusted gross income for IRA holders and non-IRA holders at each age. IRA holders have considerably higher average income at every age, implying that they are very likely wealthier and one would expect a higher EIS estimate for this population.

The model is formally rejected by a χ^2 overidentification test. This is primarily driven by the fact that the estimation procedure does not incorporate the variance of the parameters estimated in the first step (e.g., the income process parameters).[33]

We present t-statistics on the moment conditions in Table 1.6. We find large t-statistics on many of the moment conditions. This is in part the result of the fact that the model is highly stylized, and that many of the moments are estimated on samples with thousands (if not tens of thousands) of observations.

Table 1.6: Empirical moments, simulated moments, and t-statistics

Age	Empirical moment	Simulated moment	t-stat	Age	Empirical moment	Simulated moment	t-stat
50	1,578	860	13.97	63	4,654	4,502	1.75
51	1,620	937	13.00	64	4,525	5,325	−4.11
52	1,731	1,031	13.37	65	4,595	6,293	−8.90
53	1,721	1,161	11.18	66	4,306	2,393	10.27
54	1,747	1,316	8.93	67	3,847	2,873	6.92
55	1,876	1,508	9.09	68	3,542	3,461	0.96
56	2,022	1,725	6.41	69	3,613	3,992	−0.95
57	1,972	1,976	3.81	70	6,994	7,436	−1.34
58	1,794	2,287	−0.56	71	8,025	7,489	2.13
59	3,356	3,297	1.03	72	7,583	7,562	0.26
60	4,265	3,919	5.10	73	7,787	7,500	0.40
61	4,438	4,673	−0.40	74	7,670	7,454	0.32
62	4,697	3,888	5.72				

Notes: Empirical moments show the average withdrawal value at that age, conditional on having a non-0 account balance in the IRA, as estimated in the model input sample. The simulated moments are the average withdrawal value at that age, conditional on having a non-0 account balance in the IRA, as estimated in our simulated data (N = 10,000).

(CRRA) of 0.514 when using their robust weighting matrix, which implies an EIS of $1/0.514 = 1.95$ in their framework. Best et al. (2019) use Epstein-Zin-Weil preferences to separately estimate the EIS and the CRRA, which could explain some of the disparity in our estimates.

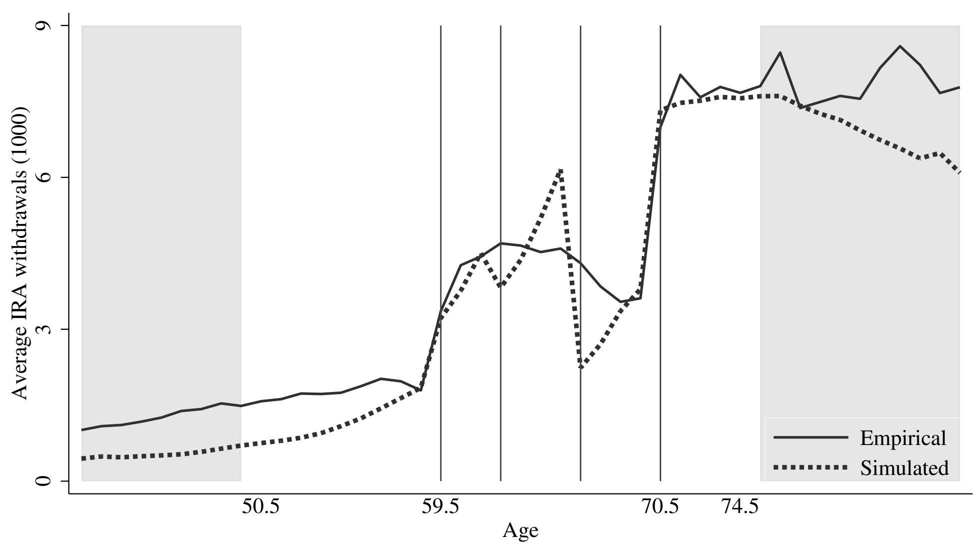
[33] Our standard errors are likely understated for the same reason.

1.6.2.4 Model fit for matched moments

Figure 1.6 compares the simulated data moments to the empirical moments. The figure shows the average withdrawal from a traditional IRA taken at each age, conditional on having a non-0 traditional IRA balance at the beginning of the period. The solid lines represent the moments from the empirical data; the dashed lines show the moments from the simulated data. The shaded regions include moments that are not targeted in the estimation procedure. The key feature of Figure 1.6 is that we accurately capture the bunching at ages $59\frac{1}{2}$ and $70\frac{1}{2}$. This is important because our counterfactual policy analysis will involve changing the two tax penalties that cause the observed bunching.

The dips in simulated average withdrawals between ages of $60\frac{1}{2}$ and $70\frac{1}{2}$ correspond with ages when the majority of our simulated individuals begin claiming their annual pension. Just over 10% of our sample starts claiming their pension at age $62\frac{1}{2}$, after which we observe a dip in average IRA withdrawals at that age (see Figure 1.4 in Section 1.6.1). By age $66\frac{1}{2}$, 90% of our sample are receiving their annual pension, which aligns with the second dip in average withdrawal value.

Figure 1.6: Comparing matched moments from the empirical and simulated data



Notes: N = 60,615 unique individuals. We simulated 10,000 individuals. The figure shows the average withdrawal from a traditional IRA taken at each age, conditional on having a non-0 IRA balance at the beginning of the period. The solid lines represent the moments from the empirical data; the dashed lines show the moments from the simulated data. The shaded regions include moments that are not targeted in the estimation procedure.

1.6.2.5 Model fit for unmatched moments

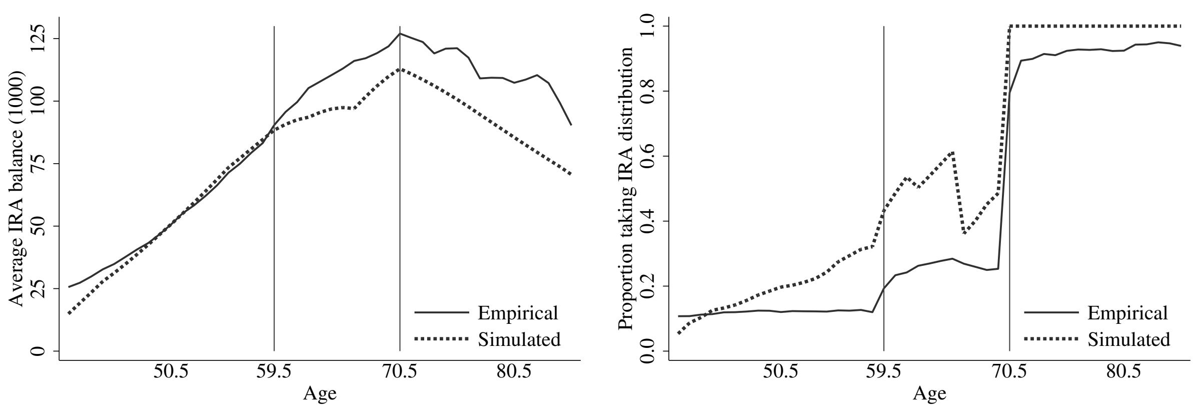
We compare the empirical and simulated moments for two additional measures of IRA behavior to further validate our model: average balance by age, and the proportion of account holders taking a withdrawal at each age. We consider the averages for each value conditional on having a non-0 traditional IRA balance at the beginning of the period.

Figure 1.7 shows the results of this exercise. The solid lines represent the moments from the empirical data; the dashed lines show the moments from the simulated data. Figure 1.7a compares average balance values. We match the empirical moments well until age $59\frac{1}{2}$, after which we mirror the shape but are a bit lower in levels. This is in part due to the fact that we do not have rollovers in the model, which would pull up average empirical account balances. After age $70\frac{1}{2}$, average balances are influenced by the value of the required minimum withdrawals, which are themselves a function of average balances. It's not a surprise that our average simulated values remain below the average empirical values in this region.

Figure 1.7b shows the proportion of individuals taking a withdrawal at each age. We are able to capture the magnitude of bunching even in this out of sample exercise.

Figure 1.7: Comparing unmatched moments from the empirical and simulated data

(a) Average IRA balances ($1,000), by age (b) Proportion taking withdrawal, by age



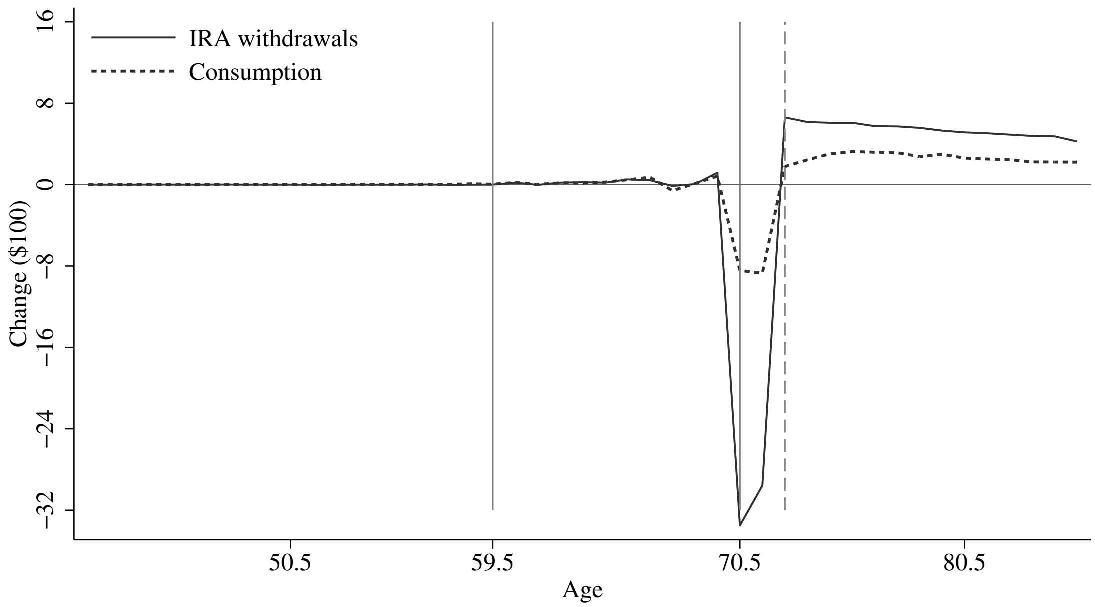
Notes: N = 60,615 unique individuals. We simulate 10,000 individuals. Panel (a) shows the average IRA balance, by age. Panel (b) shows the proportion of traditional IRA holders taking a withdrawal, by age. The solid lines represent the moments from the empirical data; the dashed lines show the moments from the simulated data.

1.7 Counterfactual policy analysis

These penalties have been changed to achieve policy goals. For example, the SECURE Act increased the age at which RMDs kick in from $70\frac{1}{2}$ to 72. This change was made with little understanding about what the long term impact would be on welfare and tax remittances. In fact, there is no evidence about whether these penalties do or do not achieve their purported goals of increasing retirement consumption while limiting the tax cost. In this section, we estimate the impact of the change to the age threshold for RMDs from the SECURE Act before considering more broadly a large number of combinations of age thresholds and penalty rates.[34]

1.7.1 Increasing the age for Required Minimum Distributions from $70\frac{1}{2}$ to $72\frac{1}{2}$

Figure 1.8: Impact of raising the age for Required Minimum Distributions from $70\frac{1}{2}$ to $72\frac{1}{2}$ on average IRA withdrawals and consumption



Notes: Each iteration includes 10,000 unique simulated individuals. The figure shows the difference in average IRA withdrawals and consumption from the base policy after we raise the age for Required Minimum Distributions from $70\frac{1}{2}$ to $72\frac{1}{2}$. A positive number means the value is greater under the counterfactual policy; a negative number means the value is lower under the counterfactual policy. The solid lines show average IRA withdrawals, and the dashed lines show average consumption levels.

The SECURE Act raised the RMD age to 72, allowing for longer accumulation in tax-benefited accounts. We examine the impact on IRA withdrawal behavior after changing the age for RMDs from $70\frac{1}{2}$ to $72\frac{1}{2}$.[35] Figure 1.8 shows the changes in average withdrawal

[34]Lawmakers have also changed these penalties during the last two major financial crises. RMDs were suspended in 2009 as part of the Worker, Retiree, and Employer Recovery Act of 2008. In March 2020, the Coronavirus Aid, Relief, and Economic Security Act (CARES Act) waived the early withdrawal penalty for withdrawals due to COVID-19-related financial hardships and suspended RMDs for 2020.

[35]We solve the model in whole periods and therefore can only consider changes in whole years. Because we consider a slightly bigger change than what was actually enacted in 2019, our estimates should be considered

amounts and consumption at each age relative to the current policy. A positive number means the value is greater under the counterfactual policy; a negative number means the value is lower under the counterfactual policy. The solid lines show average IRA withdrawals, and the dashed lines show average consumption levels.

We observe that average IRA withdrawals decrease by just over \$3,000 at age $70\frac{1}{2}$ after the age for RMDs increases. The decrease at age $71\frac{1}{2}$ is nearly as large. Average consumption also drops, though by less than \$1,000 at both $70\frac{1}{2}$ and $71\frac{1}{2}$. When individuals are required to take withdrawals from their account, they can choose to use the withdrawal to finance consumption or save the withdrawal at the standard savings rate. In contrast, when individuals are not required to take withdrawals, their decision is to take a withdrawal for consumption or continue to save that money in the tax-preferred account. Our results imply that the optimal consumption choice under those two scenarios is not the same.

Average IRA withdrawals are both higher at age $72\frac{1}{2}$ relative to the current policy, and at every age thereafter. This must be true mechanically if average IRA balances are higher due to decreased withdrawals at ages $71\frac{1}{2}$ and $72\frac{1}{2}$. We also see elevated levels of consumption at age $72\frac{1}{2}$, and for the rest of the simulated lifespan, relative to the level under the current policy.

There is essentially no change in withdrawals before age $65\frac{1}{2}$. There is similarly no change in consumption, suggesting that increasing this tax benefit may not induce people to increase their contributions to IRAs from the base level. This means that the additional tax benefit may not be sufficient to encourage individuals to save more in IRAs at the expense of current consumption or other types of precautionary savings. We observe slight changes in both average withdrawals and consumption between ages $65\frac{1}{2}$ and $70\frac{1}{2}$.

We compare four additional outcomes under the current policy and after raising the age for Required Minimum Distributions to $72\frac{1}{2}$. The first is a measure of equivalent variation, which we use to estimate the welfare consequences of raising the age for required withdrawals. To estimate this measure, we calculate the average present discounted value of lifetime utility under both the base policy and the counterfactual policy as shown in Equation 1.7.1. Let u_{it} be the utility in period t for individual i under the original policy and u'_{it} be the utility in period t under the counterfactual policy. The present discounted value of lifetime utility for individual i under the base policy is calculated by discounting the value of u_{it} back to age $41\frac{1}{2}$, and then summing up between the ages for $41\frac{1}{2}$ and $85\frac{1}{2}$.[36] We then take the average of this value to determine the average present discounted value of lifetime utility.

upper bounds.

[36]We discount to age $41\frac{1}{2}$ rather than age $40\frac{1}{2}$ because starting at age $41\frac{1}{2}$, less than 2% of our simulated individuals are constrained by our highest grid value. See Appendix A.7 for more details.

The average present discounted value of lifetime utility under the counterfactual policy is calculated similarly.

$$u_{PDV,base} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=41.5}^{85.5}\beta^{t-41.5}(u_{it}) \tag{1.7.1}$$
$$u_{PDV,cf} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=41.5}^{85.5}\beta^{t-41.5}(u'_{it})$$

Once we have determined $u_{PDV,base}$ and $u_{PDV,cf}$, we use our policy functions from solving the model under the base policy to determine how much income we would need to give to (or take away from) our simulated individuals in the first period so that $u_{PDV,base} = u_{PDV,cf}$. A positive number indicates that our simulated individuals are better off under the counterfactual policy, while a negative number indicates that our simulated individuals are better off under the base policy.

In addition to our measure of equivalent variation, we also calculate the difference in the present discounted value of lifetime tax remittances,[37] average IRA balance at age $65\frac{1}{2}$, and average bequeathed IRA balance (i.e., average IRA balance at age $85\frac{1}{2}$).

Table 1.7 shows the results of these four outcomes. As with our measure of equivalent variation, a positive number indicates a higher value in the counterfactual policy, while a negative number indicates a lower value in the counterfactual policy. We find that increasing the age for Required Minimum Distributions from $70\frac{1}{2}$ to $72\frac{1}{2}$ yields marginally higher welfare and slightly lower tax remittances. Bequeathed IRA balances are more than 6% larger after the change. The average IRA balance at age $65\frac{1}{2}$ is slightly lower, which is consistent with what we observed in Figure 1.8 (i.e., that average withdrawals at age $65\frac{1}{2}$ are higher under the counterfactual policy).

Table 1.7: Comparing the base policy to the SECURE Act

	Equivalent variation	PDV lifetime tax remittances	IRA balance at age $65\frac{1}{2}$	Bequeathed IRA balance
Levels ($)	339	-81	-186	4,283
Percent change	0.0017%	-0.0254%	-0.1908%	6.6%

Notes: Each iteration includes 10,000 unique simulated individuals. The table shows (1) our measure of equivalent variation, (2) the present discounted value of lifetime taxes remitted, (3) the value of average IRA balances at age $65\frac{1}{2}$, and (4) the value of the bequeathed IRA at age $85\frac{1}{2}$ for the base policy and the counterfactual policy where we raise the age for Required Minimum Distributions from $70\frac{1}{2}$ to $72\frac{1}{2}$.

[37] We use Equation 1.7.1 to calculate the present discounted value of total taxes remitted, with one change: instead of discounting by the discount factor β, we discount by the rate of return for the non-tax-benefited savings account: $\frac{1}{1.01}$, where 1% is an estimate of the government's borrowing rate.

As shown in Table 1.8, the decrease in tax revenue is driven by reduced excess accumulation penalty payments. As we'll discuss in Section 1.7.2, there are good reasons to believe that the relevant measure with the excess accumulation penalty is actually the change in income taxes remitted. In that case, we observe that both welfare and tax remittances increase relative to the base policy in the world where we raise the age for required withdrawals to $72\frac{1}{2}$.

Table 1.8: Changes in sources of tax revenue from raising the age for Required Minimum Distributions from $70\frac{1}{2}$ to $72\frac{1}{2}$

Average total tax revenue remitted between ages $41\frac{1}{2}$-$85\frac{1}{2}$	
Original policy	317,800.75
RMD age = $72\frac{1}{2}$	317,720.09
Total change	-80.66
Change in income taxes	27.25
Change in early withdrawal penalties	-0.26
Change in excess accumulation penalties	-107.61

Notes: Each iteration includes 10,000 unique simulated individuals. This table shows the average total taxes remitted under the original policy and under the counterfactual policy, the difference between the total, and how changes in the three sources of tax revenue contribute to that difference.

Although we do not account for redistribution in our model, the fact that this policy increases both welfare and lifetime income tax remittances suggests that there are changes to these tax penalties that could yield better outcomes than what we have now. At the very least, we could give individuals a lump sum transfer in the last period of their lives (thereby avoiding potential distortionary effects from redistribution), and they would be strictly better off. In the next two sections, we calculate our measure of equivalent variation and the impact on lifetime tax remittances and IRA balances for a large number of changes to both tax penalties. For these estimates, we change both the penalty rate, and the age at which the penalty applies.

1.7.2 Changing the excess accumulation penalty

In the previous section, we considered one specific change to the early accumulation penalty: changing the age at which the excess accumulation penalty applies from $70\frac{1}{2}$ to $72\frac{1}{2}$. We considered this particular change because it most closely mirrored the change implemented in 2019, but there is no reason to believe that this change led us to the optimal policy, conditional on the existence of these penalties.

To consider a more exhaustive set of counterfactual policies, we compare the base policy for the excess accumulation penalty (age threshold = $70\frac{1}{2}$, penalty rate = 50%) with every combination of age thresholds $68\frac{1}{2}$ and $78\frac{1}{2}$ and penalty rates 40% to 70% (in increments of 10%). The results are given in Figure 1.9; Table A.10 in Appendix A.8 gives these results both in levels (as shown below) and as percent changes from the base policy.

Figure 1.9a gives the results for equivalent variation. Each point shows the amount of additional income we would need to give our simulated individuals for the average present discounted value of lifetime utility in the base policy to be equal to that in the counterfactual policy. The straight horizontal line at 0 indicates where there is no difference from the base policy; the base policy is indicated by the black square on the 0 line. Points above the 0 line indicate a positive value relative to the base policy (i.e., the individual is better off under the counterfactual policy), while points below the 0 line indicate a negative value relative to the base policy (i.e., the individual is worse off under the counterfactual policy).

Raising the age for required withdrawals increases welfare. This is not surprising, because raising the age is equivalent to loosening a constraint and we are modeling rational agents with exponential discounting. We find a positive value for all penalty rates considered for every age threshold above the base policy. The observed increases are small in magnitude, with the largest change being an increase of income of just under \$600. The rate of change of the equivalent variation value decreases as the age threshold rises. All of the observed differences result from changing the age threshold; there is essentially no difference if we change the penalty rate.

While there is no explicit government budget constraint in the model, we can consider whether or not the policy changes we consider would be tax-revenue neutral (or tax-revenue increasing). The results for the present discounted value of total lifetime tax remittances are shown in Figure 1.9b. Raising the age for required withdrawals has a mixed impact on total lifetime tax remittances. Across penalty rates, total lifetime tax remittances increase from age $68\frac{1}{2}$ to $69\frac{1}{2}$, but then decrease as the age threshold rises. Total tax remittances are higher relative to the base policy at ages $68\frac{1}{2}$ through $72\frac{1}{2}$ for a penalty rate of 70%, at ages $68\frac{1}{2}$ through $71\frac{1}{2}$ for a penalty rate of 60%, and at ages $68\frac{1}{2}$ and $69\frac{1}{2}$ for a penalty rate of 50%. As with equivalent variation, the changes are modest in magnitude, with the largest increase falling just over \$100.

The decreases in total tax remittances are largely driven by changes in remittances of the excess accumulation penalty. Mortenson et al. (2019) find that less than 1% of traditional IRA holders who do not comply with the RMD rules file the required Form 5329 to remit the excess accumulation penalty. As most individuals do comply with the RMD rules, this does not amount to a large number of individuals not paying the penalty. It does, however, suggest that including changes in excess accumulation penalty payments may not be the most accurate measure of changes in tax remittances if we change the age threshold and penalty rate for the excess accumulation penalty, because our simulated individuals are forced to pay the penalty.[38]

Figure 1.9c shows the change in total lifetime income tax remittances. In this case, we observe positive differences from the base policy at all age thresholds through for $76\frac{1}{2}$ a penalty rate of 70%, and at all age thresholds except for $75\frac{1}{2}$ for a penalty rate of 60%. At a penalty rate of 50%, there is an increase in income tax remittances at age $69\frac{1}{2}$ and at ages $71\frac{1}{2}$ through $73\frac{1}{2}$.

All of the counterfactual policies with age thresholds at $71\frac{1}{2}$ or higher that resulted in higher total lifetime tax remittances also had positive equivalent variation values. This means that there are numerous combinations of age threshold and penalty rate that are both welfare improving while also resulting in increased tax revenue.

An important caveat to the results for taxes remitted is that we do not account for taxes paid on inherited IRAs. The rules for what happens to an inherited IRA are complicated. The amount of tax due, and when, depends on who is the beneficiary (e.g., spouse or non-spouse), the age of the beneficiary, and whether or not the original account holder had started taking RMDs (see Appendix A.1.6 for more details). While modeling these rules is not the goal of this project, we can roughly estimate whether taxes collected on inherited IRAs will increase or decrease simply by looking at whether average IRA account balances increase or decrease in the final period relative to the base policy.

Figure 1.9d shows the difference in the amount chosen to leave in the IRA as part of the bequest in the last period relative to the base policy. Inherited IRAs increase in size as the penalty rate decreases for a given age threshold, and increase in size as the age threshold increases for a given penalty rate. As with equivalent variation and tax remittances, the variation in age threshold has a much bigger impact on the average inherited IRA balance than variation in the penalty rate. Most importantly, average bequeathed IRA balances are higher for nearly every combination of age threshold and penalty rate that also yielded increased welfare and income tax remittances (the exception being an age threshold of $71\frac{1}{2}$ with a penalty rate of 70%).

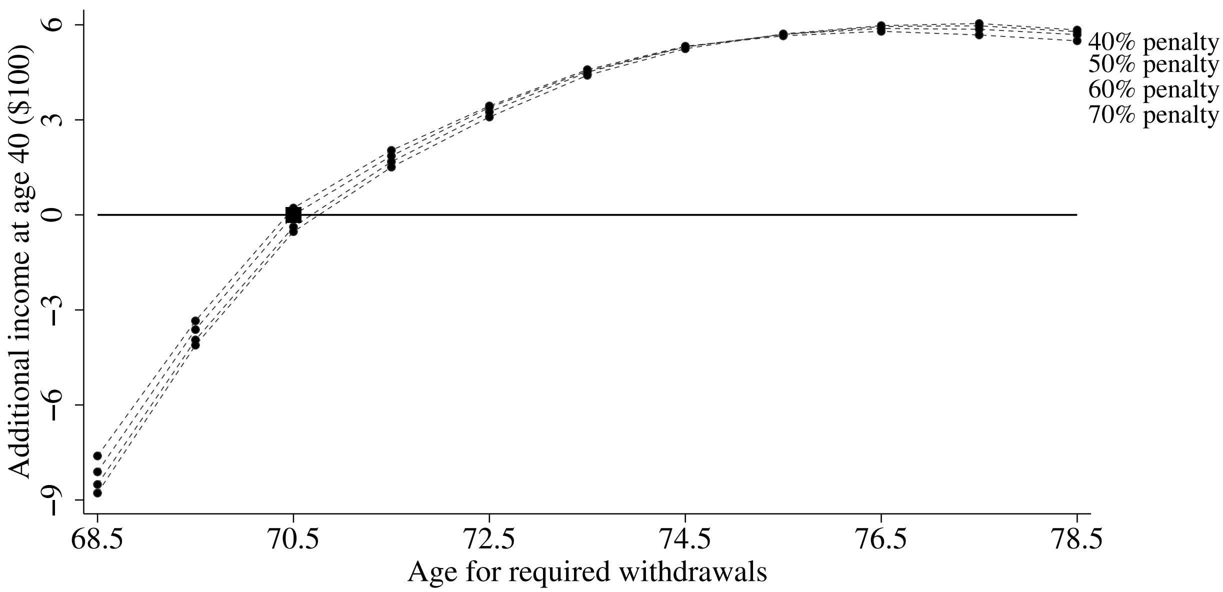
1.7.3 Changing the early withdrawal penalty

We conduct the same exercise for the early withdrawal penalty. We compare our outcomes under the base policy to every combination of age threshold between $55\frac{1}{2}$ and $65\frac{1}{2}$ and penalty rates of 5%, 10%, 20%, and 30%. The results are given in Figure 1.10; Table A.11 in Appendix

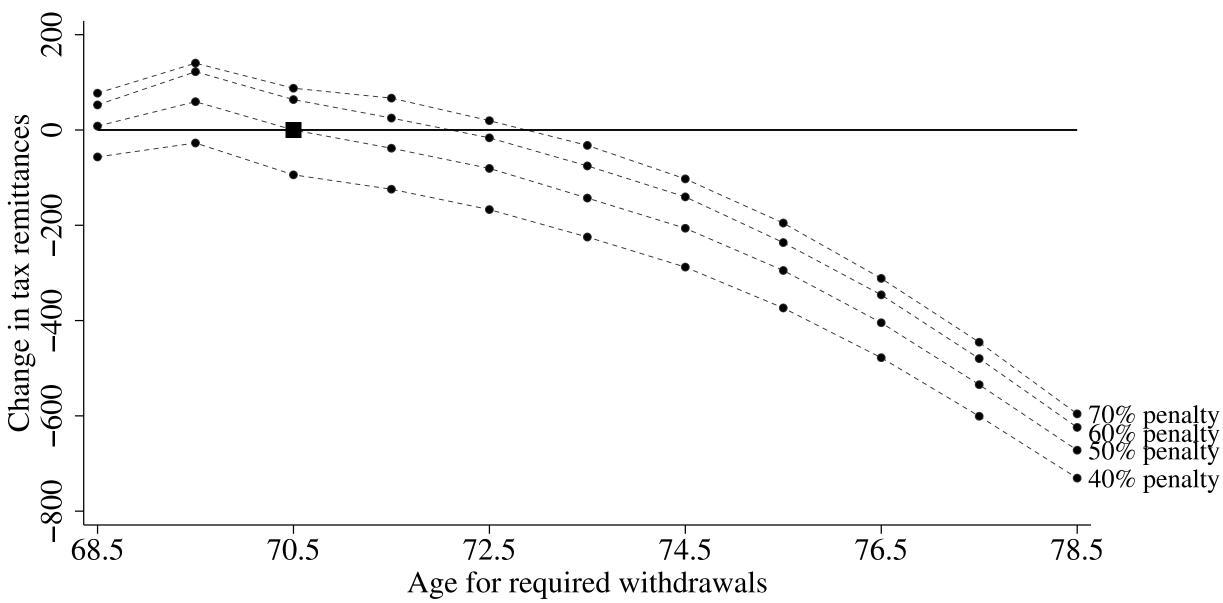
[38] Over 75% of our simulated individual-years have no excess accumulation penalty. All of our simulated individual-years with a positive excess accumulation penalty value took a positive withdrawal. About 75% of those face the penalty on a value that is less than 10% of the amount withdrawn. This suggests our overestimate of the amount of excess accumulation penalties is in part due to simulation error.

Figure 1.9: Changing the excess accumulation penalty

(a) Change in age $40\frac{1}{2}$ income in base policy to get PDV lifetime utility in counterfactual policy



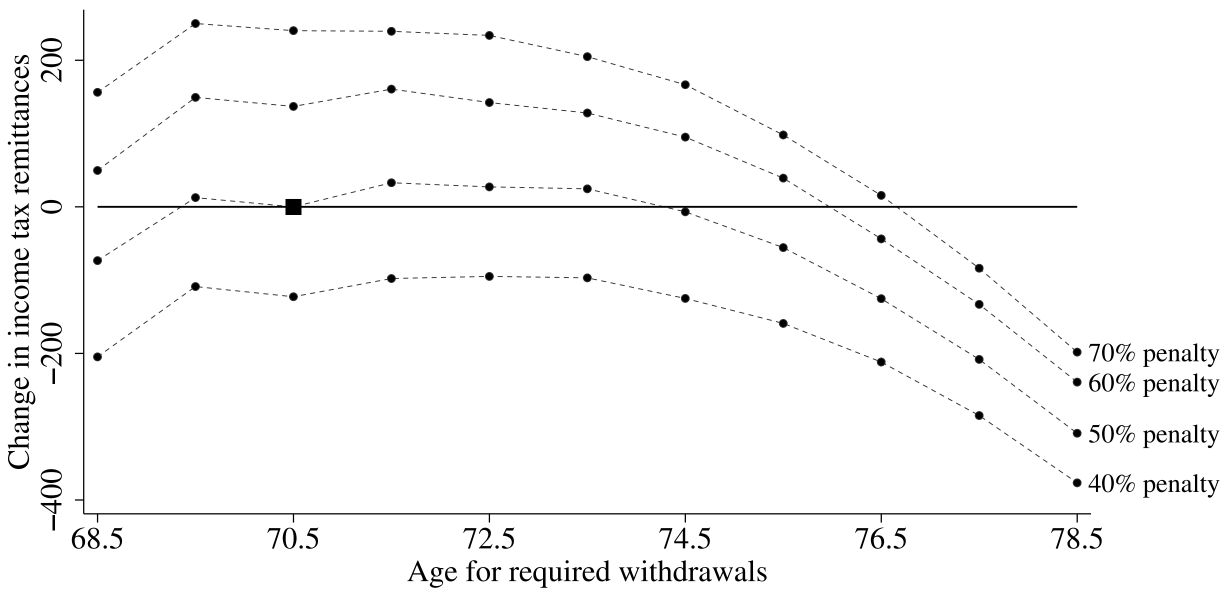
(b) Difference in lifetime taxes remitted relative to base policy ($, PDV from age $41\frac{1}{2}$)



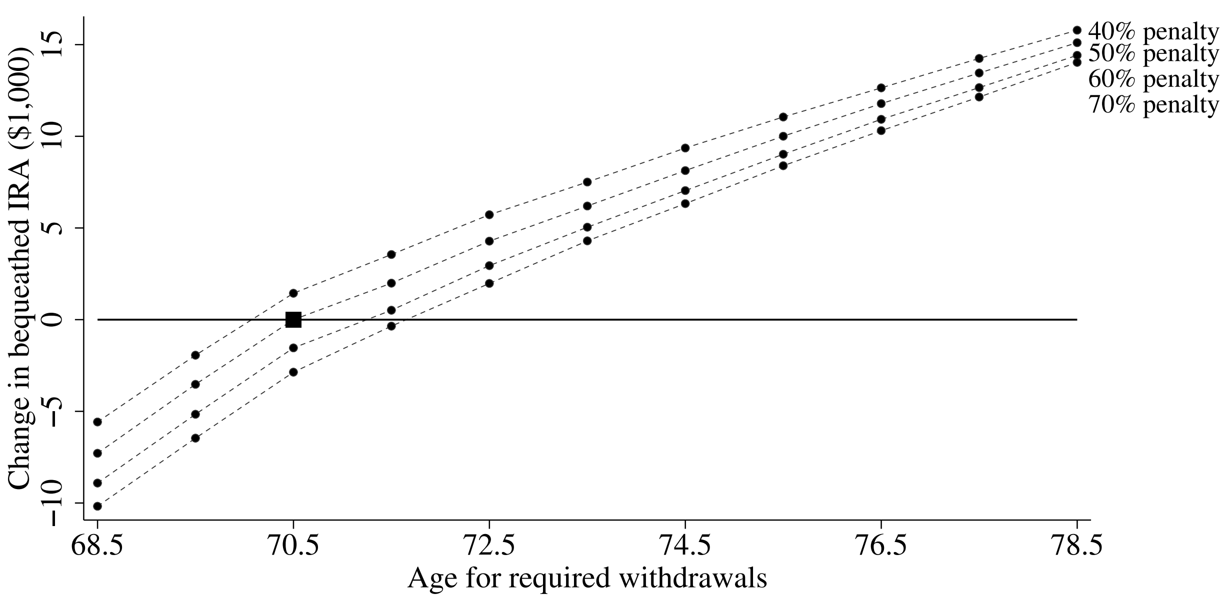
Notes: Each iteration includes 10,000 unique simulated individuals. The figure compares various counterfactual policies for the excess accumulation penalty to the base policy (age threshold = $70\frac{1}{2}$, penalty rate = 50%). Figure 1.9a shows our measure of equivalent variation. Figure 1.9b shows the change in the PDV of lifetime tax remittances. Figure 1.9c shows the change in the PDV of lifetime income tax remittances. Figure 1.9d shows the change in bequeathed IRA balances. In all panels, the base policy is indicated by a black square on the 0 line.

Figure 1.9: Changing the excess accumulation penalty, continued

(c) Difference in lifetime income taxes remitted relative to base policy ($, PDV from age $41\frac{1}{2}$)



(d) Difference in bequeathed IRA balance relative to base policy



Notes: Each iteration includes 10,000 unique simulated individuals. The figure compares various counterfactual policies for the excess accumulation penalty to the base policy (age threshold = $70\frac{1}{2}$, penalty rate = 50%). Figure 1.9a shows our measure of equivalent variation. Figure 1.9b shows the change in the PDV of lifetime tax remittances. Figure 1.9c shows the change in the PDV of lifetime income tax remittances. Figure 1.9d shows the change in bequeathed IRA balances. In all panels, the base policy is indicated by a black square on the 0 line.

A.8 gives these results both in levels (as shown below) and as percent changes from the base policy. As with Figure 1.9, the straight horizontal line at 0 indicates where there is no difference from the base policy; the base policy is marked by the black square on the 0 line. Points above the 0 line indicate a positive amount relative to the base policy, while points below the 0 line indicate a negative amount relative to the base policy.

As shown in Figure 1.10a, our equivalent variation measure decreases both as the age threshold increases and as the penalty rate increases. This follows from the fact that, for rational agents with exponential discounting, raising either the age threshold or the penalty rate tightens the lifetime budget constraint. As a result, our simulated individuals ought to be worse off in expectation. Unlike the excess accumulation penalty, where varying the age threshold had the biggest impact on welfare, changing the age threshold for the early withdrawal penalty has a smaller impact on welfare than changing the penalty rate.

The impact of changing the early withdrawal penalty on lifetime tax remittances is shown in Figure 1.10b. Increasing the penalty rate and the age threshold both lead to increased lifetime tax remittances. Increasing the age threshold has an increasingly large impact on lifetime tax remittances.

There are four age threshold-penalty rate combinations that lead to increases in both welfare and lifetime tax remittances: a penalty rate of 5% with an age threshold between $62\frac{1}{2}$ and $65\frac{1}{2}$. The increases in both measures are modest: the equivalent variation measure ranges from \$1,486 to \$2,004, while the change in lifetime tax remittances ranges from \$41 to \$412.

We present the results for lifetime income tax remittances in Figure 1.10c. We observe that income tax remittances decrease with age threshold, and with the penalty rate. There are increased income tax remittances relative to the base policy at all age thresholds facing a 5% penalty, and for all age thresholds below $59\frac{1}{2}$ facing a 10% penalty. These are the same age threshold and penalty rate combinations that led to increased welfare. As we observe individuals paying the early withdrawal penalty in the tax data, we believe that the correct measure of changes in tax remittances are total changes in tax remittances when considering the early withdrawal penalty.

Because the purported goal of the early withdrawal penalty is to improve financial security later in life, we consider the change in IRA balances at age $65\frac{1}{2}$ in Figure 1.10d. IRA balances at age $65\frac{1}{2}$ decrease with age thresholds until an age threshold of $58\frac{1}{2}$, and then increase until an age threshold of $65\frac{1}{2}$. Generally speaking, a higher penalty rate results in a higher IRA balance at age $65\frac{1}{2}$, though the higher the age threshold, the bigger the difference between penalty rates. There are numerous combinations of age threshold and penalty rate that lead to higher IRA balances at age $65\frac{1}{2}$ relative to the base policy. Most relevant to us, all three of our counterfactual policies that lead to increased welfare and total tax remittances also result in higher IRA balances at age $65\frac{1}{2}$.

1.7.4 Discussion of counterfactual exercises

We study how welfare and tax remittances change if we alter the age threshold and penalty rate for the excess accumulation penalty and the early withdrawal penalty. We find that increasing the age for Required Minimum Distributions from $70\frac{1}{2}$ to be between $71\frac{1}{2}$ and $73\frac{1}{2}$, with a penalty rate of 50% or higher, increases both welfare and the present discounted value of lifetime income tax remittances, suggesting that the increase in income taxes remitted from higher account balances may be worth the delayed tax revenue. There may be additional increases in taxes that we do not account for as a result of taxes remitted when IRAs are inherited, because bequeathed IRA balances also increase when the age for Required Minimum Distributions increases.

We find that increasing the age for penalty-free withdrawals to be between $62\frac{1}{2}$ and $65\frac{1}{2}$, while lowering the penalty rate to 5%, increases both welfare and lifetime total tax revenue. Average IRA balances at age $65\frac{1}{2}$ are also higher at these age thresholds and penalty rates. The intuition for this finding is as follows: conditional on having these penalties in place, there are benefits to encouraging individuals to keep their money in these accounts as long as possible. However, if individuals need to take early withdrawals as the result of an unexpected income shock, they can do so with minimal sanction.

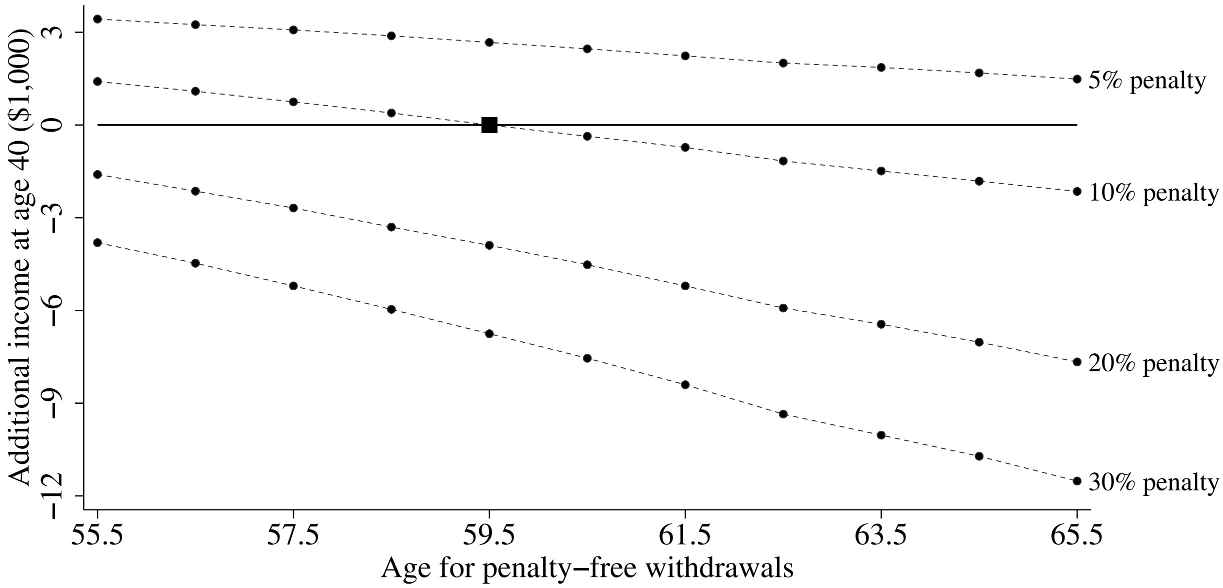
There are four things worth mentioning when considering these results. First, we have only changed one policy at a time. Our analysis of changes to the excess accumulation penalty were done keeping the base policy for the early withdrawal as is, and vice versa. There may be policies that involve changes to both penalties that yield even bigger changes in welfare and lifetime tax remittances.

Second, we have only altered two of the policy levers associated with IRAs: the two withdrawal penalties. There is a third policy lever that we have not changed: the contribution limit. The contribution limit increased substantially over the course of our sample period, from $2,000 in 1999 to $5,500 in 2015. Increasing the contribution limit would allow individuals to grow their IRA balances faster, but at the cost of putting more of their savings out of reach in the event of an unexpected income shock. Considering the role of the contribution limit is a direction for future work.

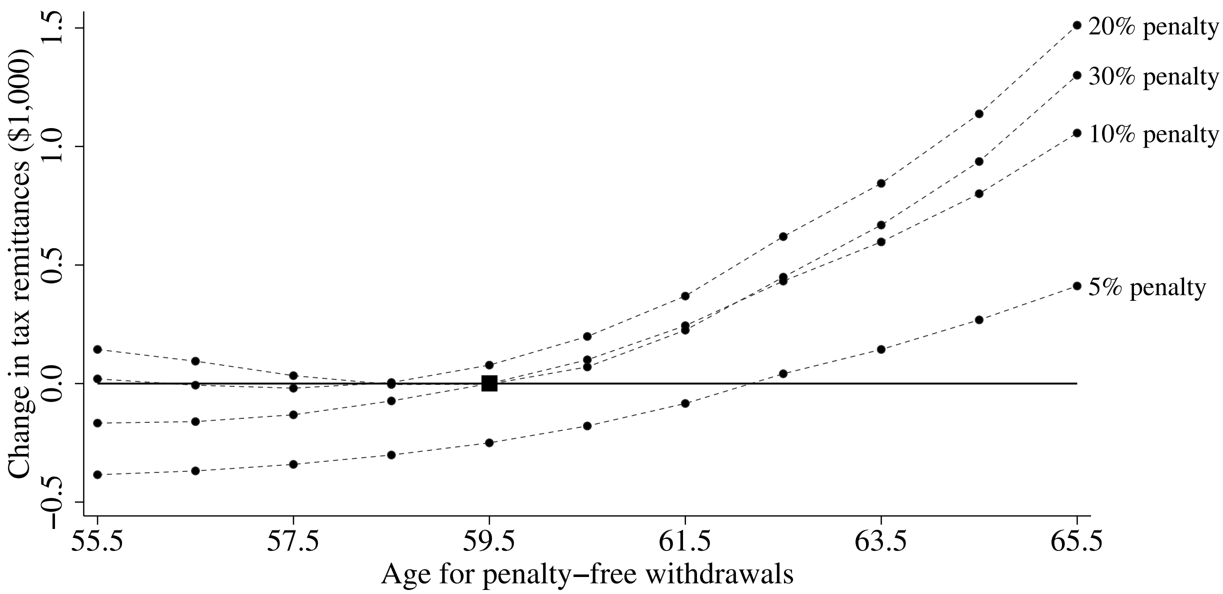
Third, these are partial equilibrium effects. Our model does not account for redistribution of tax revenue. Redistributing tax revenue would have an income effect, which could impact consumption and savings decisions. We also do not model endogenous labor supply choices. Reaching age $59\frac{1}{2}$ does not change our simulated individuals' labor income, but it's possible that individuals reduce their labor supply in response to having penalty-free access to savings

Figure 1.10: Changing the early withdrawal penalty

(a) Change in age $40\frac{1}{2}$ income in base policy to get PDV lifetime utility in counterfactual policy



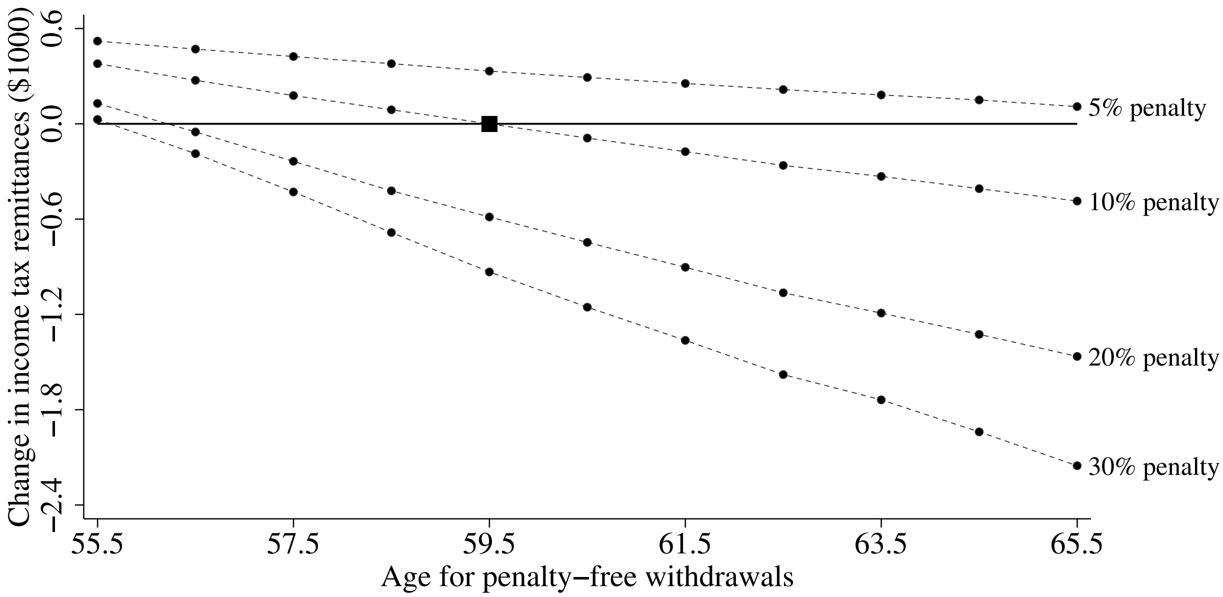
(b) Difference in lifetime taxes remitted relative to base policy ($, PDV from age $41\frac{1}{2}$)



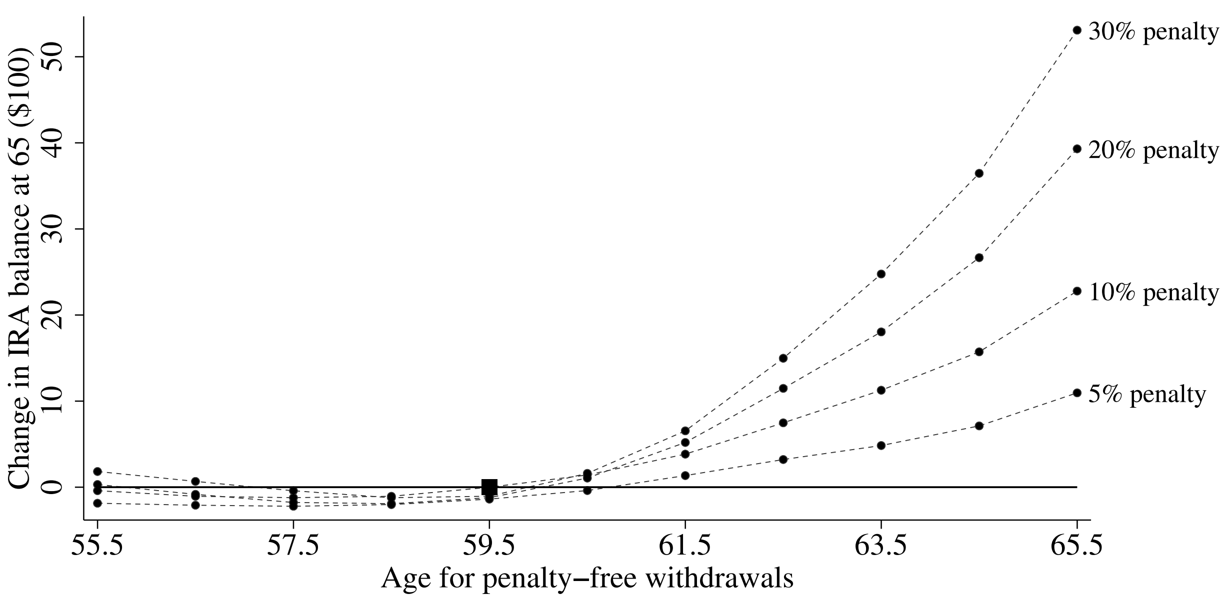
Notes: Each iteration includes 10,000 unique simulated individuals. The figure compares various counterfactual policies for the early withdrawal penalty to the base policy (age threshold = $59\frac{1}{2}$, penalty rate = 10%). Figure 1.10a shows our measure of equivalent variation. Figure 1.10b shows the change in the PDV of lifetime tax remittances. Figure 1.10c shows the change in the PDV of lifetime income tax remittances. Figure 1.10d shows the change in IRA balances at age $65\frac{1}{2}$. In all panels, the base policy is indicated by a black square on the 0 line.

Figure 1.10: Changing the early withdrawal penalty, continued

(c) Difference in lifetime income taxes remitted relative to base policy ($, PDV from age $41\frac{1}{2}$)



(d) Difference in IRA balances at age $65\frac{1}{2}$ relative to base policy



Notes: Each iteration includes 10,000 unique simulated individuals. The figure compares various counterfactual policies for the early withdrawal penalty to the base policy (age threshold = $59\frac{1}{2}$, penalty rate = 10%). Figure 1.10a shows our measure of equivalent variation. Figure 1.10b shows the change in the PDV of lifetime tax remittances. Figure 1.10c shows the change in the PDV of lifetime income tax remittances. Figure 1.10d shows the change in IRA balances at age $65\frac{1}{2}$. In all panels, the base policy is indicated by a black square on the 0 line.

in their IRAs. There could also be general equilibrium shifts in the interest rate as a result of policy-induced changes in the supply of savings.

Finally, we do not model alternative ways that individuals may finance their consumption in the face of an income shock. For example, we do not account for the scenario where an increase in the penalty rate for early withdrawals drives individuals with few other liquid savings to take on additional debt, perhaps at a high interest rate. Our results suggest that discouraging early withdrawals from retirement accounts through lower penalties that are in place for a longer period of time, and finding other policies to help individuals through unexpected financial hardship before the age of penalty-free withdrawals, may ultimately increase overall welfare.

1.8 Conclusion

We provide empirical evidence that traditional IRA holders respond to both the early withdrawal penalty and the excess accumulation penalty. We use the observed bunching response to estimate the magnitude of the response. We find that, every year, approximately 1.4% of traditional IRA holders (about 648,400 individuals) change the timing of withdrawals as a result of the early withdrawal penalty, and 16.2% of traditional IRA holders (about 7.7 million individuals) change the timing of withdrawals in response to the excess accumulation penalty. These shifts impact at least \$57.9 billion in withdrawals each year. We believe these estimates should be considered lower bounds of the number of people whose behavior changes as a result of these penalties.

After estimating our model, we consider an array of counterfactual policies in which we alter the age threshold and penalty rate for the two penalties. We find that increasing the age for penalty-free withdrawals to be between $62\frac{1}{2}$ and $65\frac{1}{2}$, while lowering the penalty rate to 5%, increases both welfare and tax revenue as well as IRA balances at age $65\frac{1}{2}$. Similarly, we find that increasing the age for Required Minimum Distributions to be between $71\frac{1}{2}$ and $73\frac{1}{2}$, with a penalty rate of 50% or higher, increases both welfare and the present discounted value of lifetime income tax remittances.

Our results have three important implications for policy makers considering using these penalties as policy levers to increase retirement savings. First, there are combinations of age threshold and penalty rate for these penalties that increase both average welfare and tax remittances.

Second, encouraging individuals to keep their money in retirement savings accounts as long as possible increases welfare, as long as the penalty for early withdrawal isn't too high. Alternative strategies to assist individuals facing unexpected financial shocks before retirement should be considered in lieu of policies that make it easier for individuals to take early withdrawals from their retirement accounts.

Finally, the people who react to the early withdrawal penalty may differ from those who only take withdrawals because of Required Minimum Distributions. Intuitively, the early withdrawal penalty affects people who are trying to access their savings earlier in life, whereas the excess accumulation penalty affects individuals who want to enjoy the tax benefits for as long as possible. In addition, IRA holders have, on average, considerably higher incomes than non-IRA holders. If we want to create policy to help people save for retirement and we are worried that people with lower income might be more at risk for not having enough savings, then changing the policy levers associated with these accounts may not get us far enough. Incorporating extensive margin decisions into the model is a direction for future work.

CHAPTER II

Does Giving Tax Debtors a Break Improve Compliance and Income? Evidence from Quasi-Random Assignment of IRS Revenue Officers (with William C. Boning, Joel Slemrod, and Alex Turk)

2.1 Introduction

Millions of U.S. taxpayers do not fully remit their taxes each year, resulting in billions of dollars of uncollected tax debt. Attempts to collect this debt loom large in taxpayers' lives. If collecting debts from taxpayers whose economic resources are insufficient to meet basic living expenses reduces their incentive to work or invest, postponing collection could reap higher future repayments.

On the other hand, collecting tax debts deters tax evasion by increasing the effective penalties for failing to report taxes. Taking into account whether tax debts are collected adds nuance to the classic Allingham and Sandmo (1972a) model of tax evasion. While the original model assumes all debts are collected, this is not true in practice. If not all debts are collected, then changes in the collection rate affect both the expected present value of the tax remitted per dollar of reported income, and of the penalties remitted per dollar of unreported income uncovered in an audit. Taxpayers choose both how much income to report and how much tax to remit (and when), given penalties both for underreporting and underpayment.

There is little evidence about the consequences of collection attempts writ large, including on how pursuing or forgoing collection affects taxpayers' future compliance and income. One exception is Miller et al. (2014), who find that, conditional on a later suspension of collection, assigning a case to an IRS field collection agent ("Revenue Officer") earlier in the collections process is associated with larger amounts collected. Other work on tax debt collection finds that much smaller interventions, such as letters highlighting financial penalties (Cranor et al. (2020)) or potential social stigma (Perez-Truglia and Troiano (2018)), modestly increase payments. Miller et al. (2016) find that selling property tax liens to investors prompts higher payments, but less so in times of economic distress.

In this paper, we exploit an institutional feature of the debt collection process to provide evidence of the causal effects of tax collection on taxpayers' subsequent behavior. To provide causal evidence, we leverage the fact that the Internal Revenue Service (IRS) pauses collection when it determines that a taxpayer's income and assets are insufficient to meet basic living expenses, and designates their debt currently not collectible (CNC). The extent to which the IRS should grant such hardship designations is an open policy question. The National Taxpayer Advocate listed among 2018's most serious tax administration problems the IRS's failure to use available data to proactively designate additional cases CNC.

Because debt collection is suspended primarily on the basis of economic hardship, it is difficult to separate the effects of suspending collection efforts from the consequences of the economic hardship that qualifies a taxpayer's case for suspension. We use variation in IRS Revenue Officers' propensities to designate similar cases CNC to study how suspending collection affects taxpayer behavior. Our approach is a version of the examiner assignment design developed in Doyle (2007) and Doyle (2008), which used quasi-random appointment of child protection investigators to cases in order identify the causal effects of foster care placement.[1] To our knowledge, this is the first time an examiner assignment design has been applied to study the causal impact of a tax enforcement policy.

Our empirical strategy takes advantage of plausibly exogenous experience of tax debt relief resulting from the quasi-random assignment of delinquent cases to Revenue Officers. The process of determining whether a case should receive a CNC designation inherently requires the Revenue Officer to use their judgment. We gauge a Revenue Officer's propensity to designate a case CNC using a residualized "leave-one-out" measure based on the proportion of cases the Revenue Officer has determined to be CNC, adjusted for any observable characteristics of the case that could alter whether a CNC designation is appropriate.[2] This measure functions as an instrument for whether or not a case was actually deemed CNC. Tests of random assignment suggest that our instrument removes much, but not all, of the bias in naive OLS estimates. Our instrument is correlated with some pre-treatment covariates, suggesting that

[1]Examiner assignment designs have since been used to estimate local average treatment effects for a wide variety of policy interventions, including receipt of Social Security Disability Insurance (Maestas et al., 2013), bankruptcy protection (Dobbie and Song, 2015; Dobbie et al., 2017), incarceration (Kling, 2006; Loeffler, 2013; Mueller-Smith, 2015), pre-trial detention (Dobbie et al., 2018), and eviction (Humphries et al., 2019).

[2]Our definition of a CNC case includes any case in which some portion of the tax debt is designated CNC. There is little variation in the fraction of tax debt designated CNC: in more than 75% of cases in which any debt was designated CNC, the designation was applied to the entire debt.

the exclusion restriction may be violated. The imbalance is small relative to OLS estimates, but because even a small imbalance is cause for concern when interpreting the results of any instrumental variables approach, we adopt an instrumented difference-in-differences approach to address the remaining imbalance. As we do not observe trends in the instrumented difference-in-difference estimates in periods before treatment, we believe these estimates provide causal evidence of the effect of suspending collection efforts due to hardship on taxpayer compliance and income.

We analyze de-identified administrative data linking over 123,000 tax debt cases to information about taxpayers' incomes, tax return filings, and contemporaneous information about the Revenue Officers who worked each case. We track three measures of income (adjusted gross income, wages, and self-employment income) as well as payments to the IRS and future tax return filing. We estimate local average treatment effects of tax collection suspensions due to hardship stemming from variation in Revenue Officer assignment. In other words, we compare cases where Revenue Officers who differ in their inherent leniency might disagree on whether the taxpayer's hardship warranted the CNC designation.

We find that suspending collection due to hardship causes taxpayers to report larger adjusted gross income in future years. The increase in income is driven by an increase in wages earned by the spouse of the taxpayer with whom the case is associated. These results contrast with simple difference-in-difference estimates, which are subject to bias due to differential trends over time for taxpayers undergoing hardship. In those simple difference-in-difference estimates, hardship designations are associated with declines in debt repayment, payment of future years' tax liability, tax return filing, wages, and self-employment income. The instrumented estimates are less precise than the un-instrumented estimates, however, and in some cases confidence intervals for the instrumented estimates include the un-instrumented point estimates. While naive OLS estimates, which are subject to selection bias, indicate that a CNC designation is associated with a lower likelihood of filing a tax return, lower wages, and less other income, in our examiner assignment design we do not find significant effects of pausing collection on these outcomes. The IV point estimates for these outcomes are generally similar to the OLS point estimates, but larger standard errors from the IV approach eliminate any statistical significance.

The remainder of the paper proceeds as follows. Section 2.2 provides an overview of the collections process. Section 2.3 outlines our data and sample frame. Section 2.4 describes our identification strategy and research design. Section 2.5 presents the results, and Section 2.6 concludes.

2.2 The collections process

2.2.1 Delinquency designation

The IRS sends taxpayers who fail to file or to remit their known tax liability a bill notifying the taxpayer of their outstanding debt.[3] These notices are the beginning of the collection process.[4] Some taxpayers contact the IRS after receiving a notice in order to pay off some or all of the debt, to dispute it, or to explain that they are unable to pay. The IRS may negotiate an installment agreement or extend the due date for a taxpayer, depending on the taxpayer's circumstances. A small fraction of taxpayers settle their outstanding tax debt for less than the amount they owe through the Offer in Compromise (OIC) program or through "partial pay" installment agreements.[5]

When taxpayers do not respond to the initial (or subsequent) notices, their accounts are considered delinquent. At the end of Fiscal Year (FY) 2017,[6] there were approximately 14 million delinquent accounts with a total of over $131 billion owed in taxes, penalties, and interest (Internal Revenue Service (2017), Table 16). Delinquent accounts may be handled either by a call site (via the Automated Collection System (ACS)) or by a field office. In the first case, ACS personnel will try to contact the taxpayer by correspondence and by phone to negotiate a payment solution.[7] In the second case, a Revenue Officer from a local collection field office will work with the taxpayer to try to resolve the outstanding debt. Cases that are assigned to ACS may ultimately be transferred to a field office if its attempts to resolve the debt are unsuccessful.

2.2.2 Revenue Officer assignment

Our understanding of the process by which cases are assigned to Revenue Officers is based on the Internal Revenue Manual (IRM), which specifies IRS administrative procedures, as well as discussions with two group managers and a former Revenue Officer. These conversations highlighted that many factors play a role in case assignment, including some

[3]Third-party information reports provide information on the income of taxpayers who fail to file.

[4]This process is outlined at https://www.irs.gov/newsroom/the-collection-process and in Appendix B.1.

[5]In 2017, taxpayers made about 62,000 Offers in Compromise, of which the IRS accepted only 25,000 (Internal Revenue Service (2017), Table 16). These 62,000 offers are less than one half of one percent of all delinquent cases at the beginning of 2017. One condition of the IRS accepting less than the outstanding tax liability is that the taxpayer must remain compliant. "Partial pay" installment agreements occur when the taxpayer reaches an installment agreement with the IRS where the 10-year statute of limitations will run out on some of the debt before the taxpayer remits it. After the statute of limitations expires, the taxpayer is no longer responsible for the debt.

[6]The fiscal year for the IRS is the same as the United States federal government: Oct. 1 - Sept. 30.

[7]The IRS does not contact taxpayers by phone without first attempting to contact them by mail. See https://www.irs.gov/newsroom/phony-irs-calls-increase-during-filing-season.

based on professional judgment, despite the fact that the Internal Revenue Manual (1.4.50.10 9) explicitly states that "[p]rofessional judgment should play a limited role in case selection." A delinquent case is assigned to the local collection field office geographically closest to the taxpayer. The group manager of the office then assigns the case to a Revenue Officer. The first factor in case assignment is the grade of the case (9, 11, 12, or 13), which reflects the expected difficulty of closing the case. The case's priority is another factor in case assignment. Group managers are required to assign high priority cases before lower priority cases. High priority cases include those with particularly high balances, or where the statute of limitations for the debt is about to run out. Revenue Officers are more likely to receive new cases in zip codes where they already have active cases.[8] The group manager also has access to other aspects of the case that may impact assignment. These include the taxpayer's history of interacting with the IRS and factors influencing how likely the individual will be able to remit the debt (e.g., older taxpayers have fewer working years ahead of them and therefore may have lower potential earnings).

Case assignment also depends on the General Schedule (GS) grade of the Revenue Officer and the current inventory of the Revenue Officer.[9] Generally speaking, grade 09 cases are assigned to GS grade 09 Revenue Officers, grade 11 cases are assigned to GS grade 11 Revenue Officers, and so on. Revenue Officers with fewer workload-adjusted cases in their current inventories are more likely to receive new cases. Average caseload varies with GS grade. GS grade 09 Revenue Officers have an average case inventory of 72 cases. For GS grade 11 Revenue Officers, the average inventory is 61 cases. GS grade 12 and 13 Revenue Officers have average caseloads of 39 and 37 cases, respectively.

Once a case is assigned to a Revenue Officer, the Revenue Officer contacts the taxpayer, conducts research to ascertain the taxpayer's ability to make payments toward their outstanding debt, and takes steps to close the case. Depending on IRS guidelines and the Revenue Officer's judgement, possible steps range from seizures, liens, and levies to full payments, agreements in which the taxpayer will repay the debt in installments, and designating the case CNC, which suspends collection efforts. More than two-thirds of cases close through full payment, an installment agreement, or a CNC designation, as Table 2.1 shows. Cases may be deemed CNC for a variety of reasons, including hardship and inability to locate the taxpayer. In some cases, a taxpayer may pay off some portion of their debt and

[8] Instances when professional judgment is appropriate in case selection include efficiently allocating the resources of the Service (e.g., geographically clustering cases to minimize Revenue Officer driving time) and addressing "the developmental needs of the Revenue Officer" (i.e., making sure the officer is exposed to a variety of case types to develop expertise in order to meet standards for promotion).

[9] The General Schedule (GS) Pay Scale is a wage schedule for federal employees. Revenue Officers may have a GS grade of 05, 07, 09, 11, 12, or 13. Grades 05 and 07 are related to training. Starting in 2017, the IRS discontinued use of the GS 09 grade for Revenue Officers.

the remaining debt may be deemed CNC, for example, due to hardship.

Table 2.1: Distribution of how cases were closed

Description	Number of cases	Percent of cases
Installment agreement	49,041	34.1%
CNC	43,783	30.5%
Other	40,396	28.1%
Full pay	10,443	7.3%
Total	123,396	100.0%

Notes: Includes all cases that meet our sample restriction criteria (discussed in Section 2.3) between November 2014 and December 2018. Other ways a case might be closed include Offers in Compromise, abatement, payment tracer (which resolves situations of missing and misapplied payment issues), and being flagged for handling outside of the usual system and procedures.

2.2.3 Currently Not Collectible (CNC)

The vast majority of CNC cases for individual taxpayers are either "hardship" (49%) or "unable to locate" (39%). Per the Internal Revenue Manual, "[a] hardship exists if a taxpayer is unable to pay reasonable basic living expenses" (IRM 5.16.1.2.9(1)). Outstanding tax liability that is designated CNC is still legally owed to the United States government.[10] When a taxpayer's debt is deemed CNC, the IRS sends the taxpayer a letter notifying them of the change in status and reminding them that the taxpayer must pay future tax liability. Changes in the taxpayer's circumstances may cause the IRS to re-evaluate whether a CNC designation is appropriate.[11] For example, when taxpayers with unpaid tax that is designated CNC file returns, the designation may be rescinded if the taxpayer's income becomes sufficiently large. Interest and penalties continue to accrue on the outstanding balance while collection efforts are suspended. If the taxpayer is owed a refund on a future tax return, the IRS will retain the refund to offset the tax debt in CNC status. Even if a taxpayer has outstanding tax debt in CNC status, new unpaid tax undergoes the notification process described in Section 2.2.1.

A priori reasoning does not lead to clear predictions about the direction of changes in tax compliance behavior by those whose debt is designated CNC. A CNC designation suspends

[10]The statute of limitations on unremitted debt is ten years, after which the IRS may no longer pursue the debt.

[11]Receipt of future returns from the taxpayer automatically initiates a review to see whether the taxpayer's circumstances have changed (e.g., a change in address, income, a new levy source). If there has been a change, the case can be re-activated and the CNC status may be revoked. When this happens, the case will either be sent back to ACS or directly to a group manager's queue. The taxpayer is not directly informed that their case has been re-activated, but the taxpayer may receive, e.g., a letter from ACS or a notice of a levy, which would indicate that the IRS has resumed pursuit of the outstanding balance.

efforts to collect some or all outstanding tax debt, such as attempts to garnish wages and assets, and thereby reduces the incentive for the taxpayer to voluntarily repay outstanding debt. In addition, a CNC designation signals that the IRS can be lenient, and might lead the taxpayer to conclude that future noncompliance will also be met with leniency. This reduction in the expected costs of noncompliance has a negative effect on compliance. In contrast, maintaining a CNC designation requires that a taxpayer comply with the requirements to file tax returns and remit tax payments for new tax years liability. Failure to comply with these requirements leads to the resumption of active collection, resulting in a positive effect on compliance from a CNC designation.

The effect of a CNC designation on the incentive to earn income is also theoretically unclear. Having some or all tax debt designated CNC provides a positive wealth effect, as it reduces the expected present value of tax liability. A permanent CNC designation would reduce the taxpayers expected tax rate, because future earnings and assets will not be garnished to meet the tax debt. However, a CNC designation is not permanent (unless the 10-year statute of limitations is reached), and the IRS retains the right, if the taxpayers financial situation improves sufficiently, to put the debt back in active collection status. Thus, a taxpayer may perceive that his or her marginal tax rate is reduced by the CNC designation for future incomes up to some unknown level that triggers the CNC status revocation, above which the marginal tax rate absent the CNC designation is restored. The combination of a negative effect on future earnings from the wealth effect and a generally positive substitution effect means that the net effect on earnings is of indeterminate sign.

Consequently, both the magnitude and the sign of the impact of a CNC designation on future behavior can only be determined via empirical analysis. We turn next to this task.

2.3 Data and sample frame

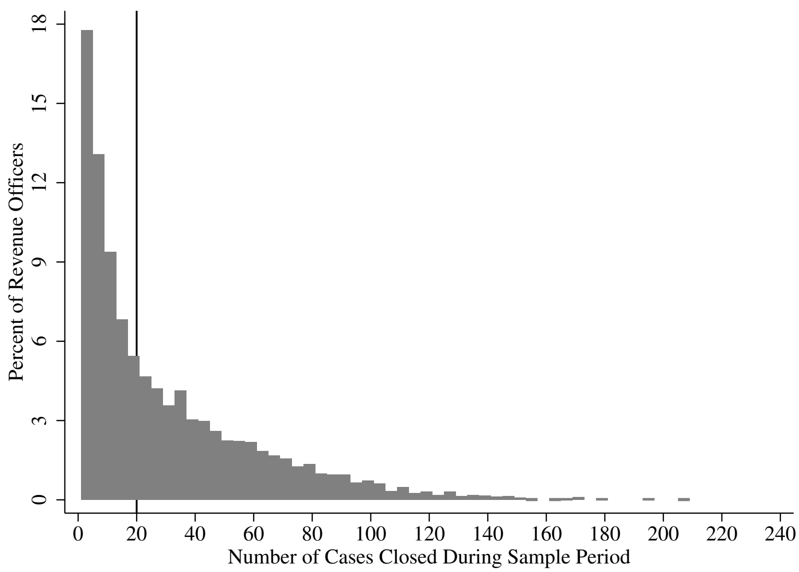
To assess the effects of decisions made during the collection process, we combine information from the collection case management system with caseload information about Revenue Officers and administrative taxpayer information, including a history of filing status (i.e., whether or not the taxpayer filed a return), monthly data on unpaid tax liability, and information from individual tax returns. Data on Revenue Officers include their GS grades, inventory levels at the end of each month, and their decisions in past collection cases. We supplement the taxpayer data with information about local allowable living expenses and whether they live in an urban or rural area.[12]

We focus on individual taxpayer cases and therefore exclude business entities other than

[12] Appendix B.2 explains the construction of these variables and the outcome variables in detail.

sole proprietorships. We begin with about 283,000 individual taxpayer cases. Approximately 123,000 taxpayers remain after excluding cases where we suspect case assignment may not have been random,[13] cases designated CNC because the Revenue Officer was unable to locate the taxpayer, and cases worked by Revenue Officers who handled fewer than 20 cases between November 2014 and December 2018.[14] We perform this final exclusion to ensure that we observe sufficient decisions made by each Revenue Officer to obtain a reliable estimate of their tendency to designate a case CNC. Figure 2.1 shows the distribution of the number of cases closed by a Revenue Officer between November 2014 and December 2018. Out of 4,808 Revenue Officers included, 2,345 (49%) meet our sample restriction criteria.

Figure 2.1: Distribution of number of cases closed per Revenue Officer over the sample period



Notes: 4, 808. Includes all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

2.3.1 Field data

To assemble our sample for analysis, we begin with the universe of cases with an assessed balance due assigned to a field collection Revenue Officer between November 2014 and December 2018. The data include the numeric ID of the Revenue Officer who worked each

[13]We discuss situations when case assignment may not be random in Section 2.3.1.

[14]The mean (standard deviation) fraction of cases designated CNC is 0.24 (0.24) among Revenue Officers who worked fewer than 20 that meet our selection criteria, and 0.30 (0.12) among Revenue Officers who worked at least 20 cases that meet our selection criteria. As a robustness check, we repeat our analysis using alternative cut-offs of 10 cases and 30 cases, and the results are qualitatively unchanged. Appendix B.3 shows the IV results for cut-offs of 10, 20, and 30 cases.

case, the date the case was assigned to field, the date the case was assigned to the Revenue Officer, and the date the case was closed. We drop cases where it appears that the Revenue Officer reported working zero hours, cases that were not closed, cases where the taxpayer lives outside of the United States or on a military base, and cases flagged as handled outside the usual system and procedures. Because we are ultimately interested in the impact of a CNC designation on future outcomes, we drop cases designated CNC because the Revenue Officer was unable to locate the taxpayer.

In addition, we exclude certain categories of cases in which the Revenue Officer's manager might exercise deliberate discretion in the assignment process. These categories were identified based on interviews about the case assignment process with field collection group managers and an individual who worked for several years as a Revenue Officer. Cases where the statute of limitations is close to expiration are dropped,[15] as these cases may be assigned to a Revenue Officer who works quickly. We use only the first time we observe a case being assigned within a given group, which eliminates cases where there is already an existing case involving the same taxpayer. New cases in a group manager's queue that involve the relevant taxpayer in an ongoing case are usually assigned to the Revenue Officer working the existing case. This also removes cases that have been returned to the group's queue. The group manager is better able to anticipate the type of work the case will require when cases return to the group's queue, particularly if the case is an installment agreement default or a review of a previous CNC decision, which may influence the assignment decision.[16] Finally, we exclude cases that are reassigned to a different Revenue Officer within the group (or in a different group), as reassignment may reflect unobserved characteristics of the case. In contrast, the group manager assigns cases that the group has not seen before on the basis of limited information about the case and the group's Revenue Officers that is observable in administrative databases.

2.3.2 Revenue Officer data

The Revenue Officer data includes monthly-level information about factors used to determine the workload to assign to the Revenue Officer, including the Revenue Officer's GS grade, monthly inventory and inventory by case grade, and the fraction of the Revenue Officer's case load that is above their GS grade level. We match this data to the field data using the Revenue Officer id number and relevant month: if a case in the field data was

[15]We consider the statute of limitations close to expiration if the expiration date is within a year.

[16]As is standard in literature using examiner assignment to identify causal estimates, we do not exclude individuals who appear more than once in our data but are assigned to a different local IRS office, or "group," who, because they are in a different location, may not have a previous history with the taxpayer. Less than 3% of our sample appear in the data more than once.

assigned to a Revenue Officer in, e.g., May 2017, we match the Revenue Officer's information from the previous month.

We make two additional exclusions from the data based on the Revenue Officer data. First, we exclude Revenue Officers whose GS grade is listed 5 or 7, which are training grades. Second, we exclude group-month combinations when only a single Revenue Officer in the group is assigned any cases. This second exclusion is necessary because there will be no variation in proclivity to designate a case CNC in that group during that month.

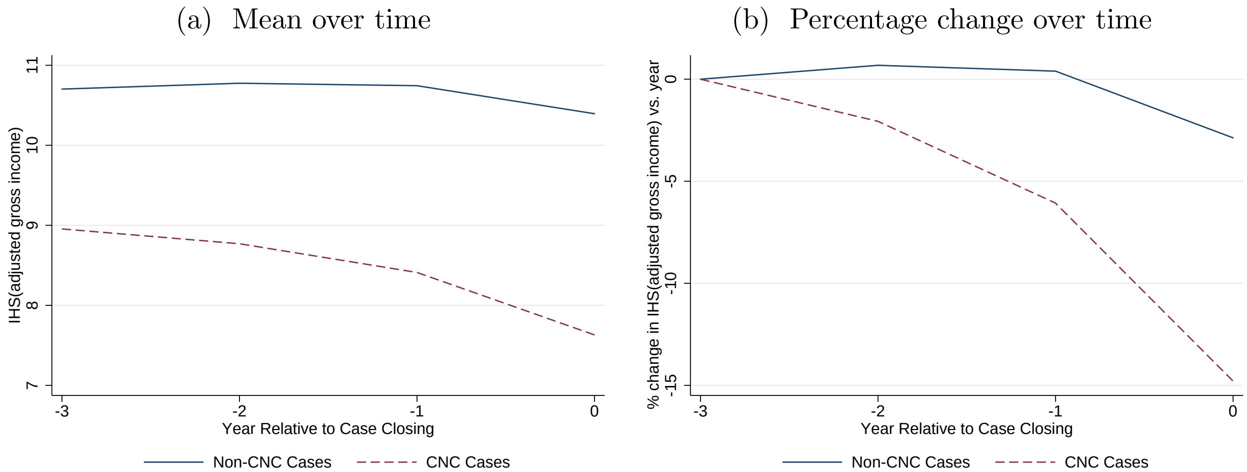
2.3.3 Taxpayer data

Our taxpayer panel spans January 2009 to December 2018. This panel contains the taxpayer's year of birth, outstanding tax debt and payments, information from annual tax returns, and information about the taxpayer's and spouse's income from third-party information reports. We measure behavior at the level of the taxpaying unit, combining the third-party-reported values for the taxpayer and, if the taxpayer filed a joint tax return in the year the case closed, the spouse listed on that return. The measures of tax compliance we construct are payments toward outstanding tax debt, payments toward current-year tax liability (including withholding from wages reported on Form W-2, quarterly estimated tax payments, and payments remitted with the tax return), an indicator for remitting all of the current year's tax liability, and an indicator for filing a return conditional on having third-party-reported wages. Our measures of income are adjusted gross income, total W-2 wages, W-2 wages broken out by whether the taxpayer or their spouse was the earner, and self-employment income. We take the inverse hyperbolic sine transformation of each of our measures of income. We then merge this panel with the field and Revenue Officer data described above.

2.4 Identification strategy

The fundamental challenge in assessing the causal effect of a CNC designation on taxpayer outcomes is disentangling the effect of the CNC designation from the effects of the hardships that might lead to such a designation. For example, a sole proprietorship might become unprofitable, fail to pay taxes on the prior year's income, and then close, leaving the proprietor with substantial tax debt and little income. In this example, determining the effect of a subsequent CNC designation is complicated by the effect of the business failure itself. Taxpayers whose cases are designated CNC face substantially larger decreases in adjusted gross income in the years leading up to their cases being closed (shown in Figure 2.2). This means that taxpayers whose cases are not designated CNC are, as a whole, a poor comparison

Figure 2.2: IHS(adjusted gross income) over time (relative to year case closed)



Notes: N = 138,452. Panel (a) shows the mean of the inverse hyperbolic sine (a transformation similar to the natural logarithm) of adjusted gross income over time across two subgroups: tax units whose cases were designated CNC in dashed red and tax units whose cases were not designated CNC in solid blue. The x-axis indicates the year relative to when each case closed, e.g., -1 is the year before the taxpayer's case closed. Panel (b) shows cumulative percentage changes in this measure relative to the value three years before the case closed. Adjusted gross income adjusted for inflation to 2017 values.

group for taxpayers receiving a CNC designation.

We first attempt to address this sample selection issue with an instrumental variable. As described in Section 2.4.1, we use variation in the propensity of the Revenue Officer assigned to a case to designate other cases CNC as an instrument for CNC designation. We test the exogeneity of this instrument by examining whether pre-treatment characteristics are balanced across cases with high and low values of the instrument. We find that even when using the instrument, some imbalance remains. We address this remaining imbalance and related concerns about instrument exogeneity with a difference-in-differences instrumental variables approach, discussed in Section 2.4.2.

2.4.1 Source of variation and instrument

Variation in CNC status across otherwise similar cases comes from randomness in which Revenue Officer is assigned to a case, combined with differences in Revenue Officers' propensity to designate cases CNC. An intuitive method for estimating Revenue Officer j's propensity to designate cases CNC would be to determine what proportion of cases Revenue Officer j designated CNC. Simply using this measure to predict whether Revenue Officer j designated case i as CNC would be biased, however, since case i would have been included in the calculation. To avoid this bias, we construct a case-specific instrument using a "leave-one-out" measure of the propensity of Revenue Officer j to designate case i CNC.[17] Specifically, the "leave-one-out" measure is

$$Z_{ij}^{S} = \frac{\sum_{k=1}^{n_j} CNC_{kj} - CNC_{ij}}{n_j - 1}. \tag{2.4.1}$$

The numerator of this expression is the number of cases designated CNC by Revenue Officer j, less one if case i was designated CNC (i.e., $CNC_{ij} = 1$). The denominator is the total number of cases handled by Revenue Officer j less one.

As described previously, case assignment is not always random. Cases are assigned to Revenue Officers based on case grade, case priority, taxpayer characteristics, and geographic considerations. For example, if higher grade cases are less likely to receive a CNC designation, then failing to control for case grade and Revenue Officer grade would make it seem like higher GS grade Revenue Officers were stricter. We follow Dobbie et al. (2018) and develop a residualized leave-one-out measure of the propensity to designate a case CNC that removes variation from our instrument that is driven by observable determinants of case assignment that may also impact CNC designation. As a result, our instrument is a measure of the Revenue Officer's tendencies rather than features of the case that may be correlated with a CNC designation. We discuss the empirical importance of residualization and potential pitfalls in Section 2.4.1.1.

We calculate the residualized leave-one-out instrument as follows. First, we regress true CNC status for each case on Revenue Officer and case characteristics that may impact case assignment:

$$CNC_{ij} = \beta_0 + \beta \boldsymbol{X}_{ij} + u_i, \tag{2.4.2}$$

where $\boldsymbol{X}_{ij}$ includes an indicator for whether or not the case is high priority, case grade fixed effects, and case characteristics (the taxpayer's year of birth, estimated ability to pay,[18] an indicator of whether the taxpayer has had their case "in the field" since 2009, and indicator variables for whether the oldest debt associated with the case is more than 12 months old or more than 36 months old.)[19]

We include fixed effects for the group of the assigned Revenue Officer.[20] To see why this

[17]This method is used in, e.g., Doyle, Jr. (2007), Doyle, Jr. (2008); Maestas et al. (2013); Dobbie and Song (2015); Dobbie et al. (2017). Other approaches to eliminating this bias include Jackknife IV (Angrist et al., 1999), split-sample two-stage IV (Angrist and Krueger, 1995), and limited-information maximum likelihood.

[18]"Estimated ability to pay' is the ratio of an estimated future income based on their previous adjusted gross income and age, and their outstanding balance with the IRS.

[19]Appendix B.2 explains how we derived the specific variables used in our analysis and provides sample statistics by CNC status.

[20]Groups are assigned cases from particular ZIP Codes in their local area. We are unable to control

might matter, consider two groups. Group A works cases in affluent Township A whereas Group B works cases in low-income Village B. We might expect that the cases in Village B are more likely to face a true financial hardship, and therefore more likely be given a CNC designation. We want to make sure the residual propensity to designate a case CNC reflects characteristics of the Revenue Officer rather than characteristics of the taxpayers who live where the Revenue Officer's cases take place. We also control for two relevant geographic variables: allowable living expenses and an urban indicator. In robustness checks we include total inventory and case grade by Revenue Officer GS grade fixed effects, which does not qualitatively change the results.

We then use the results of estimating Equation 2.4.2 to predict CNC status:

$$\widehat{CNC}_{ij} = \hat{\beta}_0 + \hat{\beta}\boldsymbol{X}_{ij}, \tag{2.4.3}$$

and calculate the residual value of CNC_i^R by subtracting $\widehat{CNC}_{ij}$ from CNC_i:

$$CNC_{ij}^R = CNC_{ij} - \widehat{CNC}_{ij}. \tag{2.4.4}$$

The case-specific, residualized leave-one-out measure of Revenue Officer j's propensity to designate case i as CNC is computed using Equation 2.4.5:

$$Z_{ij}^R = \frac{\sum_{k=1}^{n_j} CNC_{kj}^R - CNC_{ij}^R}{n_j - 1}. \tag{2.4.5}$$

The numerator of this expression is equal to the sum of the residualized CNC designation for all cases covered by Revenue Officer j less the residualized CNC designation for case i. The denominator is equal to the total number of cases handled by Revenue Officer j less one. Intuitively, this measure indicates Revenue Officer j's residual propensity to designate cases as CNC (excluding case i), holding constant observed characteristics of the case. Conditional on observable characteristics and assuming conditional random assignment, Z_{ij}^R should be correlated with the decision in case i only if Revenue Officer j has a threshold level of leniency specific to case i.

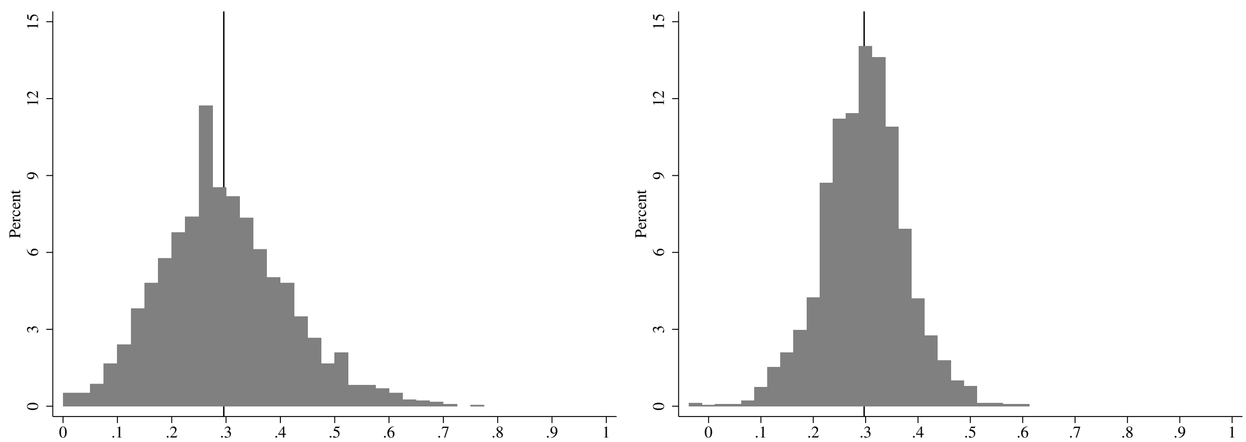
The distributions of the simple leave-one-out instrument, Z_{ij}^S, and the residualized leave-one-out instrument, Z_{ij}^R, are different; crucially, the distribution of the residualized instrument is less dispersed. Our residualized leave-one-out instrument has an average value of 0.0002 and a standard deviation of 0.0826, while the simple leave-one-out instrument has a mean of 0.3024 and a standard deviation of 0.1119. This is illustrated in Figure 2.3, where

directly for ZIP Code fixed effects because there are two few cases per zip code.

Figure 2.3: Distribution of the fraction of cases designated CNC by Revenue Officer

(a) Distribution of the fraction of cases designated CNC by Revenue Officer (Z_{ij}^{S*})

(b) Adjusted distribution of the fraction of cases designated CNC by Revenue Officer (Z_{ij}^{R*})



Notes: Unique at the Revenue Officer level ($N = 2,345$). Constructed using all cases closed by Revenue Officers between November 2014 and December 2018 who closed at least 20 cases that meet our sample restriction criteria during that time period. Panel (a) shows the fraction Z_{ij}^{S*} of cases designated CNC by each Revenue Officer, where Z_{ij}^{S*} is defined in Equation 2.4.6. Panel (b) shows the adjusted fraction R_{ij}^{S*} of cases designated CNC after accounting for residualization, as defined in Equation 2.4.7.

Panel (a) shows how much more variation there is in the value of the simple leave-one-out instrument compared to the residualized leave-one-out instrument in Panel (b). The figure plots a slightly adjusted version of the instrument to include each Revenue Officer only once.[21]

2.4.1.1 Tests of instrument exogeneity

Using Revenue Officer assignment to instrument for CNC status relies on the assumption that Revenue Officer assignment is uncorrelated with future outcomes, conditional on included covariates. This assumption allows for assignment of Revenue Officers based on certain

[21]Panel (a) shows the fraction Z_j^{S*} of cases designated CNC by each Revenue Officer j, where

$$Z_j^{S*} = \frac{\sum_{n_j}^{k=1} CNC_{kj}}{n_j}. \tag{2.4.6}$$

Z_j^{S*} is equivalent to Z_{ij}^{S} (calculated in Equation 2.4.1) without omitting case i. Similarly, Panel (b) shows the adjusted fraction Z_j^{R*} of cases designated CNC by each Revenue Officer j, where

$$Z_j^{R*} = \frac{\sum_{n_j}^{k=1} \widehat{CNC_{ij}}}{n_j}. \tag{2.4.7}$$

Z_j^{R*} is equivalent to $Z_j^{S*} - \frac{1}{n_j}\sum_{n_j}^{k=1} CNC_{kj}^{R}$, where CNC_{ij}^{R} is calculated as in Equation 2.4.4. We plot Z_j^{R*} rather than $\frac{1}{n_j}\sum_{n_j}^{k=1} CNC_{kj}^{R}$ to facilitate a visual comparison: the resulting object has the same mean as true CNC status, as is illustrated by the vertical lines in the figure.

observable characteristics (such as case grade) but, once those observable characteristics are controlled for, Revenue Officer assignment should be random. As mentioned previously, group managers are explicitly instructed that "[p]rofessional judgment should play a limited role in case selection" (IRM 1.4.50.10 9).

Our discussions with group managers underlined certain aspects of cases that might result in group manager professional expertise playing a larger role in case assignment. Some of these factors are not included in the case data and therefore cannot be included in the creation of our instrument, raising the concern that the instrument may not be uncorrelated with future outcomes. We test whether Revenue Officer assignment is conditionally random by comparing pre-assignment case and taxpayer characteristics across Revenue Officers with high and low values of the instrument. We regress pre-assignment characteristics on, in turn, an indicator for CNC designation, the simple instrument (Z^S_{ij}, defined in Equation 2.4.1), and the residualized instrument (Z^R_{ij}, defined in Equation 2.4.5), following Equation 2.4.8:

$$Y_i = \eta_0 + \eta Z_i + u_i, \tag{2.4.8}$$

where Y_i is the pre-assignment outcome for taxpayer i and Z_i is the value of the relevant CNC indicator or instrument.

We test for balance using five pre-assignment characteristics of the case:[22] model score (an estimate of how likely the taxpayer is to remit their outstanding liability), the taxpayer's average wages (from Forms W-2 provided by employers), an indicator for whether the taxpayer filed a return in the year before their case was assigned to a Revenue Officer, and, if the taxpayer did file a return, the adjusted gross income reported.[23] We transform W-2 wages and adjusted gross income by taking the inverse hyperbolic sine, which can be interpreted like a log transformation and enables us to include non-positive values.

As one would expect, cases that were designated not collectible had lower model scores (indicating a lower estimated probability of collection), lower wages, lower adjusted gross income, and were less likely to file before assignment. These associations, shown in Table 2.2, are statistically significant at the 0.1% level.

The simple leave-one-out average of Revenue Officer CNC designations does not solve our

[22]We cannot use the variables included in the creation of our residualized leave-one-out instrument, all of which were chosen based on our reading of the Internal Revenue Manual and our conversations with group managers and an ex-Revenue Officer.

[23]One potential concern arises with using average adjusted gross income to calculate our estimate of the taxpayer's ability to pay off their balance, one of the variables used in the residualization process. This means that adjusted gross income appears on both the left- and right-hand sides of the regression for two of our balance tests. Any bias this step in the residualization might introduce is of minimal concern because the coefficient on these outcomes is statistically insignificant regardless of whether or not we residualize the instrument.

Table 2.2: Test of instrument exogeneity

	Model Score	Ave. W-2 Wages (IHS)	Ave. AGI (IHS)	AGI before assignment (IHS)	Filed before assignment
OLS	-0.065***	-1.913***	-1.505***	-2.096***	-0.160***
SE	(0.001)	(0.037)	(0.030)	(0.037)	(0.003)
F-stat	2,263.104	2,744.591	2,527.897	3,264.277	2,369.145
IV (Simple)	-0.161***	-2.057***	-1.245***	-1.727***	-0.167***
SE	(0.012)	(0.225)	(0.240)	(0.265)	(0.018)
F-stat	188.746	83.554	26.858	42.613	89.157
IV (Resid.)	-0.033*	-0.822**	-0.372	-0.569*	-0.040
SE	(0.014)	(0.274)	(0.224)	(0.256)	(0.024)
F-stat	5.679	9.022	2.767	4.925	2.853
N	91,112	120,753	117,289	116,803	120,754

Notes: Standard errors, in parentheses, are clustered at the Revenue Officer level. Includes all cases that meet our sample restriction criteria between November 2014 and December 2018. Limited to cases worked by Revenue Officers who closed at least 20 cases that met our sample restriction criteria between November 2014 and December 2018. Variables are defined in Appendix B.2. Values for average wages, average adjusted gross income, and and adjusted gross income in the year before case assignment are given using the inverse hyperbolic sine transformation. We measure adjusted gross income and W-2 wages at the level of the taxpaying unit: if a taxpayer filed a joint tax return during the year in which their case was closed, we add the value for the taxpayer's spouse to the taxpayer's own value. If the taxpayer did not file a joint tax return in that year, the outcome value is equal to the value for the taxpayer alone.
*** Significant at the 0.1% level; ** significant at the 1% level; * significant at the 5% level.

sample selection problem. All of the coefficients continue to be significant at the 0.1% level, and the magnitude of the coefficients hardly change (the exception being the coefficient on model score, for which the magnitude of the coefficient increased substantially).

The residualized Revenue Officer instrument reduces the imbalance but does not fully correct it. The magnitudes of the coefficients for all five pre-treatment characteristics shrink considerably. The coefficients between the residualized instrument and model score, and between the residualized instrument and pre-assignment adjusted gross income, are only statistically significant at the 5% level. The coefficient between the residualized instrument and average wages is statistically significant at the 1% level. While any violation of the exclusion restriction is cause for concern, these coefficients are small in economic terms. Assignment to a Revenue Officer whose value of the instrument is one standard deviation higher is associated with a model score that is, on average, 0.0027 lower, 6.7% lower wages, and 4.7% lower adjusted gross income.[24] As the signs of the correlations with the residualized

[24]The standard deviation of our residualized instrument is 0.082 and the coefficient on average wages in the balance test is -0.822: $-0.822 \times 0.082 = -0.067$ log points, about 6.7%. Similarly, the coefficient on adjusted gross income before assignment is -0.569 ($-0.569 \times 0.082 = -0.047$), and the coefficient on model score is -0.033 ($-0.033 \times 0.082 = -0.0027$).

instrument are the same as the signs of the correlations with CNC status, one would expect any bias in the IV results to be toward the OLS estimates. Concerns about the remaining imbalance motivate our use of the difference-in-differences instrumental variables design detailed in Section 2.4.2.

2.4.2 Difference-in-differences instrumental variables design

We employ a difference-in-differences specification in which we instrument for CNC status with the residualized Revenue Officer assignment instrument.[25] We adopt the event-study version of difference-in-differences, where coefficients on treatment are estimated separately for each year relative to the year the case closes. In the first stage, we estimate the equation:

$$CNC_{it} = \sum_k 1(t = k)[\alpha_k Z^R_{ij} + \gamma_k \boldsymbol{X}_{it} + \delta_k] + \zeta_i + \eta_{it}, \tag{2.4.9}$$

where CNC_{it} is equal to one if the case is ever designated CNC and zero otherwise,[26] k indexes years relative to case closing, Z^R_{ij} is the residualized Revenue Officer instrument, $\boldsymbol{X}_{it}$ are case-specific variables that may influence Revenue Officer assignment, and η_{it} is the error term. This equation instruments for CNC status with the Revenue Officer instrument on a per-period basis, producing one instrument and one instrumented value of CNC for each year relative to case closing.

Table 2.3: First stage results

	(1)
Residualized Revenue Officer instrument	0.509***
	(0.018)
Controls * Year FE	Yes
N	936,839
F-Statistic	756

Notes: Standard errors, in parentheses, are clustered at the Revenue Officer level. Includes all cases that meet our sample restriction criteria between November 2014 and December 2018. Limited to cases worked by Revenue Officers who closed at least 20 cases that met our sample restriction criteria between November 2014 and December 2018. Descriptions of the outcome variables and variables included in residualization are included in Appendix B.2. Controls include a high priority indicator, case grade fixed effects, an urban indicator, an estimate of ability to pay, year of birth, an indicator for previously assigned to field, indicators for debt older than 12 and 36 months, and group fixed effects, all of which are interacted with a full set of indicators for year relative to case closing.
*** Significant at the 0.1% level; ** significant at the 1% level; * significant at the 5% level.

[25] This approach is similar in spirit to, among others, Duflo (2001).

[26] While CNC_{it} is not time varying, the continuous predicted value we construct, $\widehat{CNC}_{it}$, can vary by time within i due to the effect of changes in $\boldsymbol{X}_{it}$.

To generate our estimates of the effect of suspending tax debt collection due to hardship, we suppose that taxpayer behavior is determined as follows:

$$Y_{it} = \sum_k 1(t=k)[\pi_k CNC_{it} + \theta_k \boldsymbol{X}_{it} + \mu_k] + \kappa_i + \epsilon_{it}, \tag{2.4.10}$$

where Y_{it} is the outcome variable, μ_k are event-time fixed effects, and ϵ_{it} is the error term. We replace CNC_{it} with the predicted values $\widehat{CNC}_{it}$ from Equation 2.4.9. The resulting second stage relationship we estimate is then

$$Y_{it} = \sum_k 1(t=k)[\phi_k \widehat{CNC}_{it} + \tau_k \boldsymbol{X}_{it} + \nu_k] + \iota_i + e_{it}. \tag{2.4.11}$$

This design relaxes the instrumental variables conditional exogeneity assumption, replacing it with exogeneity conditional on case and event-year fixed effects and interactions between the pre-treatment control variables and event-year fixed effects. While relying on this weaker assumption has the potential to address the imbalances described above in Section 2.4.1.1 (which invalidate the simple instrumental variables assumption), there is no guarantee that we are able to control for all of the time-varying factors associated with the instrument that could affect the outcome. We test whether this alternative exogeneity assumption holds by including periods before case closing in our regressions. If the assumption holds, coefficients on the instrumented value of CNC should be zero before case closing. While this test is informative, it cannot guarantee that similar relationships would have held after case closing absent CNC status.[27,28]

[27] Another possible violation of the conditional exogeneity assumption would occur if Revenue Officers who are more or less lenient in their propensity to designate cases CNC also systematically differ in their approach to cases in other ways, for example in their attempts to collect tax debts prior to making a CNC determination. Such differences would likely arise prior to case closing, so our test of trends before closing is also informative about this potential confounding treatment.

[28] The additional assumptions in an instrumental variables specification are relevance, monotonicity, and the exclusion restriction. Relevance means that Revenue Officer assignment must be correlated with CNC designation (formally, $\mathbb{C}\text{ov}[CNC_{ij}, Z_{ij}] \neq 0$). The correlation between CNC status and the residualized leave-one-out instrument is 0.1702. In our setting, monotonicity means that cases designated CNC by low-propensity (high-σ_j) Revenue Officers would always have also been designated CNC by high-propensity (low-σ_j) Revenue Officers. Similarly, cases that were not deemed CNC by high-propensity (low-σ_j) Revenue Officers would also not have been deemed CNC by high-propensity (low-σ_j) Revenue Officers. This assumption would be violated if, e.g., some Revenue Officers were more likely to designate cases as CNC if the taxpayers were older, while some Revenue Officers were more inclined to designate cases as CNC if the taxpayer were younger. Finally, the exclusion restriction means Revenue Officer propensity to designate a case CNC must only affect taxpayer outcomes through the variation in having a case designated CNC, that is, $\mathbb{C}\text{ov}[Z_i, h_i] = 0$. This assumption would be violated if Revenue Officer propensity to designate a case CNC is correlated with unobservable determinants of future taxpayer outcomes. This assumption is also violated if Revenue Officer propensity to designate a case CNC impacts future taxpayer outcomes through means other than CNC designation (e.g., if Revenue Officers that are more willing to deem a case CNC are also more likely to provide information to help a taxpayer avoid being in this situation in the future, or make more intense attempts to

2.5 Effects of Suspending Tax Debt Collection Due to Hardship

We next consider how suspending tax debt collection affects measures of taxpayer compliance and income. We measure behavior at the level of the taxpaying unit: if a taxpayer filed a joint tax return during the year in which their case was closed, we add the value for the taxpayer's spouse to the taxpayer's own value. If the taxpayer did not file a joint tax return in that year, the outcome value is equal to the value for the taxpayer alone.

We show the results from both an uninstrumented difference-in-differences regression and from difference-in-differences regressions instrumenting for treatment with the Revenue Officer instrument. We include cases worked by Revenue Officers who closed at least 20 cases during our sample period.[29] The IV difference-in-difference results show the effect of a CNC designation on cases for which the CNC designation hinged on the Revenue Officer assigned. Despite a strong first-stage relationship between the Revenue Officer instrument and CNC status, these estimates are far less precise than simple difference-in-difference results because only a fraction of cases hinge on the instrument, removing much of the effective sample size.

We set the year before the year in which the case closed as Year 0. Coefficients are reported relative to Year 0. This ensures that all behavioral changes due to CNC status are properly included in the post period. For example, if a case closed in January of Year T, it may impact whether the taxpayer filed their Year T-1 tax return by April of Year T. One consequence of this approach is that behavioral responses may appear delayed: if a case closes in December, only a small fraction of Year 1 is post-case closing, and the first full year in which CNC status can affect behavior is Year 2.

We estimate effects for years from two years before Year 0 (Year -2) to four years after Year 0 (Year 4). Years -2 and -1 allow us to visually assess whether or not the parallel trends assumption holds before treatment. Coefficients in the years before treatment are not statistically significant, except for in Year -1 for two related outcomes discussed in Section 2.5.1.2 below.

collect unpaid tax before designating a case CNC). Although the exclusion restriction is ultimately untestable, it is reasonable to assume that Revenue Officers are not systematically differentially interacting with taxpayers whose cases they may or may not ultimately designate CNC.

[29]There is a critical trade-off in choosing this threshold. As shown in Figure 2.1, the number of Revenue Officers excluded from our analysis quickly increases as we raise the case count threshold. This reduction in sample size reduces the power of the estimates. Lowering the case count threshold means that we include Revenue Officers for whom we have a less precise measure of their latent tendency to designate cases CNC. Results are similar if we use alternative case count thresholds of 10 or 30 cases, though standard errors are slightly different.

2.5.1 Effects on compliance

If taxpayers whose cases are deemed CNC come to believe that their tax compliance is subject to less scrutiny than they had previously thought, they may be less compliant in the future. This reduced compliance could take the form of reduced payments to the IRS, or a lower probability of filing a tax return. On the other hand, if taxpayers try to avoid jeopardizing their CNC status and triggering renewed attempts to collect, they may be more compliant in the future.

In this section, we show the effect on marginal cases of a CNC designation for three measures of compliance: payments toward outstanding tax debt (i.e., the debt that caused the taxpayer's case to be assigned to a Revenue Officer), payments toward the current tax year's tax liability, and whether or not the taxpayer filed a return conditional on receiving W-2 wages. We find that a CNC designation leads to increased withholding from wages, but has no discernible effect on other, more voluntary payments to the IRS for either past or current tax liability and no effect on tax filing.

2.5.1.1 Payments toward outstanding tax debt

Figure 2.4 shows changes in the dollar value of payments against outstanding tax debt made by taxpayers whose cases are designated CNC compared to taxpayers whose cases are not.[30] A standard difference-in-differences regression implies that payments from taxpayers whose cases are designated CNC fall substantially relative to other taxpayers with a decline that is largest in Year 2 before recovering partially. The IV regression results display a similar pattern in their point estimates but have 95% confidence intervals that include zero in each year, and are thus uninformative about whether payments towards outstanding tax debt decline following a CNC designation.

2.5.1.2 Payments toward current tax liability

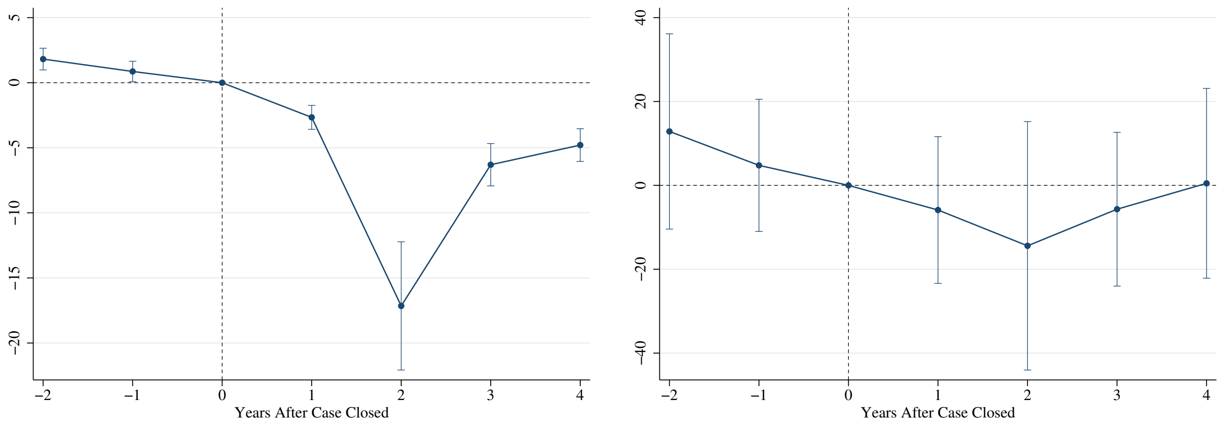
Next, we consider the dollar value of various payments toward the current year's tax liability: withholding from wages, quarterly estimated tax payments, and remittances submitted with a tax return.

Figure 2.5 shows that a CNC designation results in a substantial increase in withholding from wages reported on Form W-2, which peaks at around $30,000 three years after the case closed. This point estimate and the slightly smaller point estimate four years after the case closed are both statistically significant, with 95% confidence intervals that exclude zero. Income tax withheld from wages on Forms W-2 dips considerably in Year 0 relative to

[30]This measure does not include payments against tax debt accrued after the case closed.

Figure 2.4: Effect of a CNC designation on payments toward outstanding tax debt ($1,000)

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 139,246 for difference-in-differences and 121,923 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. Payment values adjusted for inflation to 2017 values.

both the years before and after, which is puzzling given that there is not a corresponding decrease in W-2 wages (see Section 2.5.2.2) or in the likelihood that the taxpayer or their spouse receives a W-2. This dip, which results in a statistically significant positive effect on withholding in the year before the case closed, is a concern for the validity of the identification strategy for the W-2 withholding estimates.

In contrast, there is no effect on either quarterly estimated tax payments or remittances with returns, with relatively tight confidence intervals in the post periods that exclude changes of more than a few thousand dollars in the quarterly estimated tax payments and of more than a few tens of dollars in the remittances with returns.

We also consider whether or not the taxpayer fully paid the current year's tax liability. Figure 2.6 shows the instrumented effect of a CNC designation on the probability taxpayers remit their entire tax liability for a given year. As with W-2 withholding, which is one substantial input into whether the taxpayer paid in full, the coefficient in Year -1 is positive and statistically significant, raising concerns about the validity of the estimation strategy. After treatment, there is no statistically significant effect, though the standard errors are large, and a ten percentage point change is often within the 95% confidence intervals.

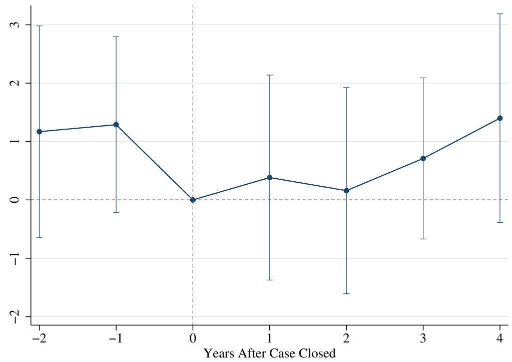
2.5.1.3 Filing tax returns

We also study the effect of a CNC designation on tax compliance as measured by filing a tax return. Taxpayers whose cases are assigned to the field for collection often have incomes

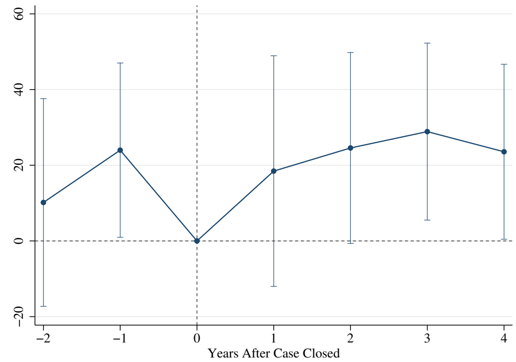
Figure 2.5: Effect of a CNC designation on payments toward current tax liability ($1,000)

W-2 Withholding

(a) Difference-in-differences

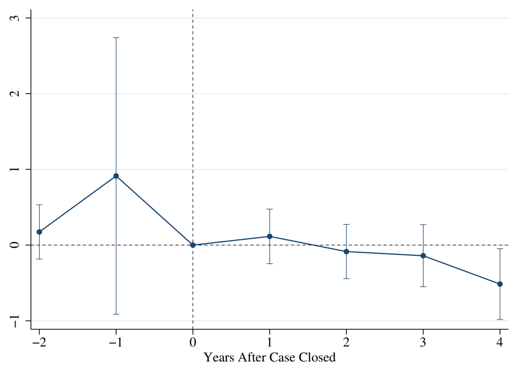


(b) IV with residualized instrument

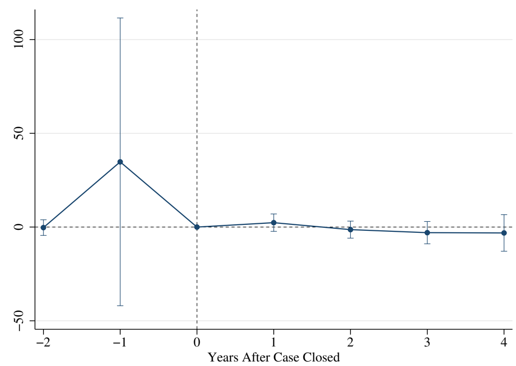


Quarterly Estimated Tax Payments

(c) Difference-in-differences

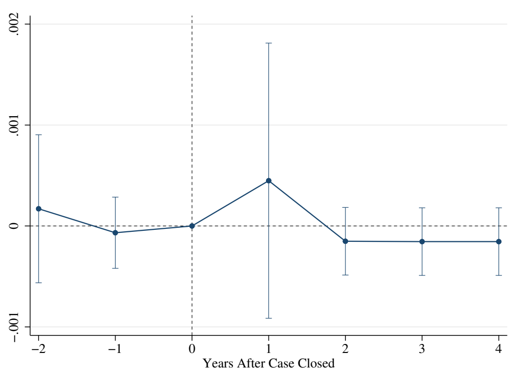


(d) IV with residualized instrument

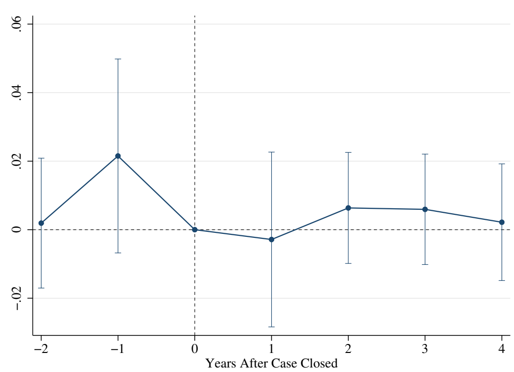


Remittances with Returns

(e) Difference-in-differences



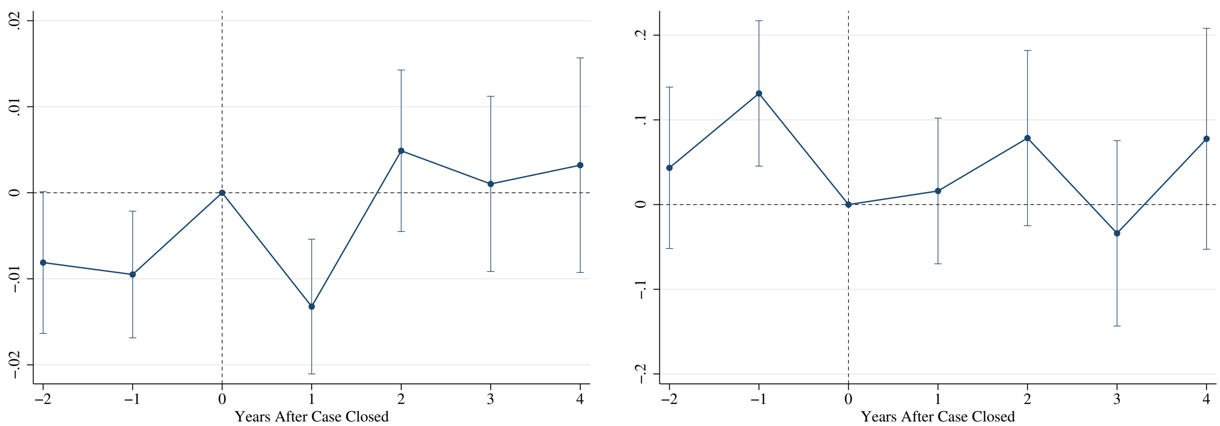
(f) IV with residualized instrument



Notes: Number of cases is 139,246 for W-2 withholding difference-in-differences and 121,923 for the IV specification, 139,246 and 121,923 for quarterly estimated tax payments, and 139,246 and 121,923 for remittances with returns. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. All payment variables adjusted for inflation to 2017 values.

Figure 2.6: Effect of a CNC designation on whether or not the taxpayer remitted their entire current tax liability

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 123,541 for difference-in-differences and 104,139 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

below the threshold at which they would be required to file.[31] As a result, we examine filing behavior among those with wage income reported on Form W-2, who are both more likely to be required to file and more likely to be detected if they fail to file.

The filing results are shown in Figure 2.7. While the standard difference-in-differences results suggest that a CNC designation leads to a reduced likelihood of filing, the IV results show that taxpayers who received a CNC designation were no less likely to file a return in the following years.

2.5.2 Effects on income

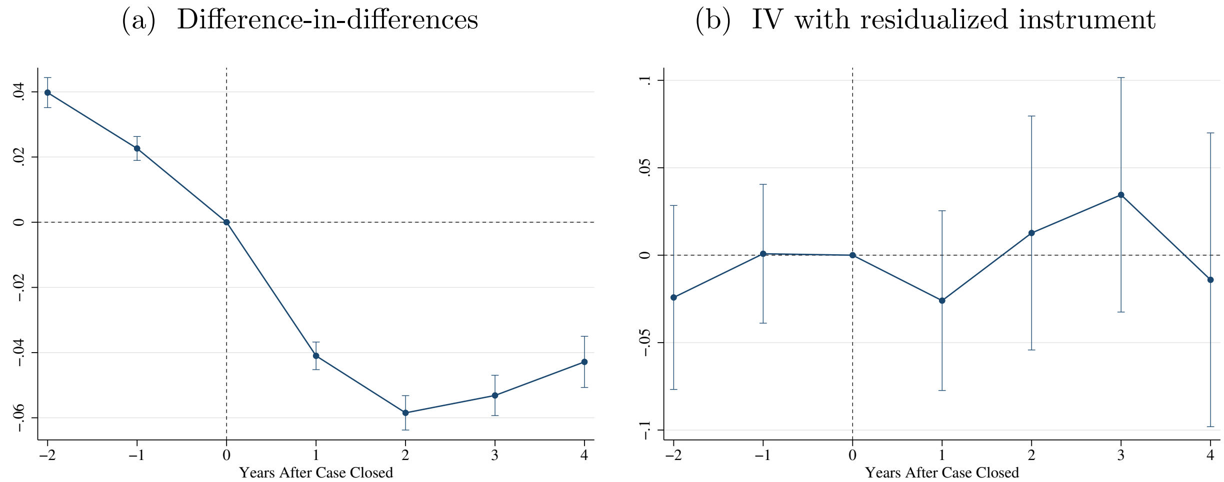
Suspending collection efforts due to hardship does not imply that a taxpayer's debt is forgiven. Taxpayers whose cases are given a CNC designation may use the relief to make investments to increase their future wage or self-employment income, or increase labor supply with the expectation that the income they earn is less likely to go toward tax debts. On the other hand, a CNC designation could reduce labor supply through an income effect, as more of taxpayers' income becomes available for uses other than tax debt repayment.

We examine how a CNC designation affects adjusted gross income, wages reported on Form W-2, and self-employment income reported on Form 1040 Schedule C. We take the inverse hyperbolic sine of each of these measures of income.[32]

[31] For example, we observe that, even among households receiving wages reported on Form W-2, about 40% had wage income below the filing threshold.

[32] The inverse hyperbolic site transformation is: $y_{IHS} = \ln(x_i + (x_i^2 + 1)^{\frac{1}{2}})$. This is approximately equal to

Figure 2.7: Effect of a CNC designation on filing a tax return



Notes: Number of cases is 139,246 for difference-in-differences and 121,923 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

In standard difference-in-differences results CNC status is associated with large, lasting, and statistically significant decreases in adjusted gross income, wages, and self-employment income. Our instrumental variables difference-in-difference approach finds statistically significant increases in adjusted gross income and wages as a result of a CNC designation. This increase is driven by an increase in W-2 wages earned by the spouses of married taxpayers in years three and four after the case closes. We do not observe a statistically significant change in self-employment income.

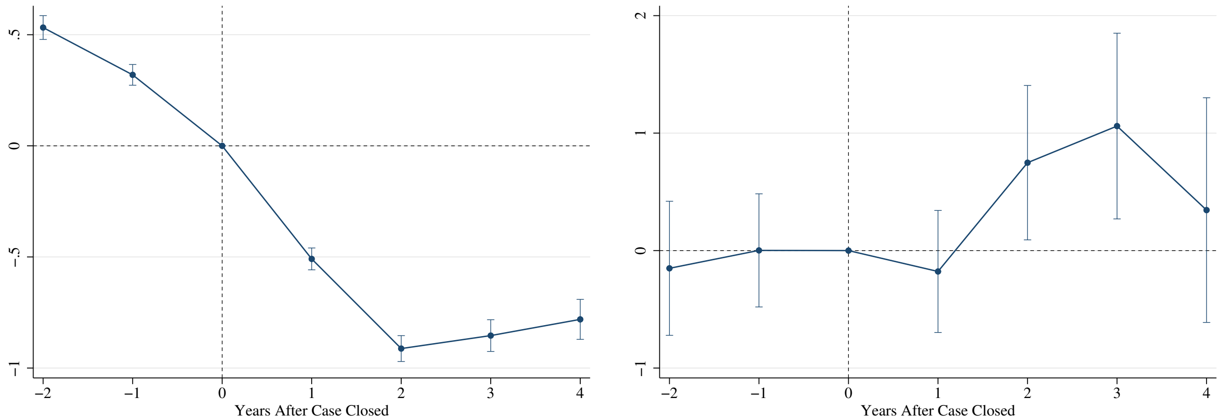
2.5.2.1 Adjusted gross income

We construct a version of adjusted gross income (AGI) that combines adjusted gross income reported on tax returns with third-party reported income of non-filers. Figure 2.8 shows that, in a standard difference-in-differences regression, a CNC designation has a significant negative effect on household income, implying that receiving hardship relief decreases household income by as much as 60% two years after Year 0. In contrast, the instrumented difference-in-differences specification suggests that a CNC designation leads to statistically significant increases in adjusted gross income, which more than doubles three years after Year 0. The stark difference between the two approaches arises because

$\ln(2x_i) = \ln(2) + \ln(x_i) = 0.69 + \ln(x_i) \approx \ln(x_i)$ except for small values of $\ln(x_i)$. As a result, coefficients on these variables can be interpreted as we would for standard logarithmic dependent variables, but allows for the transformation of values equal to 0 and negative values. To transform the coefficients back into meaningful values, we use the approximation % change $\approx (\exp(\beta) - 1) * 100$ (see Bellemare and Wichman (2020) for more information about the inverse hyperbolic sine transformation).

Figure 2.8: Effect of a CNC designation on adjusted gross income (IHS)

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 138,842 for difference-in-differences and 121,208 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Adjusted gross income adjusted for inflation to 2017 values.

CNC designations reflect selection on past income and expectations of low future income, which differencing does not remove, while instrumenting effectively restricts to taxpayers whose circumstances are similar at the time of the CNC decision. In spite of the statistical imprecision of our estimates, among taxpayers on the margin of a CNC designation, those given a CNC designation have much higher income in the following years.

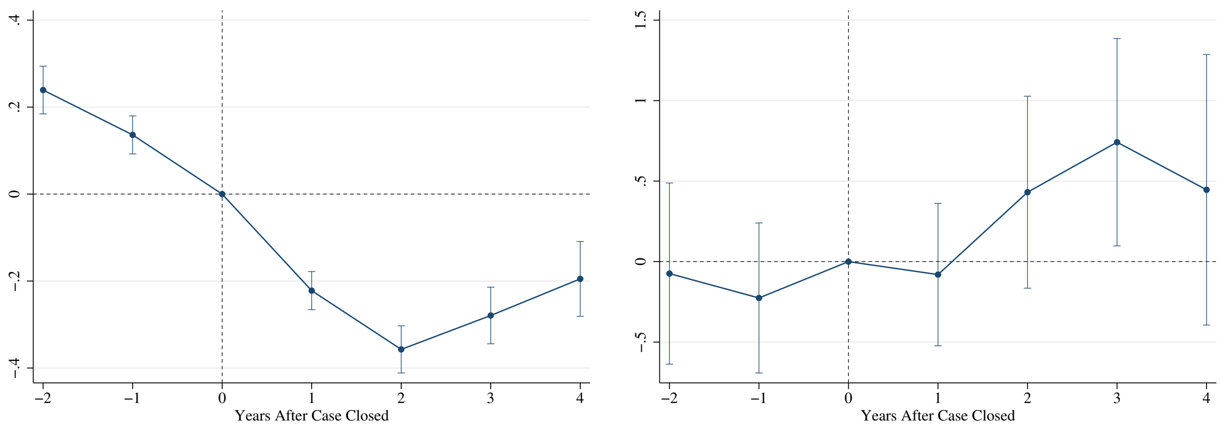
2.5.2.2 W-2 wages

Figure 2.9 shows the effects of CNC designation on the sum of the taxpayer's and their spouse's wages reported on Form W-2. The patterns in these results closely resemble the patterns for adjusted gross income: the standard difference-in-differences specification shows wage decreases after Year 0, while the instrumented difference-in-differences specification shows large increases in W-2 wages. The increase in Year 3 is statistically significant at the 95% level.

To examine the source of the large rise in income, we split household wages into the taxpayers' own wages and taxpayers' spouses' wages. The increase is attributable to taxpayers' spouses. Figure 2.10 shows that in the instrumented specification wages do not rise for the individual whose case is designated CNC (Figure 2.10b), but there is a substantial increase in the wages received by the spouses of married taxpayers (Figure 2.10d), which more than double 3 and 4 years after the case closed. There is a similar pattern in the standard difference-in-differences specification for spouses' wages, although for individuals' own wages

Figure 2.9: Effect of a CNC designation on W-2 wages (IHS)

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 139,246 for difference-in-differences and 121,923 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. W-2 income adjusted for inflation to 2017 values.

the standard difference-in-differences shows a decline that begins well in advance of treatment.

2.5.2.3 Self-employment income

Finally, we consider self-employment income, the sum of all Schedule C profits and losses reported with a taxpayer's tax return.

The impact of a CNC designation on self-employment income is shown in Figure 2.11. A standard difference-in-differences regression shows self-employment income falling after a CNC designation, while the coefficients of the instrumented specification hover around zero and are imprecisely estimated.

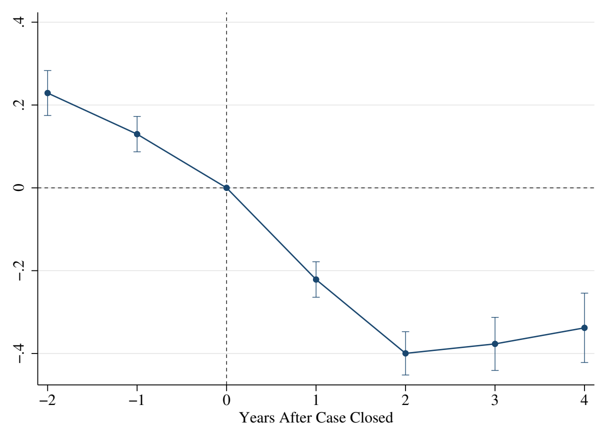
2.5.3 Summary

Suspending debt collection due to inability to pay basic living expenses leads to a large proportional increase in the taxpayer's spouse's (but not the taxpayer's own) W-2 wages three and four years after the case closed, accompanied by a rise in tax withheld from those wages and reported on Form W-2. This increase in income is consistent with the incentives a CNC designation provides by reducing the effective marginal tax rate (at least over the range where additional income does not trigger the revocation of the CNC designation), and suggests that the increased keep rate on higher earnings has a greater effect on income than the wealth effect from setting debt aside. The married taxpayers in our sample are mostly male. If we assume their spouses are more likely to be female, our results are consistent

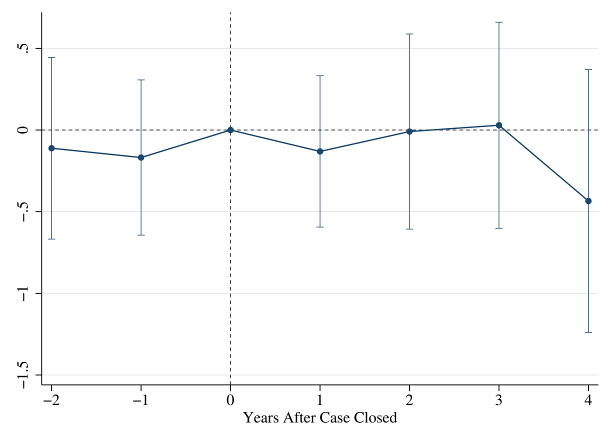
Figure 2.10: Effect of a CNC designation on W-2 Wages (IHS): Taxpayer vs. Spouse

Taxpayer whose case was designated CNC

(a) Difference-in-differences

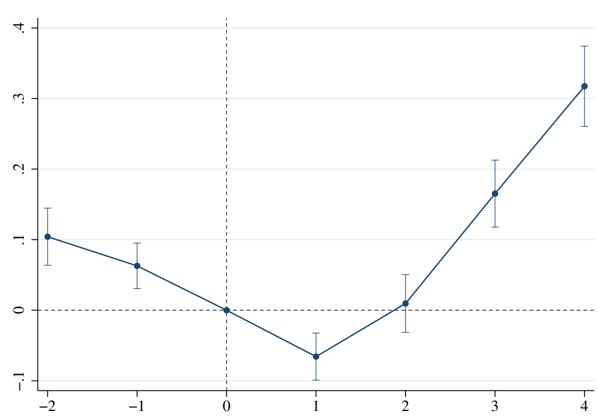


(b) IV with residualized instrument

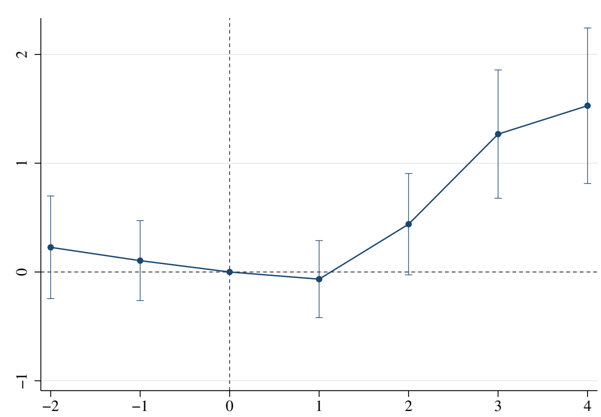


Spouse

(c) Difference-in-differences



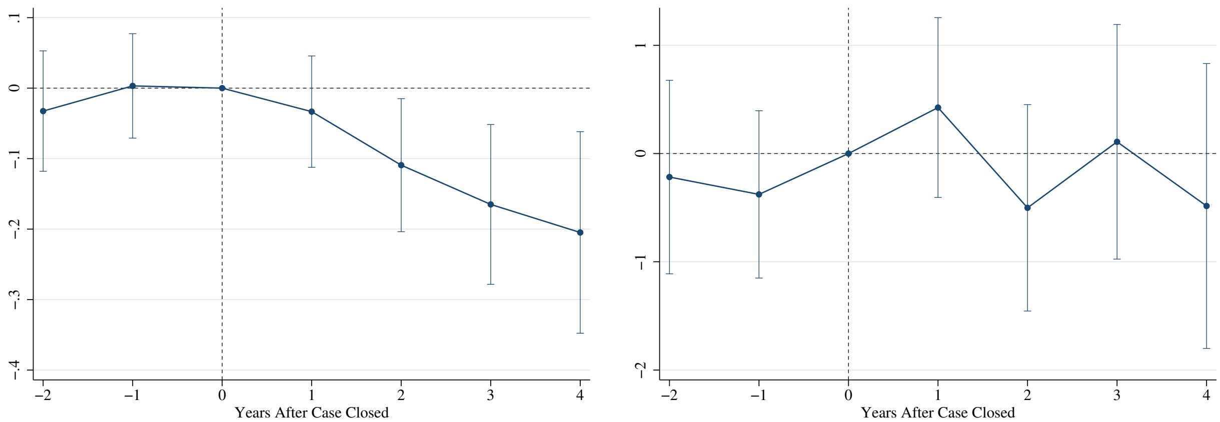
(d) IV with residualized instrument



Notes: Number of cases for W-2 wages is 139,246 for difference-in-differences and 121,923 for the IV specification, and the case numbers for the spouse's W-2 wages are 139,246 and 121,923. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Spouses include individuals married to taxpayers in the year in which their case was closed. W-2 income adjusted for inflation to 2017 values.

Figure 2.11: Effect of a CNC designation on self-employment income (IHS)

(a) Difference-in-differences (b) IV with residualized instrument



Notes: The number of cases is 123,412 for difference-in-differences and 103,795 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Self employment income adjusted for inflation to 2017 values.

with the standard finding that married women's labor supply is more elastic than men's labor supply (Keane, 2011). An alternative explanation for the increase in spouse's wages is through a liquidity channel: by enabling investments that facilitate working (e.g. a second car or child care), the positive wealth shock from a CNC designation could increase labor supply in this population of taxpayers. Given the rise in income, the increase in tax withheld on Form W-2 is largely mechanical rather than the result of an active change in voluntary compliance, as default withholding rates generally align with tax liability and taxpayers rarely set withholding rates near zero.

A CNC designation does not detectably change payments toward outstanding debt, tax filing, quarterly estimated tax payments, remittances with returns, whether the taxpayer remits all current year tax liability, or self-employment income. The lack of significant estimated effects on tax compliance could result from the imprecise nature of the IV estimates, from an offset between the increase in compliance stemming from the potential for debt suspension to be revoked and the decrease in compliance due to the perception that enforcement is lenient, or from an absence of such effects.

2.6 Conclusion

Random and quasi-random assignment of cases to administrative officers provides an opportunity to study the causal effects of policy interventions when administrative officers' discretion can determine who receives treatment. This paper uses such an approach to assess the effects of suspending attempts to collect a taxpayer's unpaid taxes. Variation comes from assignment of cases to Revenue Officers of differing "leniency," the inherent tendency to designate cases currently not collectible due to taxpayer hardship. To our knowledge, this is the first application of this research design to issues of tax enforcement.

The effects of tax debt collection policies on taxpayer behavior are of interest to economic researchers and policymakers. This paper provides the first causal estimates of the impact of suspending efforts to collect tax debt on future taxpayer behavior, and addresses the challenge posed by the selected sample of taxpayers subject to such action. We find that, following the suspension of collection efforts, taxpayers have higher incomes, driven by increases in their spouses' wages. Unlike naive difference-in-difference results, in which suspending collection is strongly associated with declines in debt repayment, tax filing, and both wage and self-employment income, we find that these behaviors either increase or do not decline detectably.

In this setting, the ability of an examiner assignment design to elicit a causal effect of debt forgiveness not polluted by sample selection bias is tempered by the fact that the assignment of case managers to cases is only conditionally random and the average number of cases per Revenue Officer is not very large. But, because quasi-random assignment of tax officers to cases is widespread, and the research availability of administrative data is growing rapidly, the potential for this research design to provide insight into the causal effects of tax enforcement actions is substantial, and we look forward to it being applied to other settings where the empirical caveats are less severe.

CHAPTER III

Tax Preparers and High Income Earners (with Giacomo Brusco, Yeliz Kaçamak, and Mark Payne)

3.1 Introduction

The top of the income distribution has become a point of increased interest for both economic researchers and policy makers. As income inequality has increased over the past several decades, so too has the interest in understanding the economic behavior of this population. A recent point of focus in this conversation is the extent to which this population avoids and evades its tax liability. Concurrently, there has been a resurgence of attempts to understand the role of paid tax professionals in tax compliance.

Although both topics have been analyzed separately in the economics literature, the role paid tax professionals play in noncompliance of the rich has not been explicitly studied. Given that over 90% of individuals in the top 1% of the income distribution use a paid tax preparer, understanding the role professional tax preparation services play in the tax compliance (or noncompliance) of the very rich is a critical piece of the puzzle.

This paper bridges the gap between the literature on top income earners (and, in particular, the tax compliance of top income earners) and the literature on tax preparers. We use administrative tax data from random audits to provide novel descriptive statistics documenting how paid-prepared returns compare to self-prepared returns across the income distribution. Our data include both the originally reported amounts and the post-audit corrected amounts at the line-item level. We find that, among the bottom 99% of the income distribution, self-prepared returns have smaller post-audit corrections relative to total income, but among the top 1%, paid-prepared returns experience smaller corrections relative to self-prepared returns.

We then investigate why not all individuals in the top 1% use paid tax preparation services given that they tend to have smaller post-audit corrections. We find that individuals in the top 1% who are themselves an accountant or other type of financial advisor, but who self-prepare their own returns, experience the same smaller corrections as do paid-prepared returns among the top 1%. We also find that individuals in the top 1% who have a higher proportion of their total income from third-party reported income (e.g., wages) are less likely to use tax preparation services. This suggests that there are both supply and demand considerations at play.

We provide a brief overview of the literatures on tax preparers and on income and wealth inequality in the U.S. in Section 3.2. These literatures have highlighted how tax systems can contribute to or hinder the secular increase of economic inequality, not just through tax rates but also through decisions on tax bases and enforcement. This underlines the importance of examining paid tax preparer usage by top earners. Unlike the use of tax preparation software, which has seen large increases for low- and middle-income taxpayers over the past two decades, tax preparer usage has remained more or less constant, with much higher usage among high income individuals than the rest of the population.

We use administrative data from the National Research Program (NRP) of the U.S. Internal Revenue Service (IRS) to explore what distinguishes the use of tax preparers among the very rich from their use in the rest of the population. These data provide the original and corrected values, by line-item, for a stratified random sample of the population that were selected for a full audit of their tax return.

The results of this descriptive analysis are given in Section 3.3. We find that paid preparer use increases with income, particularly so among the top 1%. We document trends about which types of income are associated with greater paid preparer usage across the income distribution, and conclude that self-employment income plays a bigger role for the bottom 99%, but income from capital gains, dividends, and partnerships and S-corporations play a bigger role for the top 1%. We find that post-NRP audit corrections are smaller among self-prepared returns in the bottom 99%, but smaller among paid-prepared returns in the top 1%.

We also document trends among paid-preparer users across the income distribution. In particular, the top 1% are considerably more likely to hire Certified Public Accountants (CPAs) and attorneys as tax preparers relative to the bottom 99%, while the bottom 99% are more likely to hire Enrolled Agents or Supervised Registered Tax Preparers. The top 1% also tend to pay dramatically more for tax preparer services in levels, but dramatically less for tax preparation services as a proportion of income.

Given that we observe smaller post-NRP audit corrections among top 1% returns prepared a paid preparer, the natural follow-up question is why does anyone in the top 1% *not* use a paid preparer? How do the 10% of the top 1% who self-prepare their returns differ from the rest of the top 1%? To investigate this question, we develop a simple, stylized model in the style of Allingham and Sandmo (1972b) in Section 3.4. The model provides some intuition and justification for our empirical approach, which yields evidence for two potential answers to this question: the effect of a specialized occupation, and the effect of having income subject to third-party reported information. To test the former, we examine how the relative size of post-NRP audit corrections experienced by the top 1% who use preparers compare to self-prepared returns among the top 1% who have occupations in accounting or financial planning. We find that these populations experience the same relative size of corrections.

To test the second answer to our question, we consider whether individuals in different income groups are more or less likely to have hired a tax preparer based on the proportion of their income stemming from seven different income types: wages, self-employment income, capital gains, ordinary dividends, income from partnerships and S-corporations, rental income, and other income not reported elsewhere on the tax return. We find that, for the top 1%, the proportion of income coming from wages–a third-party reported income source–is the only statistically significant predictor of this type for whether or not someone in the top 1% hired a professional tax preparer (and that having a great proportion of total income coming from wages is associated with a lower probability of hiring a tax preparer).

We provide some concluding discussion, caveats, and thoughts for future research on this topic in Section 3.5.

3.2 Background and data

3.2.1 The role of paid tax preparers in tax compliance

Individual taxpayers in the United States generally either prepare their tax returns themselves ("self-prepared" returns) or hire professionals to help them ("paid-prepared" returns). More than half of individual taxpayers (and nearly all corporations) in the United States receive paid professional assistance in filing their taxes.

Despite their ubiquity, the role of tax preparers in tax compliance is not well understood. There are two intuitive reasons why an individual might seek professional help in preparing their taxes. If taxpayers are worried about making mistakes on their returns, they may hire professionals to improve the accuracy of their returns. On the other hand, taxpayers may pay professional tax preparers to help them learn about additional avenues for tax avoidance and evasion.

The empirical evidence documents a negative correlation between tax preparer usage and reported income conditional on observable characteristics (Klepper and Nagin, 1989; Long and Caudill, 1987; Erard, 1993, 1997). However, inferring a causal relationship between tax compliance and tax preparers is a difficult task because of the potential selection bias. It is reasonable to believe that taxpayers who use tax preparers are substantially different than taxpayers who do not. In fact, previous works in the literature suggest that the use of professional tax services is positively correlated with income, self-employment, the complexity of tax return, and age (Slemrod and Sorum, 1984; Long and Caudill, 1987; Slemrod, 1989; Dubin et al., 1992), which are characteristics that also are correlated with tax compliance.[1]

Despite the selection bias challenge, many researchers have attempted to investigate the possible causal relationship between the tax preparer usage and tax compliance. One strategy in the accounting literature has been to attempt to compare compliance on more or less legally ambiguous line items (Klepper and Nagin, 1989; Hite and McGill, 1992; Fleischman and Stephenson, 2012). Another approach was to analyze the effects of regulation on tax preparer aggressiveness (Ayres et al., 1989; Hansen and White, 2012). These experimental studies yielded mixed results and, due to the hypothetical decision context and low sample size, may have limited external validity.

There has been a recent resurgence in attempts to estimate the causal impact of tax preparers on tax compliance. Battaglini et al. (2019) use a seven-year panel of Italian administrative tax records to study potential "social spillovers" through the use of tax preparers by sole-proprietors. They find that a taxpayer's level of under-reported income is strongly correlated to the other clients of his or her accountant, and that this relationship holds even if the taxpayer changes tax preparers (that is, a taxpayer's under-reported income at time $t-1$ is correlated with under-reported income of their tax preparer's other clients at time t). They explore two mechanisms through which these social spillovers may occur: self-sorting by tolerance for noncompliance, and information externalities.

Yuskavage et al. (2019) use the density of tax preparation services in a zip code as an instrument for tax preparer use to address endogenous selection into tax preparer use. Controlling for zip-code, income, and complexity of the tax return, they find that returns filed with the help of a paid tax preparer have larger adjustments upon audit. This effect is not present with the Volunteer Income Tax Assistance (VITA) for similar tax returns, which implies that variance in tax preparer motives may affect tax compliance outcomes. The density of tax preparation services is potentially problematic as an instrument, however: the density of tax preparers could be the result of demand for tax preparation services (and therefore also a function of unobservables associated with hiring a tax preparer), rather than demand being higher as a result of greater supply.

[1] Paid tax preparation service use is also correlated with characteristics of low-income individuals, such as claiming the EITC. For example, Masken et al. (2008) find that tax preparers can facilitate access to bank products such as Refund Anticipation Loans (RALs) and Refund Anticipation Checks/Cards (RACs), which give taxpayers quicker access to their refunds.

3.2.2 The distribution of income and wealth in the United States and the role of the tax system

There is a long-standing literature that documents an increase in the top 1% share of (pre-tax and pre-transfer) income, especially in the English-speaking world.[2] There are several potential explanations for this phenomenon. Perhaps most prominent among these is the idea that small differences in skill at the top could be getting magnified into much bigger differences in salary, and that this phenomenon has become more widespread in the past few decades thanks to technological developments (Rosen, 1981; Berman et al., 1998).

The public economics literature has recently been concerned with how tax systems have reacted and contributed to these changes in income distribution. Alvaredo et al. (2013) point out that, in the English-speaking world, top income tax rates have been moving in the opposite direction as the share of income accrued to the top 1% of the distribution. They argue that this tendency might have operated in conjunction with other factors to contribute to the increase in inequality. Others, like Scheuer and Slemrod (2020), have stressed the importance of thinking about tax rates in conjunction with other aspects of the tax system, such as the tax base and resources for enforcement, both of which influence the extent to which changes in tax rate end up impacting government revenue.

Concerns over the erosion of the tax base, particularly among the top of the income distribution, have been highly prominent in the recent public economics literature. Alstadsæter et al. (2018), for instance, suggest that about 10% of global GDP is held in tax heavens. Alstadsæter et al. (2019) argue that increasing tax enforcement efforts in Norway has been successful in raising substantially more revenue from top income earners as a result of newly disclosed offshore assets. Guyton et al. (2021) estimate that accounting for previously undetected offshore accounts and income from pass-through businesses among the top 1% considerably increases their share of fiscal income. Given the almost ubiquitous usage of tax preparers among top U.S. income earners, they could potentially play a crucial role in facilitating sophisticated methods of avoidance and evasion.

3.2.3 Data

We use data from the IRS's National Research Program (NRP). The NRP is the IRS's random audit program. Starting in 2006, the IRS has chosen a stratified random sample of the population of U.S. taxpayers each year. These taxpayers' returns are then audited. NRP audits cover all items on a filer's tax return, as opposed to an operational audit, which may only cover the specific items of interest. Populations of particular interest for tax compliance

[2] See Atkinson et al. (2011) for a historical perspective.

measures are over-sampled (in particular, high-income taxpayers, or taxpayers claiming the Earned Income Tax Credit (EITC)). The NRP data form the base for measures of both income under-reporting and the tax gap.

We use the 2006-2014 waves of the NRP. Each wave contains approximately 12-15 thousand individuals. We include all individuals except those who are given 0 probability weight,[3] and those whose returns were prepared through the Volunteer Income Tax Assistance (VITA) program.[4] Our final sample includes 117,303 taxpayers. A little over 70% of our sample filed a return using a paid professional.

The NRP data include line-item-level information from seven tax schedules and forms, briefly described in Table 3.1. For each included line item, we have the original amount reported on the return and the "corrected," post-audit amount. If no change was made, the original and corrected amounts are equivalent. We focus on in particular on income totals across the following categories: wage income (from W-2s), self-employment income (Schedule C income), income from capital gains and dividends (Schedule D income), and income from partnerships/S-Corporations and rental income (Schedule E income). We inflate dollar values to 2014 values.[5]

Table 3.1: Forms and schedules included in the NRP data

Form 1040	The "U.S. Individual Income Tax Return."
Schedule A	Itemized allowable deductions, including charitable contributions. In lieu of filling out Schedule A, taxpayers may choose to take the standard deduction.
Schedule C	Income and expenses the result from self-employment.
Schedule D	Documents capital gains and losses incurred throughout the tax year, as well as income from ordinary dividends.
Schedule E	Income and expenses that result from the rental of real property, royalties, or from pass-through entities (e.g., trusts, estates, partnerships, or S Corporations)
Schedule F	Income and expenses related to farming.
Form 2106	Work-related expenses for Armed Forces reservists, qualified performing artists, fee-basis state or local government officials, or employees with impairment-related work expenses.

[3]All individuals in the NRP are assigned a probability weight for re-scaling measures of income under-reporting and of the tax gap to the full population. Having a 0 probability weight indicates that the return was excluded from the study for one of a variety of reasons, including but not limited to disaster, combat, or being unable to locate the individual.

[4]The VITA program, managed by the IRS, offers free, basic tax preparation assistance to qualified individuals (in particular, those making less than $57,000 a year and/or individuals with disabilities and/or individuals with limited English). Excluding VITA-prepared returns eliminates 1,203 individuals from our sample.

[5]We use the CPI Inflation Calculator provided by the U.S. Bureau of Labor Statistics (available at https://www.bls.gov/data/inflation_calculator.htm). We compare December of each year to December 2014 to generate an "inflation adjustment" for each year.

We assign each individual in our sample to a "corrected adjusted gross income (AGI) percentile." To do this, we use the population of taxpayers in each sample year to determine year-specific 5% percentile cut-offs (e.g., 0-5%, 5-10%, etc.). We then assign each individual in our sample to a percentile based on their *corrected* (i.e., post-audit) AGI in the relevant year. While the choice to use a paid preparer is endogenous (and may have an impact on reported AGI), using corrected AGI will mitigate this problem to an extent. Corrected AGI is not able to fully eliminate the problem if, for example, paid-preparer use is associated with different rates of under-reported income across the income distribution. Another caveat is that even the corrected AGI variable reflects true income with noise rather than the true income itself.

We use a variety of percentile groupings, but focus primarily on income deciles through 90%, then split the top 9% into 90-95% and 95-99%. The top 1% is divided into 99-99.5%, 99.5-99.9%, and 99.9% and above. In some figures and analyses, we group 0-99% together for comparison against the top 1%. See Table 3.2 for subtotals by corrected AGI percentile.

In our analysis, we separate out individuals who had negative corrected AGI. As we discuss in Section 3.3, individuals who have negative corrected AGI "look" more like the top 1% of the income distribution than the bottom 99% along some important dimensions. They also tend to have large adjustments as part of the audit process. If these individuals are included at the bottom of the income distribution, both the size of adjustments and the extent to which paid preparers are used by the bottom 99% are overstated.

We merge on additional information from other tables in the administrative tax database to the NRP data. Before the Tax Cuts and Jobs Act of 2017, tax preparation fees were tax deductible and listed on Schedule A. We estimate tax preparation fees using information from Schedule A.[6]

We include the taxpayers' year of birth, state of residence for filing purposes, and filing status (from which we derive marital status). We use the 3-digit occupation code from the taxpayer's highest grossing W-2 to create an indicator of whether or not the taxpayer works as an accountant or financial manager.[7] Finally, we include indicators of whether the tax preparers self-identified as Certified Public Accountants (CPAs), attorneys, Enrolled Agents, or Supervised Registered Tax Preparers (SRTPs).

It is important to note that the NRP data are not able to give a perfect measure of errors on tax returns submitted to the IRS. First, there is no indication given as to whether the

[6]In particular, we subtract Line 21 ("unreimbursed employee expenses") and Line 23 ("other expenses") from Line 24 ("gross deduction amount", the sum of lines 21-23). This calculation yields Line 22 ("tax preparation fees") as our measure of fees paid to professional preparers.

[7]We assign an accountant indicator equal to 1 for occupation codes 122 (financial manager), 141 (accountant, auditor, other financial specialist), and 151 (certified public accountant).

Table 3.2: Subgroup counts

	Paid-prepared	Self-prepared	Total
Corrected AGI Percentile: 0-99 Percent			
0-5 Pctl	688	691	1,379
5-10 Pctl	1,364	1,173	2,537
0-10 Pctl	2,052	1,864	3,916
10-15 Pctl	1,658	1,315	2,973
15-20 Pctl	1,972	1,291	3,263
10-20 Pctl	3,630	2,606	6,236
25-25 Pctl	2,044	1,324	3,368
25-30 Pctl	2,223	1,134	3,357
20-30 Pctl	4,267	2,458	6,725
30-35 Pctl	2,467	1,223	3,690
35-40 Pctl	2,507	1,275	3,782
30-40 Pctl	4,974	2,498	7,472
40-45 Pctl	2,565	1,362	3,927
45-50 Pctl	2,671	1,375	4,046
40-50 Pctl	5,236	2,737	7,973
50-55 Pctl	2,936	1,509	4,445
55-60 Pctl	3,078	1,516	4,594
50-60 Pctl	6,014	3,025	9,039
60-65 Pctl	3,246	1,462	4,708
65-70 Pctl	3,550	1,409	4,959
60-70 Pctl	6,796	2,871	9,667
70-75 Pctl	3,874	1,507	5,381
75-80 Pctl	4,229	1,552	5,781
70-80 Pctl	8,103	3,059	11,162
80-85 Pctl	4,542	1,715	6,257
85-90 Pctl	4,994	1,947	6,941
80-90 Pctl	9,536	3,662	13,198
90-95 Pctl	5,948	2,130	8,078
95-99 Pctl	13,669	4,124	17,793
0-99 Pctl	70,225	31,034	101,259
Corrected AGI Percentile: Top 1 Percent			
99-99.5 Pctl	2,986	424	3,410
$99.5 - 99.9$ Pctl	6,819	704	7,523
99.9+ Pctl	3,161	193	3,354
r1	14,544	1,500	16,044
Corrected AGI Percentile: Negative AGI			
AGI $<$ 0	1,578	179	1,757
Total	84,769	32,534	117,303

Notes: Includes individuals from the 2006-2014 NRP waves with positive probability weights.

noted mistakes were honest mistakes or attempted tax avoidance or evasion. Second, not all corrections are themselves necessarily accurate. The auditor may fail to detect under- (or over-) reported income, or may "correct" a line item that was originally accurate. Finally, NRP audits may fail to detect certain types of income, such as income from undeclared foreign bank accounts and pass-through businesses (Guyton et al., 2021). We discuss how this limitation may affect our analysis of post-audit corrections and the role of preparers in the relevant sections below.

3.3 The use of paid preparers across the income distribution

Figure 3.1a shows the proportion of individuals with paid-prepared versus self-prepared returns by corrected AGI percentile. The proportion of each decile using a paid preparer steadily increases in the bottom 99% from a little over 50% in the $0-10$ percentile group to over 75% in the $95-99$ percentile group. There is a noticeable jump in the proportion of individuals using a paid preparer in the top 1%: in the $99-99.5$, $99.5-99.9$, and 99.9+ percentile groups, 88%, 91%, and 94% used a paid preparer, respectively. This jump is mirrored by individuals who had negative corrected AGI, with 90% using a paid preparer.[8]

The observation that the proportion of individuals using a paid tax preparer increases with corrected AGI percentile is fairly stable over our sample period, as shown in Figure 3.1b. In particular, the considerable difference in the proportion of individuals using a paid preparer in the bottom 99% of the income distribution compared to the top 1% does not vary over this time period.

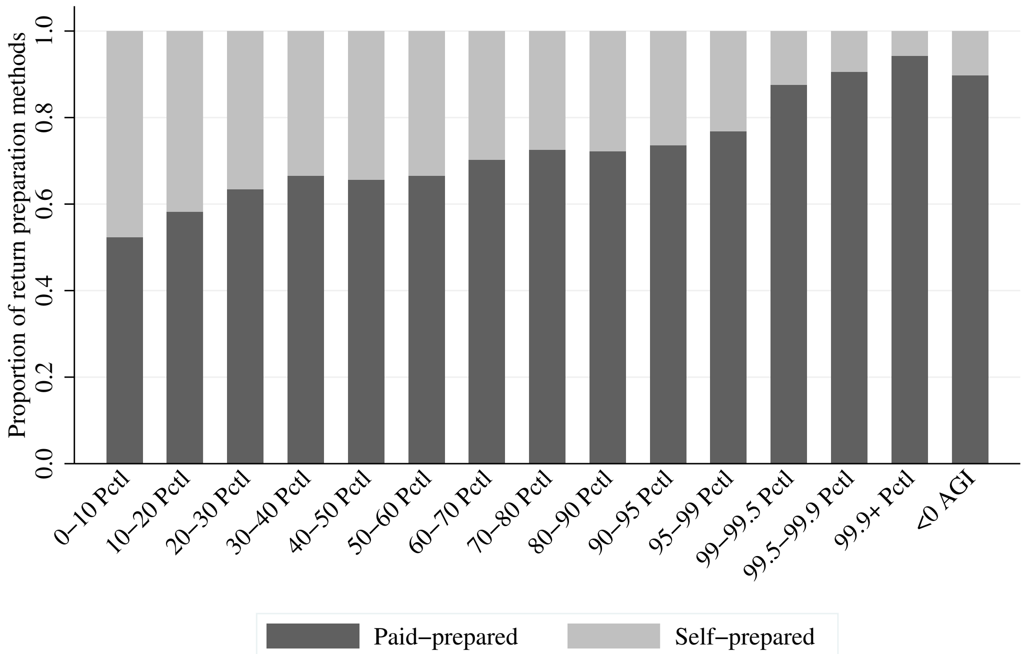
It is worth noting that there has been a shift in how self-prepared returns are prepared over the sample period. Figure 3.2 shows trends in the use of tax-preparation software among self-prepared returns by corrected AGI percentile. Figure 3.2a shows that, while tax-preparation software use generally increases with corrected AGI percentile, the differences are small compared with the range of paid-prepared versus self-prepared across corrected AGI percentiles (in this case, the smallest percentage is 76% among the negative AGI group, while the highest percentage is 91% among the $95-99$ percentile group). The share of self-prepared returns that are prepared using tax-preparation software has been growing over time, however. Figure 3.2b shows that, particularly among the $0-99$, $99-99.5$, and $99.5-99.9$ percentile groups, there has been a noticeable upward trend in tax-preparation software use.[9]

[8]The observation that individuals who had negative corrected AGI "look" more like the top 1% will hold for some, but not all of the characteristics we consider in this section. In some cases, including individuals with negative corrected AGI is so distorting to the figures that we do not show them. For some more thoughts on the group of individuals with negative corrected AGI, see Appendix C.2.1.

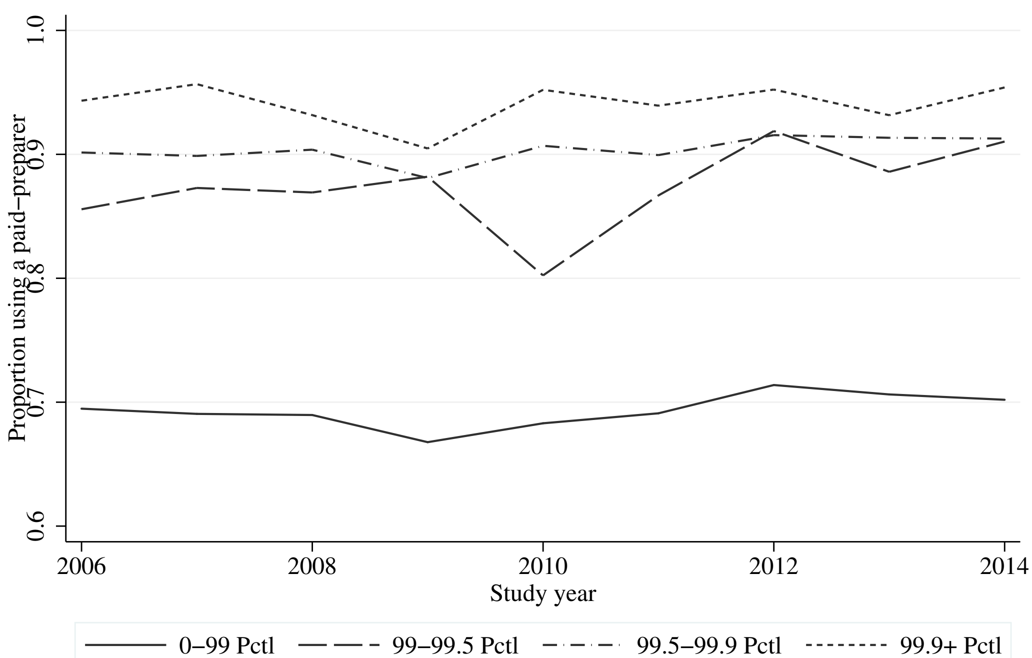
[9]This is consistent with population-level trends in the use of paid preparers, self-prepared by hand, and self-prepared using tax software. See Appendix C.2.2 for additional details.

Figure 3.1: Paid-preparer use by corrected AGI percentile

(a) Proportion



(b) Proportion over time



Notes: N = 117,303 individuals. See Table 3.2 for subtotals by corrected AGI percentile.

While we focus on the distinction between paid-prepared and self-prepared returns, this shift toward increased use of tax-preparation software suggests that understanding the role of tax-preparation software is also an increasingly important piece of the puzzle in understanding trends in tax compliance and tax avoidance.

3.3.1 How do paid-prepared and self-prepared returns vary across the income distribution?

We next consider how the differences between paid-prepared and self-prepared returns vary across the income distribution. We consider the proportion of individuals itemizing their deductions, as well as how the proportion of total income from different income sources vary. We then turn to one of the key advantages of using the NRP data and consider the differences in the size of corrections.

3.3.1.1 Proportion with itemized deductions

Figure 3.3 gives the proportion of paid-prepared returns and self-prepared returns with itemized deductions by corrected AGI percentile. There is a clear difference between the top 1% and the bottom 99% of the income distribution: the vast majority of taxpayers in the top 1% itemizes, regardless of whether they hire a tax preparer or not, but individuals in the bottom 99% who use a tax preparer are more likely to itemize than those who do not.

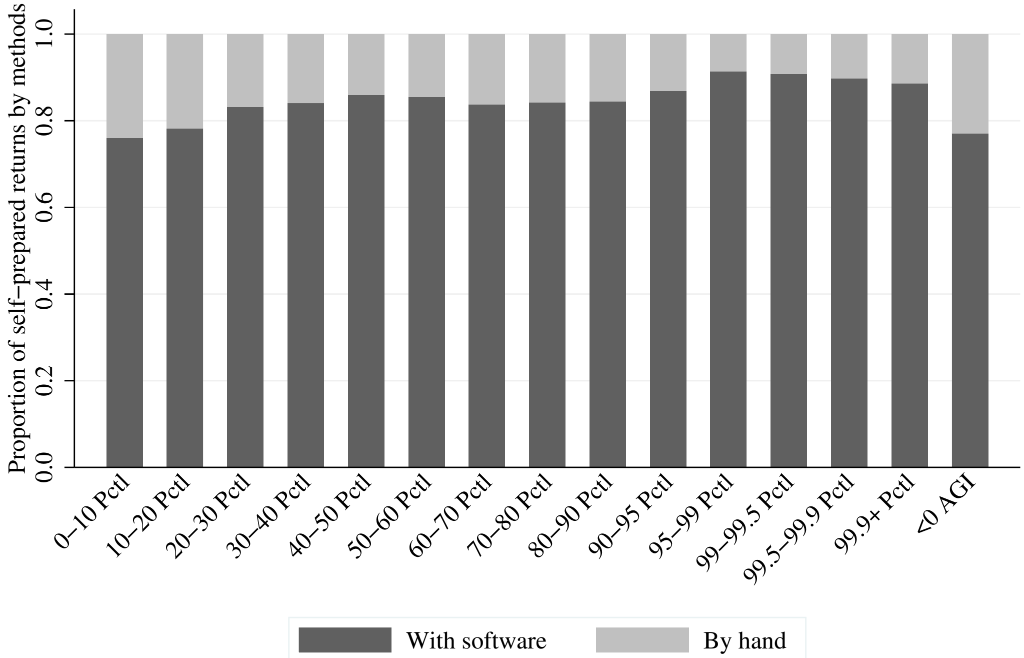
Across paid-prepared and self-prepared returns, the proportion itemizing increases with corrected AGI percentile. The difference in the proportion of individuals itemizing does not vary significantly between paid-prepared and self-prepared returns, but there are some discernible trends. The proportion itemizing is higher among paid-prepared returns for the corrected AGI deciles between $0-10$ and $60-70$. At that point, the trend shifts such that the proportion of returns with itemized deductions is higher among self-prepared returns. The difference between the proportion of returns with itemized deductions that are paid-prepared or self-prepared is smallest for individuals in the top 1%, and largest for individuals with negative corrected AGI.

3.3.1.2 Sources of income

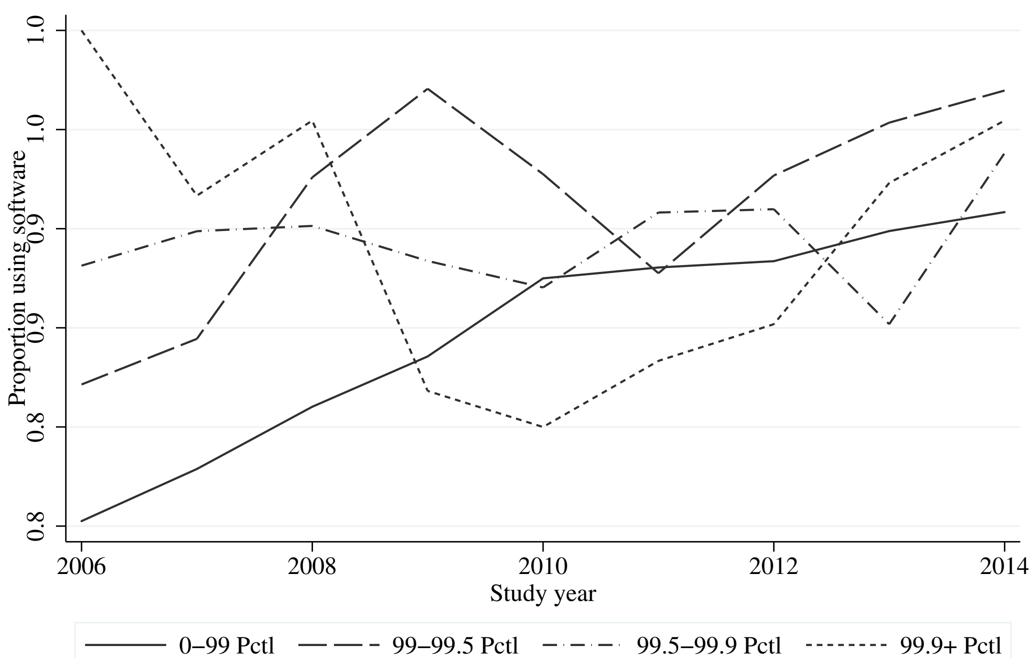
Another dimension that varies across both paid-prepared and self-prepared returns, and across the income distribution, is how many different types of income are reported on the returns. Figure 3.4 shows average count of non-zero income from seven different sources of income: (a) wages, (b) self-employment income, (c) income from capital gains, (d) income from ordinary dividends, (e) income from partnerships and S-corporations, (f) rental income,

Figure 3.2: Self-prepared returns using tax-preparation software

(a) Proportion

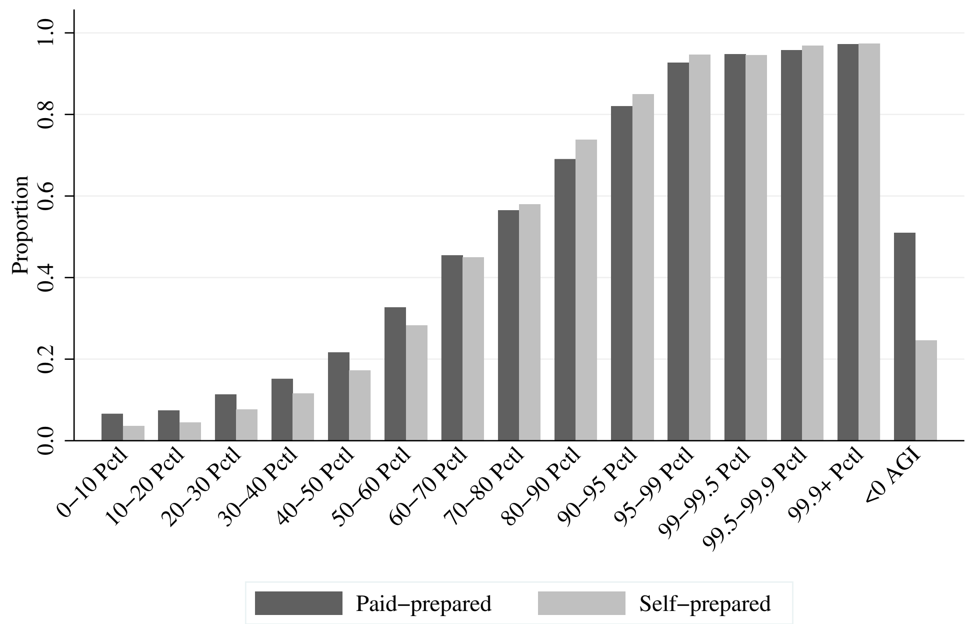


(b) Proportion over time



Notes: N = 32,534 individuals. Limited to returns that were self-prepared. See Table 3.2 for subtotals by corrected AGI percentile.

Figure 3.3: Proportion itemizing deductions



Notes: N = 117,303 individuals. See Table 3.2 for subtotals by corrected AGI percentile.

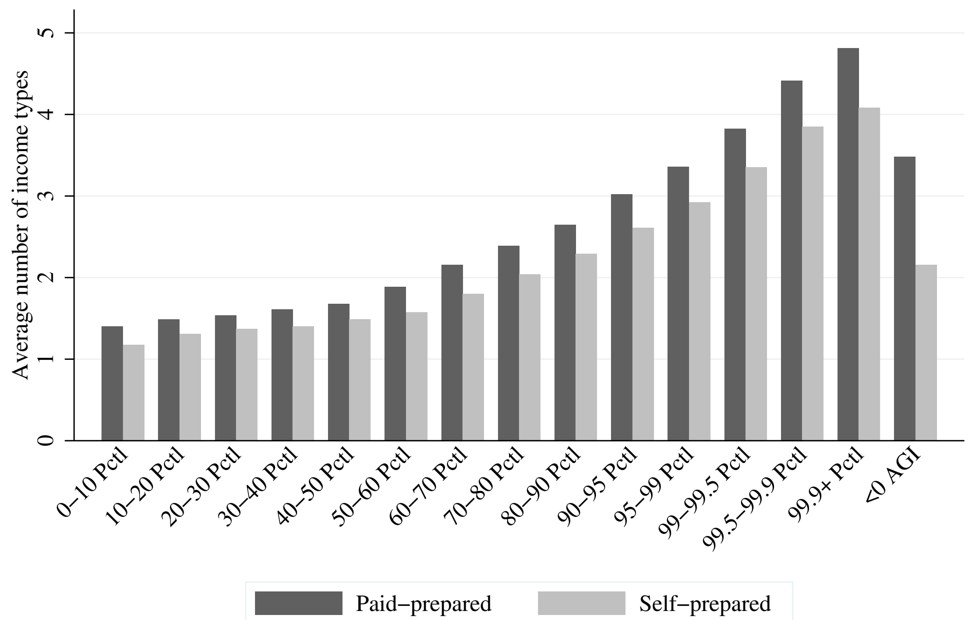
and (g) "other" income.[10] There are two clear trends: on average, returns that are prepared by professionals include more types of income compared to self-prepared returns, and returns from higher in the income distribution include more types of income relative to returns from lower in the income distribution.

In addition to the number of different types of income, the proportion of total income that comes from these different sources varies both by preparation type, and across the income distribution. Figure 3.5 shows the proportion of total income that comes from the seven different types of income we consider. If a taxpayer does not report any of a specific type of income, we assume the proportion of their total income from that type of income is 0. Because they frequently distorted the graphs beyond readability, we do not show the $0 - 10$ percentile group or the negative AGI group.

There are some fascinating trends when we consider the proportion of total income made up by various income sources, and how that varies both across the income distribution and between self-prepared and paid-prepared returns. In Figure 3.5a, it is clear that wage income makes up a considerably smaller proportion of total income for individuals using paid-preparers than individuals who are self-preparing their returns. What is even more striking is that the difference between paid-prepared and self-prepared grows as the corrected

[10] "Other income" is reported on Form 1040. It is the amount of any income not reported elsewhere on Form 1040 or any attached schedules.

Figure 3.4: Average count of income sources



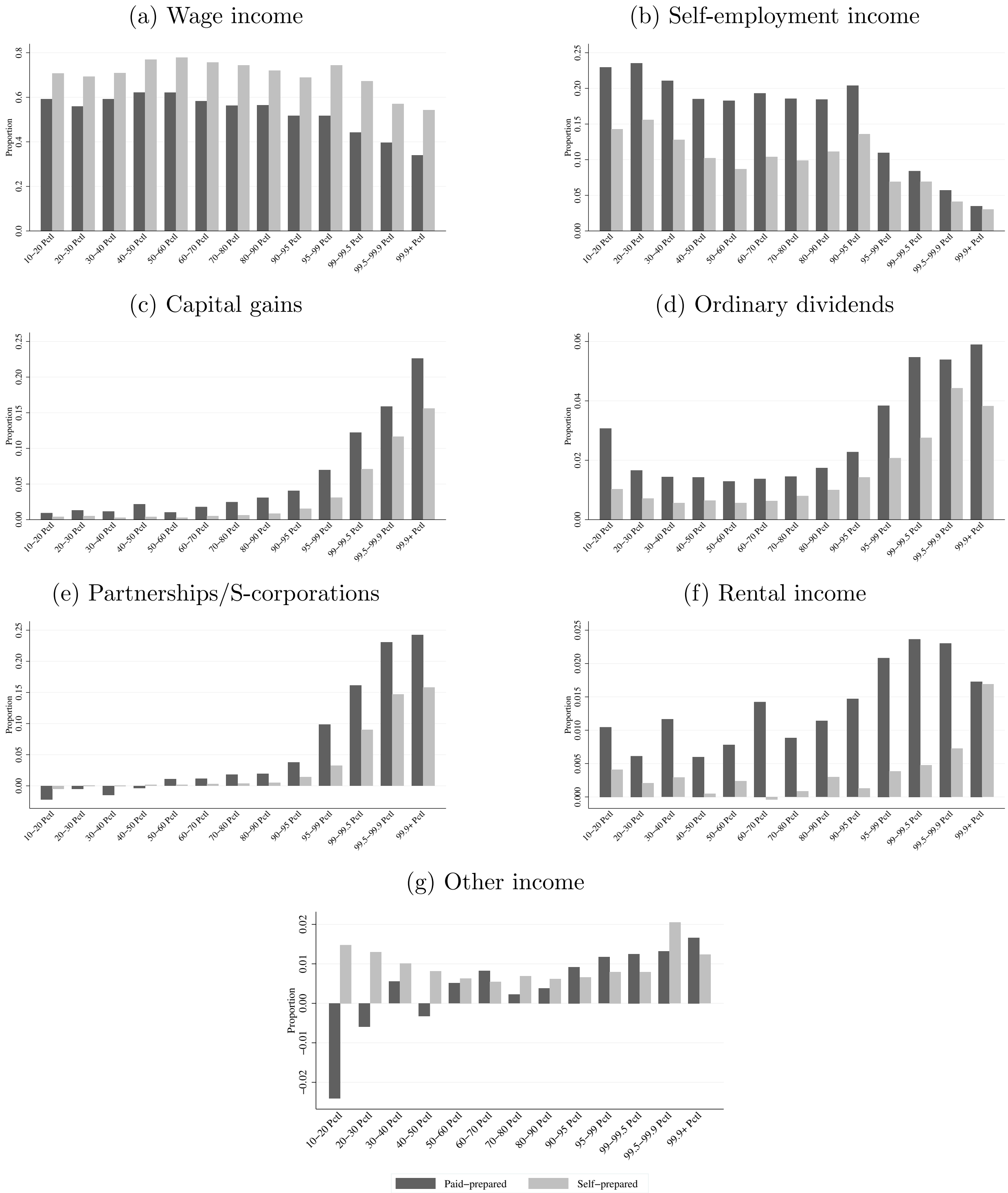
Notes: N = 117,303 individuals. See Table 3.2 for subtotals by corrected AGI percentile.

AGI percentile increases. While the proportion of total income from wages decreases as the corrected AGI percentile increases for self-prepared returns, the proportion of total income from wages increases even more for paid-prepared returns.

In contrast, the proportion of total income from the other sources of income we consider is almost always higher among paid-prepared returns compared to self-prepared returns. Figure 3.5b shows the relative proportions for self-employment income. While self-employment income makes up a larger proportion of total income for paid-prepared returns across the income distribution, the differences between paid-prepared and self-prepared are much smaller among the top 1% than the bottom 99%. The proportion of income from self-employment income generally decreases with corrected AGI percentile; the decrease is much starker among paid-prepared returns than self-prepared returns. The scale for this type of income is also considerably different than for wages: whereas the proportion of total income coming from wages peaked at close to 80%, the proportion of total income coming from self-employment income does not reach 25% for any income percentile.

The proportion of total income from capital gains sits on the same scale as self-employment income, but one of the two trends is reversed from self-employment income. As shown in Figure 3.5c, paid-prepared returns are consistently more likely to report a higher proportion of total income from capital gains than self-prepared returns, but, unlike self-employment income, the proportion of total income from capital gains increases with corrected AGI

Figure 3.5: Proportion of total income by income type



Notes: $N = 111{,}630$. See Table 3.2 for subtotals by corrected AGI percentile.

percentile (as does the difference in proportion between paid-prepared and self-prepared returns). These trends are mirrored (and on the same scale) for the proportion of income from partnerships and S-corporations (see Figure 3.5e). The proportion of income from partnerships and S-corporations may be understated: Guyton et al. (2021) suggest that random audits often do not capture under-reported income from pass-through businesses.

The other three sources of income make up considerably smaller proportions of total income relative to wages, self-employment income, capital gains, and partnership/S-corporation income. The highest proportion observed for ordinary dividends, rental income, and other income is just over 5%; the vast majority of these proportions are well under 2%. In general, the proportion of income from ordinary dividends and rental income are higher for paid-prepared returns versus self-prepared returns (Figures 3.5d and 3.5f, respectively). The proportion of income from other income generally increases across the income distribution (Figure 3.5g), but there is not a clear trend between paid-prepared and self-prepared returns. The fact that higher proportions of total income from these sources are associated with using a paid-preparer is consistent with the empirical evidence that more complicated tax returns are more likely to be prepared by a tax professional (Slemrod and Sorum, 1984; Long and Caudill, 1987). We further analyze the trends in preparer use by income source with regression analysis in Section 3.4.3.

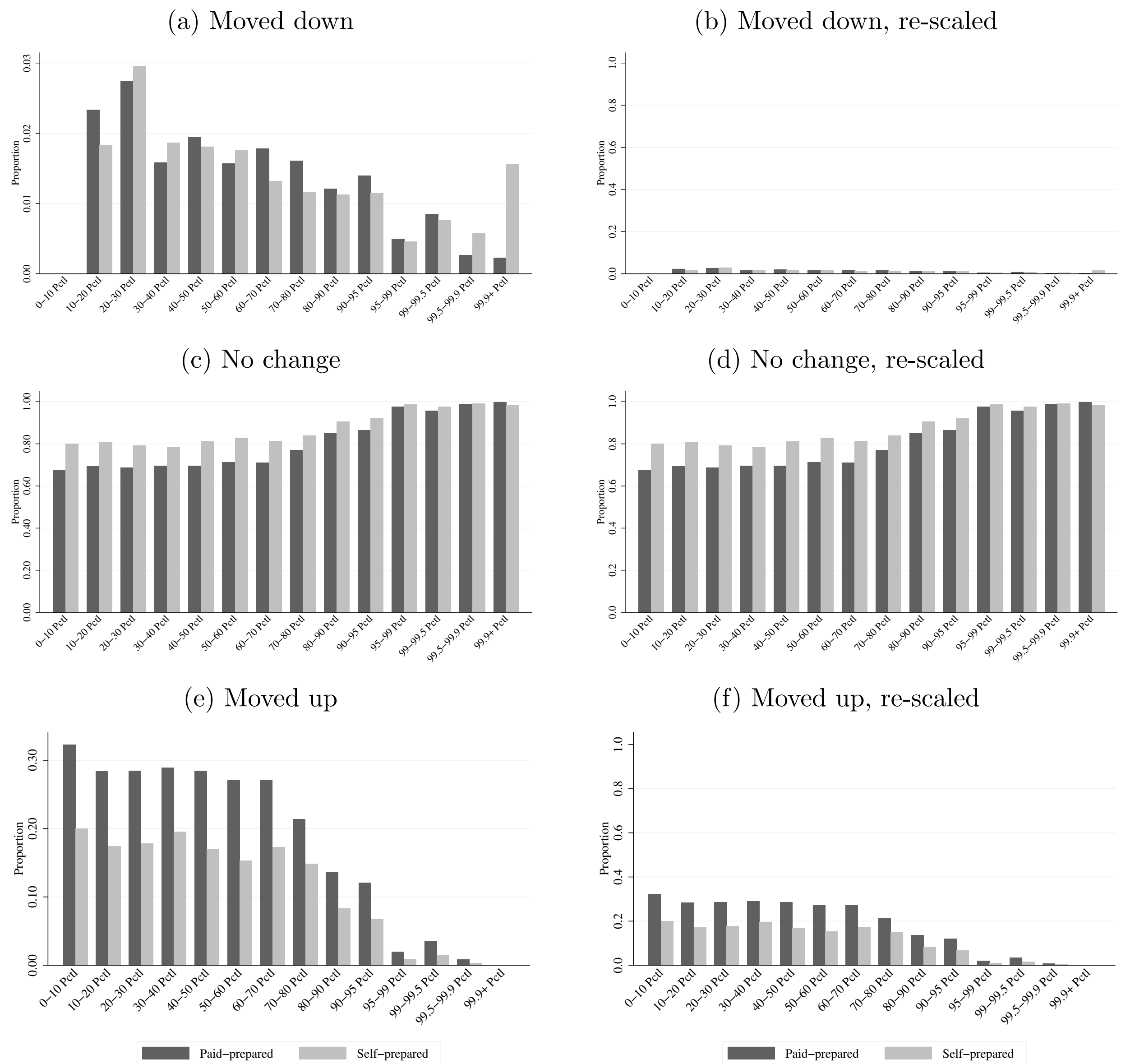
3.3.1.3 Differences in NRP audit corrections

One advantage of the NRP data is that we are able to see both the originally reported amount and the corrected amount for a host of line items. Corrections could either be fixing honest mistakes, or detecting noncompliance. While this is a huge benefit, the corrections post-NRP audit are not a perfect measure of mistakes on tax returns, and provide no information on whether mistakes are voluntary or not. Not all "corrections" are correct, and some errors are easier to detect than others.

We consider how post-NRP audit corrections vary by paid-prepared and self-prepared returns, and how those differences vary across the income distribution. First, we examine the extent to which individuals' corrected AGI percentile group changes after their return is audited. We then analyze how the size of post-audit corrections varies across the income distribution.

Figure 3.6 shows the proportion of individuals whose corrected AGI percentile group went down (a and b), did not change (c and d), or moved up (e and f). For each possible direction of movement, we show a figure with a y-axis that makes it easier to see any trends within a type of movement, and a figure with a y-axis re-scaled to 1 for easier comparison across types of movement. The x-axis of these figures is the taxpayer's original AGI percentile group (as opposed to their corrected AGI percentile group as is used in every other figure in this paper). Note that it is not possible for individuals who were originally in the $0 - 10$ percentile group to move down, nor is it possible for individuals in the 99+ percentile group to move up.

Figure 3.6: Changes in position in AGI distribution post-NRP audits



Notes: N = 114,243. See Table 3.2 for subtotals by corrected AGI percentile. Excludes individuals who either originally reported negative AGI, or who had negative corrected AGI.

Figure 3.6 yields a few important observations. While the likelihood of a taxpayer's corrected AGI percentile group being lower than their original AGI percentile group generally decreases as original AGI percentile group increases (Figure 3.6a), the overall proportion of our sample whose corrected AGI percentile group is lower than their original AGI percentile group is tiny: less than 3% for most original AGI percentile groups (Figure 3.6b).

In contrast, the majority of taxpayers do not change AGI percentile groups after their

NRP audit (Figure 3.6c). The proportion of taxpayers whose AGI percentile group does not change is consistently higher among self-prepared returns (the exception being the 99.9+ percentile group), and consistently higher as original AGI percentile group increases.

Taxpayers who used a paid preparer were consistently more likely to move up in AGI percentile group after their NRP audit than taxpayers whose returns were self-prepared. The differences are considerable: Among the bottom 70% of the original AGI percentiles, there was a 10% difference in the proportion that moved up who used a paid-preparer versus those that had self-prepared returns. More than a quarter of those in the bottom 70% who used a paid preparer saw an increase in their AGI percentile after the NRP audit. The proportion of taxpayers whose AGI percentile went up decreases as original AGI percentile decreases. Unlike with downward movements, this may be mechanical: there are fewer percentile groups to move up the higher up in the distribution one starts.

We also look at the size of corrections across overall income and different types of income. Because the size of corrections is correlated with the size of the originally reported value, we construct a scaled version, as shown in Equation 3.3.1:[11]

$$\text{Scaled correction} = \frac{\text{Corrected amount} - \text{original amount}}{|\text{Average amount for corrected AGI percentile}|} \tag{3.3.1}$$

The larger this value in magnitude, the bigger the correction. Positive values indicate the corrected value was higher than the original; negative values indicate the corrected value was lower than the original. To construct the relevant denominator, we use the finest-grain AGI percentile groups that we are able to given how our AGI percentile data are constructed: 5% increments through 95% (i.e., $0-5$, $5-10$, $10-15$, etc.), then $95-99$, $99-99.5$, $99.5-99.9$, and 99.9+. Finally, we winsorize this value at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles within the same increments used to construct the denominator in order to avoid distortionary effects from outliers.

We start by considering some high-level outcomes in Figure 3.7: total income, adjusted gross income, total taxable income, and total tax liability. In these figures, we group the bottom 99% together and separately consider the $95-99$, $99-99.5$, $99.5-99.9$, and 99.9+ percentile groups, as well as the group with negative corrected AGI.

These figures are striking. Across all four outcomes, the scale is dictated primarily by the bottom 99%. Corrections to total income are larger for self-prepared returns in the top 1%,

[11] We refrain from using the Voluntary Reporting Percentage (VRP) measure used by, e.g., Klepper and Nagin (1989). The VRP for line item i is equal to $\frac{\sum_{n=1}^{N} reported_{n,i}}{\sum_{n=1}^{N} true_{n,i}}$. We want to take advantage of the fact that we have individual-level data, which Klepper and Nagin (1989) did not have. However, there are cases where the corrected amount is 0 when the reported amount is not 0. As a result, the VRP measure cannot be constructed at the individual level.

but larger for paid-prepared returns in the bottom 99% (Figure 3.7a). The scaled corrections decrease with each increasing top 1% group, with paid-prepared returns among the top 1% experiencing less than a 5% increase in total income on average (with the value being less than 1% for the 99.9+ percentile).

This pattern is mirrored for both AGI (Figure 3.7b) and total tax liability (Figure 3.7d), with scaled corrections in general being much larger for the bottom 99%, and larger for paid-prepared returns among the bottom 99%, whereas scaled corrections are larger for self-prepared returns among the top 1%.

Scaled corrections to taxable income are given in Figure 3.7c.[12] We again see that the magnitude of scaled corrections is considerably larger among the bottom 99%, with positive scaled corrections for paid-prepared returns but negative scaled corrections for self-prepared returns. In contrast, scaled corrections are positive across the cuts of the top 1%, but again larger for self-prepared returns.

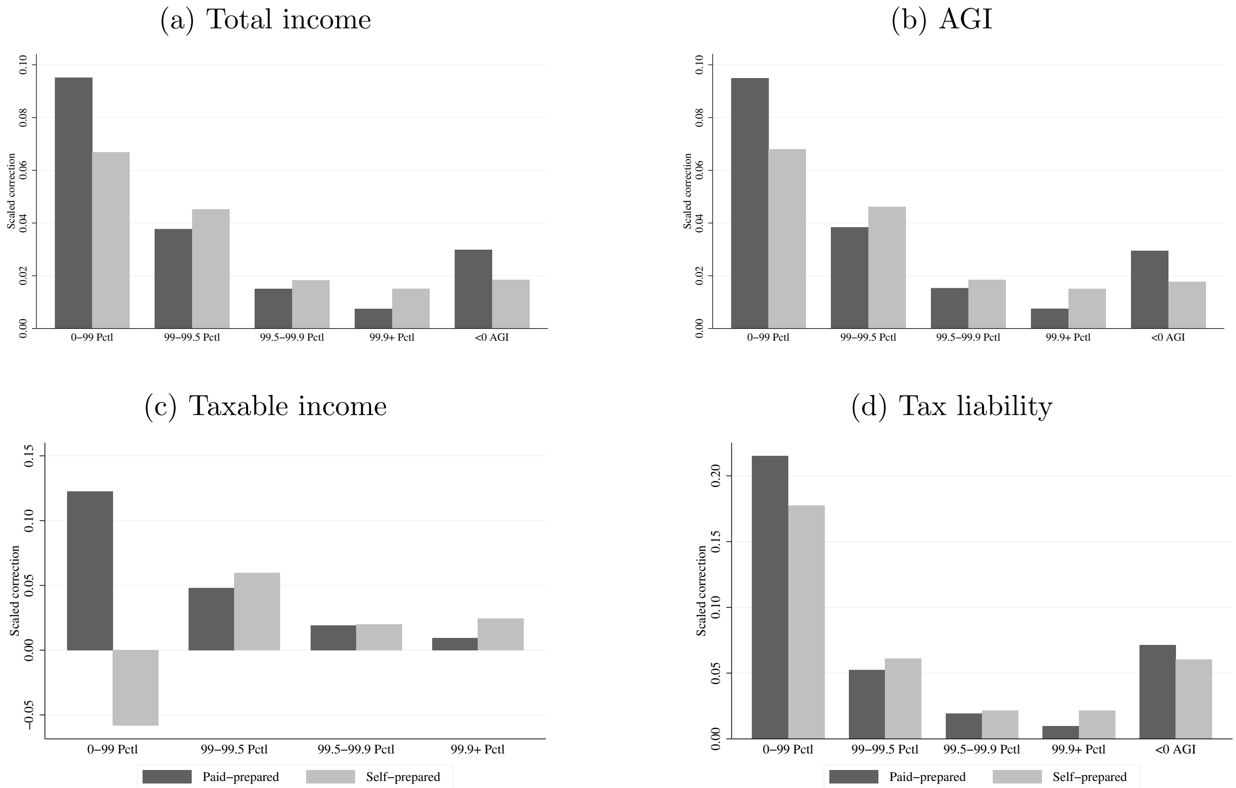
The relative size of scaled corrections among the top 1% compared to the bottom 99% may be overstated if random audits fail to detect under-reported income from sources such as undeclared foreign bank accounts and complicated pass-through business structures (Guyton et al., 2021). Individuals in the top 1% are more likely to have income from other sources that are not included in post-NRP audit corrections, which would artificially reduce the size of their scaled corrections relative to the bottom 99%. To the extent tax professionals assist in hiding income, this would also influence the relative size of scaled corrections between self-prepared and paid-prepared returns within the top 1% subgroups. In particular, it may no longer be true that scaled corrections were higher for self-prepared returns (and it may even be true that scaled corrections would be higher for paid-prepared returns).

We next consider the size of corrections across different types of income in Figure 3.8. As above, with these figures we group the bottom 99% together and separately consider the $95-99$, $99-99.5$, $99.5-99.9$, and 99.9+ percentile groups, as well as the group with negative corrected AGI.

There is considerable variation in the size of scaled corrections by income type. The largest corrections to wage income are seen by self-prepared returns in the bottom 99%, but these are, on average, less than 0.06% scaled corrections. (Figure 3.8a). Scaled corrections to wages are well under 0.03% within the top 1% of the income distribution, with corrections being generally smaller among paid-prepared returns. It is not surprising that the scale of these corrections is so small, given that wage income is subject to third-party reporting

[12] Figure 3.7c excludes individuals in the $0-5$ corrected AGI percentile group from the bottom 99% because the values were extremely distortionary. Note that there is no negative AGI group for this figure because these individuals do not have taxable income.

Figure 3.7: Size of scaled corrections, overall



Notes: Figure 3.7a N = 117,303. Figure 3.7b N = 117,303. Figure 3.7c N = 115,546. Figure 3.7d N = 117,303. See Table 3.2 for subtotals by corrected AGI percentile. These figures show the mean value for the scaled correction as calculated in Equation 3.3.1. The larger this value in magnitude, the bigger the correction. Positive values indicate the corrected value was higher than the original; negative values indicate the corrected value was lower than the original. The analyzes values are winsorized at the 1$^{\text{st}}$ and 99$^{\text{th}}$ percentiles within the finest-grain AGI percentile groups in our data.

through W-2s. This result is consistent with the literature, which finds that taxpayers are generally compliant when it comes to income subject to third-party reporting (Kleven et al., 2011; Phillips, 2014).

In contrast, there are relatively large scaled corrections made to self-employment income (Figure 3.8b). The corrections are larger among paid-prepared returns in the bottom 99%, and generally larger among self-prepared returns in the top 1%. As with wage income, the size of the scaled corrections decreases as corrected AGI percentile group increases.

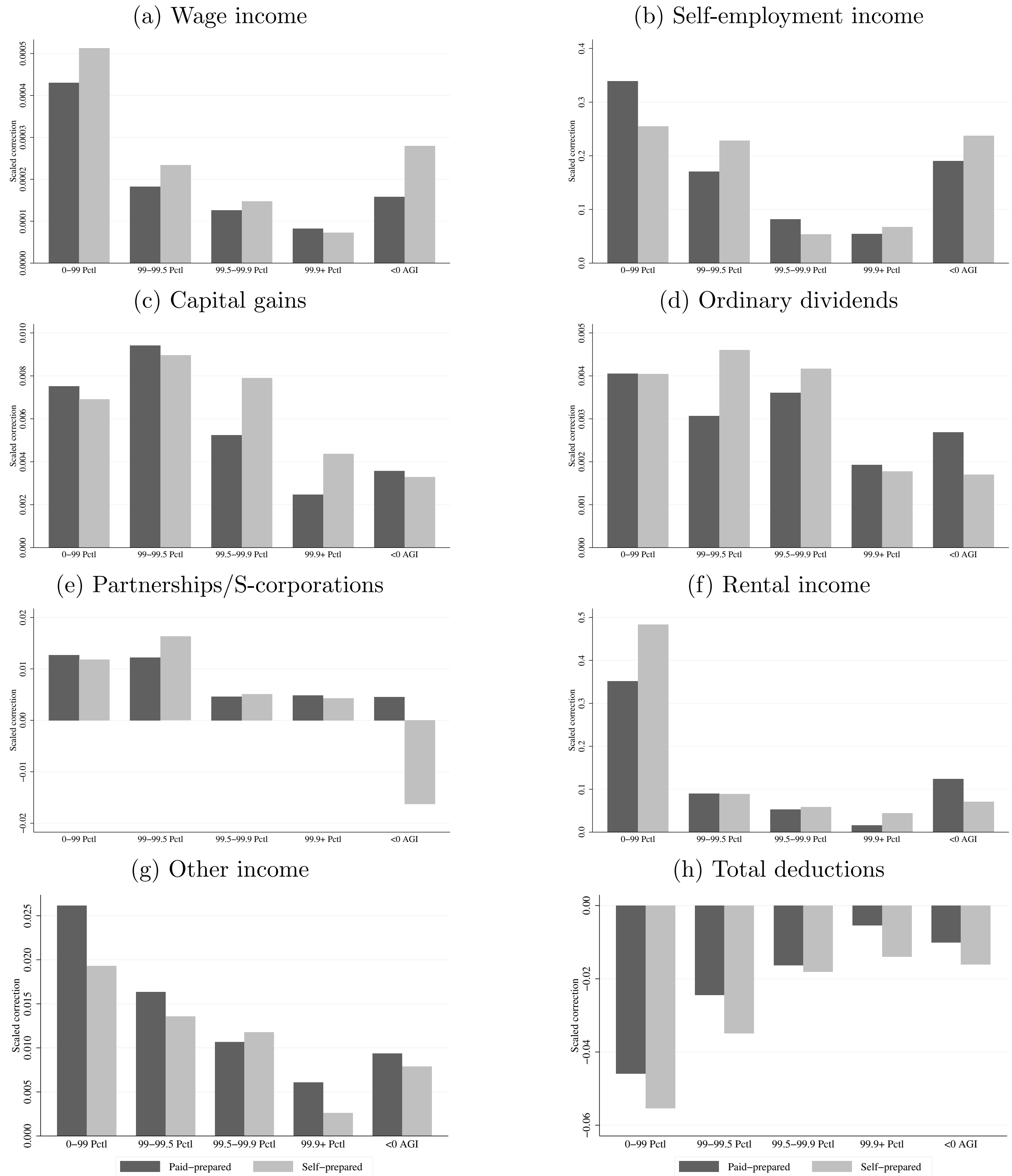
We next turn to Schedule D income: capital gains (Figure 3.8c) and ordinary dividends (Figure 3.8d). The size of scaled corrections for both of these types of income is an order of magnitude smaller than self-employment income, but 1-2 orders of magnitude greater than scaled corrections to wage income. Capital gains is the one case where average scaled corrections among the top 1% are higher than in the bottom 99% (in particular, among the $99 - 99.9$ percentile group). There is not a huge difference in scaled corrections between self-prepared and paid-prepared returns among the bottom 99% and the $99 - 99.5$ percentile group; among the $99.5 - 99.9$ and 99.9+ percentile groups, the scaled corrections among self-prepared returns are noticeably larger. In general, scaled corrections made to paid-prepared returns are higher than for self-prepared returns for other income, except among the $99.5 - 99.9$ percentile group. There are similarly very small adjustments made to ordinary dividends as to wage income, with similar trends across the income distribution and between self-prepared and paid-prepared returns (Figure 3.8d).

From Schedule E, we consider income from partnerships and S-corporations (Figure 3.5e) and rental income (Figure 3.8f). There are not large discrepancies between scaled corrections between paid-prepared and self-prepared in the top 1% for either type of income. For rental income, there is a striking difference between the scale of corrections between the bottom 99% and the top 1%. Across the income distribution, self-prepared returns have larger scaled corrections to rental income than paid-prepared returns.

Finally, we consider corrections to other income not reported elsewhere in the tax return (Figure 3.8g) and changes to total deductions (Figure 3.8h). Across the income distribution, paid-prepared returns generally have higher scaled corrections to other income, though the magnitude is small (an order of magnitude smaller than the size of corrections to self-employment income). In contrast, the size of the scaled deductions is larger for self-prepared returns than for paid-prepared. As with the types of income, the magnitude of the change indicates the size of the correction (a larger value indicates a bigger change post-audit), though in this case, the negative values mean that the corrected deduction amounts were, on average, lower than the original deduction amounts (i.e., total deductions went down).

There are several key takeaways from Figures 3.7 and 3.8. The first is that scaled corrections are larger overall for the bottom 99% than for the top 1%. This is true both across measures of comprehensive income and individual income types, as well as total deductions. Second, scaled corrections to comprehensive measures of income are larger for paid-prepared returns among the bottom 99%, but smaller for paid-prepared returns among the top 1%. This is consistent with the idea that preparers hired by the top 1% may be better at correcting mistakes and/or hiding noncompliance. We discuss differences in trends among paid-preparers hired across the income distribution in Section 3.3.2. Third, the scale of corrections varies widely between different income types. In particular, the scale for corrections to wages (which are subject to third-party reporting through the submission of W-2s) is one order of magnitude smaller than the scale of corrections to capital gains and ordinary dividends, two orders of magnitude smaller than corrections to income from partnerships and S-corporations, and three orders of magnitude smaller than the scale of corrections to self-employment income

Figure 3.8: Size of scaled corrections, by income type



Notes: Figure 3.8a N = 117,303. Figure 3.8b N = 117,303. Figure 3.8c N = 117,303. Figure 3.8d N = 117,303. Figure 3.8e N = 45,562. Figure 3.8f N = 45,562. Figure 3.8g N = 117,303. Figure 3.8h N = 117,303. See Table 3.2 for subtotals by corrected AGI percentile. These figures show the mean value for the scaled correction as calculated in Equation 3.3.1. The larger this value in magnitude, the bigger the correction. Positive values indicate the corrected value was higher than the original; negative values indicate the corrected value was lower than the original. The analyzes values are winsorized at the 1^{st} and 99^{th} percentiles within the finest-grain AGI percentile groups in our data.

and rental income. This has important implications for which individuals we might expect to hire a preparer. If self-prepared returns are subject to the same very small scale corrections for wage income, we might expect individuals with a greater proportion of their income coming from wages to be less likely to use professional tax preparation services.

3.3.2 Conditional on using a preparer, how does preparer use vary across the income distribution?

In addition to analyzing how the difference between paid-prepared and self-prepared returns varies across the income distribution, we can also examine trends specific to paid-prepared returns. We consider three dimensions along that may vary across the income distribution: average fees paid to preparers, the average number of returns prepared by the preparers used, and the professions held by the preparers used.

3.3.2.1 Differences in fees paid to tax preparers

Before the Tax Cuts and Jobs Act of 2017, tax preparation fees were tax deductible and listed on Schedule A of the individual tax return. The average amount of money deducted for tax preparation fees varies widely across the income distribution, as shown in Figure 3.9. It is important to remember that we are only able to observe preparer fees for individuals who itemized their deductions; as we showed in Figure 3.3, many individuals in the lower end of the income distribution do not itemize their deductions even if they use a paid preparer.

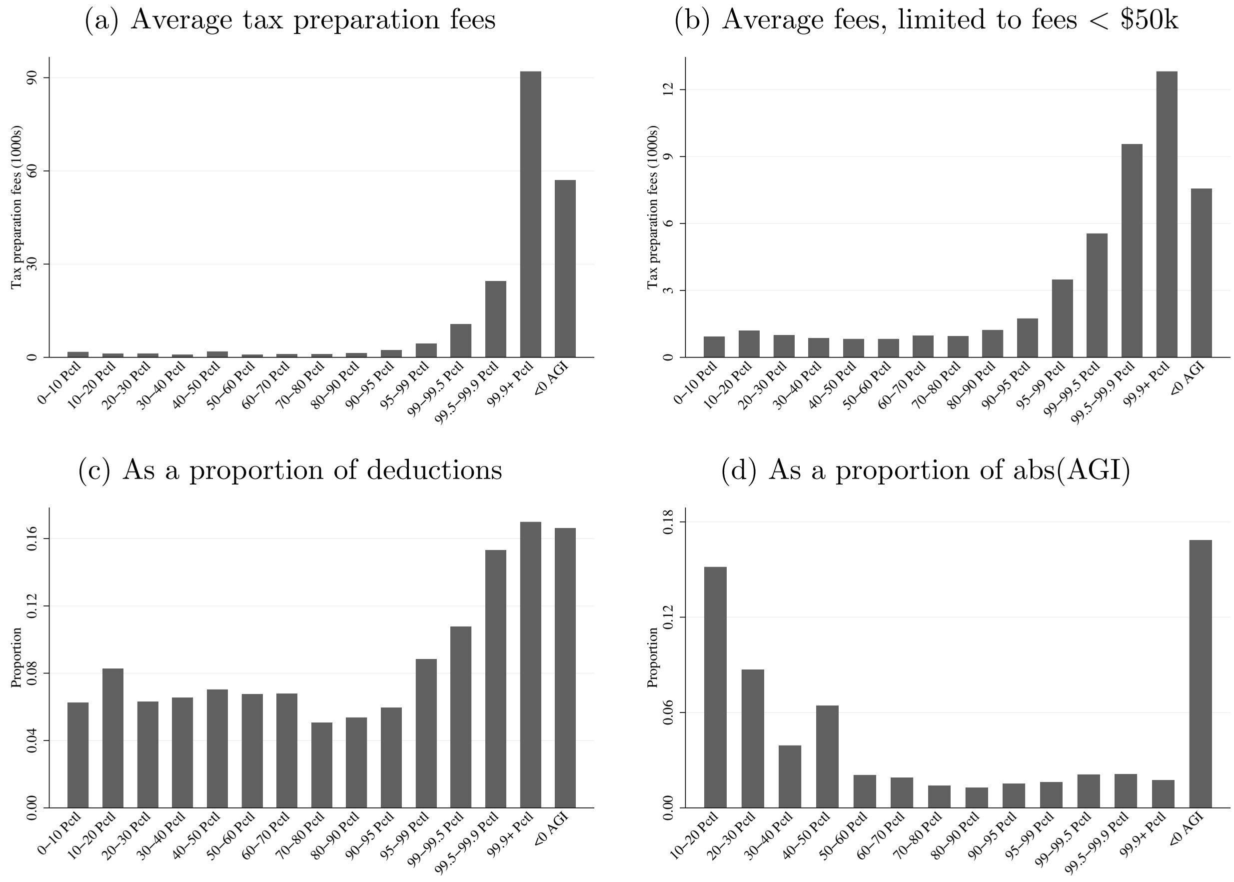
We present four alternative measures of the fees paid: (a) the average amount of tax preparation fees deducted, (b) the average amount of tax preparation fees deducted when limited to fees less than \$50,000, (c) fees as a proportion of total deductions, and (d) fees as a proportion of the (absolute value of) AGI.

The average amount of fees deducted increases considerably with corrected AGI percentile. The average fees among the bottom 99% do not break \$5,000 (and are generally considerably lower than that), while the average fees are over \$10,000, \$25,000, and \$90,000 for the $99-99.5$, $99.5-99.9$, and 99.9+ percentile groups, respectively. The group with the second highest average fees is the negative AGI group, with average fees deducted of over \$55,000.

Even when we exclude fees over \$50,000 (Figure 3.9b), the average fees paid increases dramatically in the top 1% relative to the bottom 99% of the income distribution. The average fees paid for the $99-99.5$ percentile group are nearly double those paid by the $95-99$ percentile group; the average fees paid by the $99.5-99.9$ and 99.9+ percentile groups are nearly triple and quadruple those paid by the $95-99$ percentile group.

We also consider two scaled versions of the fees paid: fees as a proportion of total

Figure 3.9: Tax preparation fees by corrected AGI percentile



Notes: N = 25,753. Limited to returns with a paid preparer with greater than 0 estimated fees paid for professional tax services. See Table 3.2 for subtotals by corrected AGI percentile.

deductions (Figure 3.9c) and fees as a proportion of (the absolute value of) AGI (Figure 3.9d). In the first case, we again see the trend that higher corrected AGI percentiles have a higher amount of their total deductions be due to tax preparation fees (with the top 1% attributing over 10% of their total deductions to tax preparations fees, on average).

In contrast, tax preparation fees tend to decrease as a proportion of AGI, as corrected AGI percentile increases. Income deciles above the median tend to spend a relatively constant proportion of their income on tax preparation fees (around 2-3%), which is half (or even less) than that spent by income decile groups below the median.[13] The proportion of income spent on tax preparation services may be overestimated if our data do not fully capture

[13]This final figure excludes the $0-10$ percentile group because it distorts the graph beyond readability. This decile group seems to spend an inordinate amount on tax preparers: 200% of their AGI on average. We suspect this group might include a few individuals similar to those declaring negative AGI (in that they have low AGI due to losses rather than, e.g., low wages). If true, this could artificially inflate this measure of fees paid by pairing high fees in the numerator with low AGI at the denominator, distorting the average. The version of the figure with the $0-10$ percentile group included is given in Figure C.3 in Appendix C.2.3.

under-reported income. If true, this may impact our estimates for the top 1% more than other income groups depending on the types of undetected income.

Overall, these results highlight that spending on tax preparation fees increases quite dramatically at top incomes. While this observation holds when considering fees as a percentage of overall deductions, the same does not hold for fees as a percentage of AGI. This last difference might be indicative of the fact that what is a "typical" tax return varies throughout the income distribution.

3.3.2.2 Average number of returns prepared

Figure 3.10 shows the average number of returns signed by the preparers used in each corrected AGI percentile group.[14] As corrected AGI percentile group increases, the average number of returns prepared by the preparers used by that percentile group generally decreases. While there is no substantial variation in the bottom 90%, the rate of decrease in the average number of returns prepared becomes larger in the top 10%, so that the average number of returns prepared by a preparer used in the top 0.1% is about half that of the average number prepared by preparers used in the middle of the income distribution.

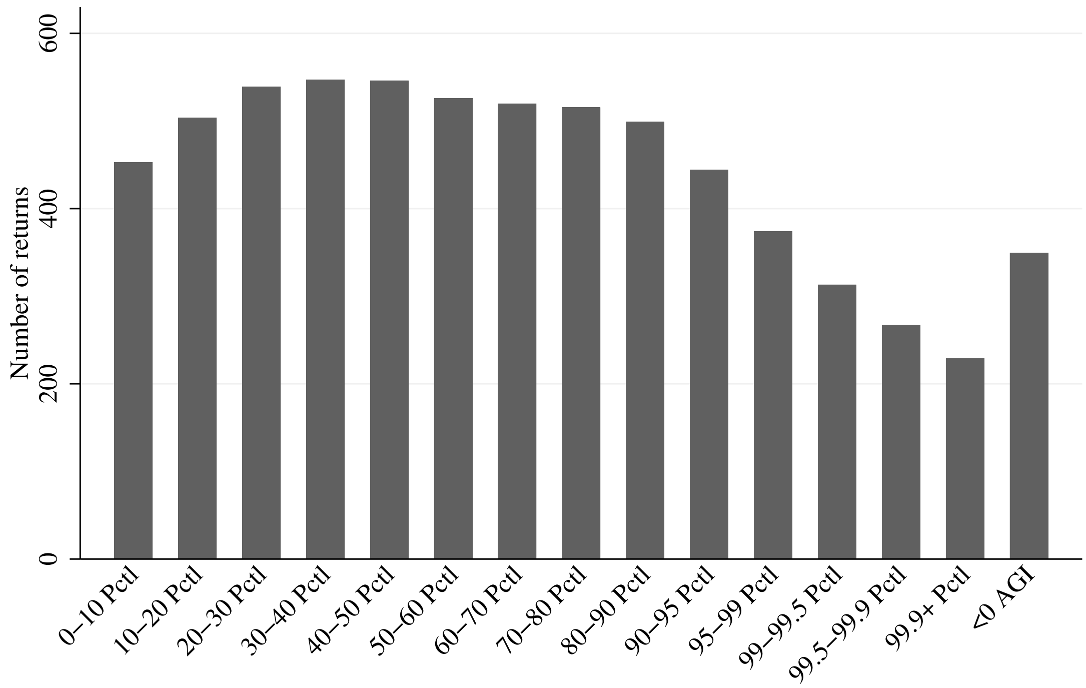
While preparers used by the top 1% tend to prepare fewer returns on average, they also elicit considerably higher fees (as we saw in Figure 3.9a). What is unknown is why these professionals prepare fewer returns. It could be because the tax returns of the top 1% are more complicated. Alternatively (or possibly in addition), it could be that the top 1% expect more attention to be paid to their return given the amount of money they are spending on the service. Either of these interpretations is consistent with the notion that paid tax preparation services for the richest taxpayers may be a different product than tax preparation services among the bottom 99%, which requires more time and specialization but yields considerably higher pay.

3.3.2.3 Tax preparer professions

Several professions may serve as paid tax preparers: Certified Public Accountants (CPAs), Enrolled Agents, tax attorneys, and Supervised Registered Tax Preparers (SRTPs). A CPA holds a license to offer accounting services to the public in a particular state. Tax attorneys are lawyers concerned with issues related to tax liability and taxation. Like CPAs, attorneys

[14]We exclude the handful of preparer ID numbers associated with 30,000 or more returns. We present the same figure as Figure 3.10 with alternative cut-offs of 1,000, 5,000, and 10,000 in Appendix C.2.4. When we use these alternative cut-offs, the trends mirror what we observe in Figure 3.10: as corrected AGI percentile group increases, the average number of returns prepared by the preparers used by that percentile group decreases.

Figure 3.10: Average number of returns prepared, by corrected AGI percentile



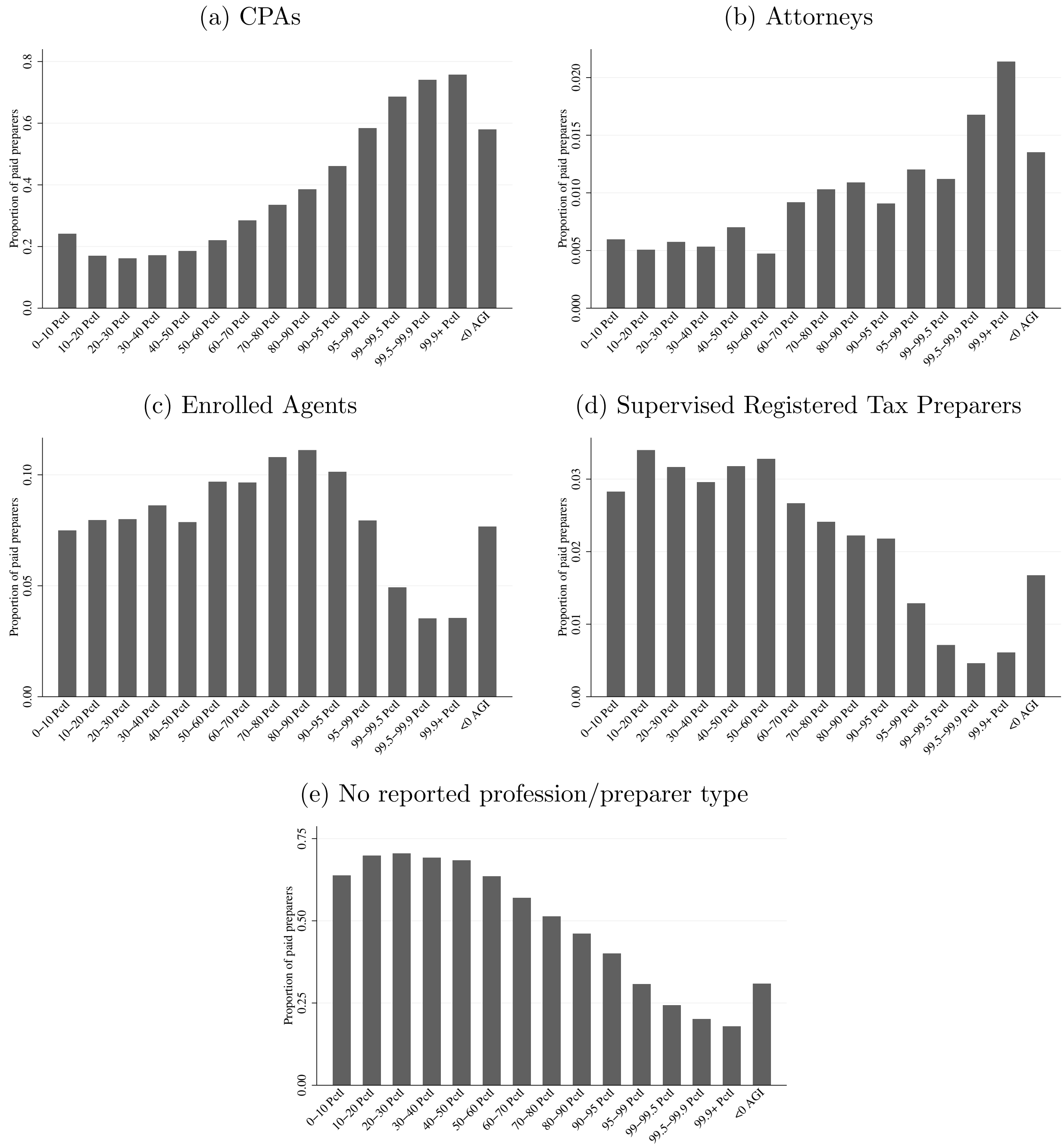
Notes: N = 83,308. See Table 3.2 for subtotals by corrected AGI percentile. Excludes preparers who are listed as preparing 30,000 or more returns.

are licensed to practice in a particular state. Enrolled Agents are tax advisors who have been federally authorized by the U.S. Department of the Treasury. The "Enrolled Agent" credential is recognized in all 50 states. SRTPs do not sign tax returns themselves; instead, they work for an organization owned by CPAs, attorneys, and/or Enrolled Agents, and are supervised by a CPA, attorney, or Enrolled Agent with a PTIN.

Figure 3.11 shows the proportion of paid preparers that fall into each of these four categories by income group, as well as the proportion that is not associated with any of these four titles.[15] We observe that higher income groups are considerably more likely to hire a preparer that is a CPA (Figure 3.11a) or attorney (Figure 3.11b) relative to lower income groups, while lower income groups are more likely to have their returns prepared by an Enrolled Agent (Figure 3.11c) or SRTP (Figure 3.11d). Lower income groups are also much more likely to have their return prepared by someone who does not self-report their profession as any of these four (Figure 3.11e). These figures make it clear that different income groups are more or less likely to hire different types of tax preparers, and reinforces

[15]The IRS began collecting this information through self-reporting by tax preparers in 2011. We are not able to identify the profession of our full set of tax preparers both because our sample starts in 2006, and because the data are self-reported.

Figure 3.11: Tax preparer professions



Notes: $N = 84{,}769$. See Table 3.2 for subtotals by corrected AGI percentile. In our data, we observe 33,849 CPAs, 838 attorneys, 6,979 Enrolled Agents, and 1,785 Supervised Registered Tax Preparers. Of the tax preparers included in our data, 39,848 did not self-report a profession or preparer type.

our conjecture at the end of Section 3.3.2.2: tax preparation services used by the top 1% look like a different product than that used by the bottom 99%.

3.4 Why would someone in the top 1% not use a tax preparer?

Given that paid preparers seem to lead to smaller corrections for the top 1% (so either do a better job at not making mistakes or at finding grey areas), why don't we observe everyone in the top 1% of the income distribution using a paid preparer?

In section 3.4.1, we extend the Allingham and Sandmo (1972b)-style model developed by Guyton et al. (2021) to incorporate the decision to hire a tax preparer. We analyze this stylized model to provide intuition and justification for the empirical findings we present in Sections 3.4.2 and 3.4.3. There are two important caveats to this model. First, the model assumes "effective concealment" is the only reason a taxpayer will hire a tax preparer. In reality, taxpayers might have other reasons to hire a tax preparer, such as opportunity cost or a desire to decrease compliance costs. Second, the model is silent about whether the tax preparers "knowingly" increase noncompliance or not. As a decrease in the probability of detection as a result of hiring a tax preparer does not have to be interpreted as "malicious activity."

3.4.1 Intuition from a stylized model

The set-up of the model is as follows. Individuals decide on how much of their true exogenous income (y) to evade, where evasion is denoted by e. Individuals can take a binary concealment action $a \in \{0, 1\}$ which decreases the probability p of that any evasion is detected. The concealment action has a fixed cost, κ. We interpret the concealment decision to be the decision to hire a professional tax preparer who can help the taxpayer evade more successfully.

Our descriptive analysis suggests that the top 1% and bottom 99% of the income distribution do not use the same type of tax preparer.[16] We extend the model to include a third action which will decrease the probability of detection further, but at a higher cost. In other words, while choosing how much to evade the individual can also take action $a \in \{0, 1, 2\}$ such that $p_0 > p_1 > p_2$ and $\kappa_2 > \kappa_1 > \kappa_0 = 0$ [17]. We refer to the action $a = 0$ as self-preparation, the action $a = 1$ as hiring a generalized tax preparer, and the action $a = 2$ as hiring a specialized tax preparer.

We make a few assumptions as part of differentiating between generalized and specialized tax preparers. First, we assume that specialized tax preparers are better able to reduce the probability of detection of evasion. Second, we assume that the fixed cost of hiring specialized tax preparers is greater than that of hiring a generalized tax preparer. This assumption is

[16]See, e.g. Figure 3.9 and Figure 3.11.

[17]The assumption of $\kappa_0 = 0$, can easily be relaxed and the results will be unaffected. To be consistent with the original Allingham-Sandmo model and Guyton et al. (2021), we keep this assumption for the rest of the model

likely to be true if specialized tax preparers spend more time on each return they prepare, resulting in those preparers accepting fewer clients and thus requiring higher fees for their services.

The individual's optimization problem is given as follows.

$$\max_{\{e\in[0,y],a\in\{0,1,2\}\}}(1-p(a))u((1-\tau)y+\tau e-\kappa_a)+p(a)u((1-\tau)y-\tau e\theta-\kappa_a) \qquad (3.4.1)$$

subject to

$$(1-\tau)y-\tau e\theta-\kappa_a>0 \quad \forall a,$$

where τ denotes the income tax rate and $\theta > 0$ is the penalty rate levied on the evaded tax. $u(\cdot)$ is a standard utility function that is strictly increasing and concave.

Let $g_a(y, p)$ denote the optimal income evaded ratio if agent were to choose $a \in \{0, 1, 2\}$. When $a = 0$, the model boils down to the standard Allingham-Sandmo framework. As a result, the behavior of optimal evasion and the fraction of income evaded decisions when $a = 0$ are the same as summarized in Lemma 1 of Guyton et al. (2021). Let $A(c) := -\frac{u''(c)}{u'(c)}$ and $R(c) := -\frac{cu"(c)}{u'(c)}$ denote the absolute and relative risk aversion. We present a generalized version of the lemma below.

Lemma 1. *The Allingham-Sandmo Tax Gap*

- *If the individual is risk averse, for any $a \in \{0, 1, 2\}$, g_a is decreasing over p, $\frac{\partial g_a}{\partial p} < 0$*
- *If absolute risk aversion is decreasing, $A' < 0$, evasion e is increasing in true income y.*
- *If relative risk aversion is constant, $R' = 0$, for any $a \in \{0, 1, 2\}$, g_a is constant over income y, $\frac{\partial g_a}{\partial y} = 0$.*
- *If relative risk aversion is decreasing, $R' < 0$, for any $a \in \{0, 1, 2\}$, g_a is decreasing over income y, $\frac{\partial g_a}{\partial y} < 0$.*
- *If relative risk aversion is increasing, $R' > 0$, for any $a \in \{0, 1, 2\}$, g_a is increasing over income y, $\frac{\partial g_a}{\partial y} > 0$.*

Proof. See Allingham and Sandmo (1972b) for the case of $a = 0$. However, all else constant, $a = 0$ can be trivially extended to $a = 1$ and $a = 2$ cases. □

In order to accommodate the third costly action we include above, we introduce the following modified Assumption 1 and Lemma 2:

Assumption 1. *As y becomes arbitrarily large, $g_0(y, p_1)$ and $g_1(y, p_2)$ approach strictly positive constants respectively.*

Lemma 2. *Under Assumption 1, as y becomes arbitrarily large:*

- $g_1(y,p_1) - g_0(y,p_1)$ *converges to* 0,
- $g_2(y,p_2) - g_1(y,p_2)$ *converges to* 0.

Lemma 2 suggests that as income y increases, given a fixed detection probability, the fixed cost of hiring a tax preparer and the additional fixed cost of hiring a "better" tax preparer diminishes.

Further, in order to guarantee that each action will be optimal for some agent, we need to ensure that no action is particularly "disadvantaged" with respect to the cost-benefit bundle that it offers. The following assumptions ensures that the cost for a generalized preparer is not too high:

Assumption 2.

$$\begin{aligned}\kappa_1 < min\Bigg\{ & \frac{u'\big(c_N(p_1,0,y)\big)}{u'\big(c_N(p_1,\kappa_1,y)\big)}\frac{(p_0-p_1)}{1-p_1}\tau\theta g(p_0,\kappa_0,y)y, \\ & \kappa_2 - \left(\frac{u'\big(c_D(p_1,\kappa_1,y)\big)}{u'\big(c_D(p_2,\kappa_2,y)\big)}\frac{(p_1-p_2)}{p_1}\tau g(p_2,\kappa_2,y)y\right)\Bigg\}\end{aligned} \tag{3.4.2}$$

We can now state the main result:

Proposition 1. *Under Assumption 1,*

- *There is a cutoff income level, y', such that all else equal, agents with income higher than this cutoff will prefer to hire a generalized tax preparer over self-preparation,*
- *There is a cutoff income level, y'', such that all else equal, agents with income higher than this cutoff will prefer to hire a specialized tax preparer over generalized tax preparer.*
- *If, additionally, Assumption 2 is satisfied, then $y'' > y'$.*

The first statement of the proposition is equivalent to Proposition 1 in Guyton et al. (2021). The second statement is proved by a slight modification of the proof for the first statement. The proofs are included in Appendix C.3. Note that Proposition 1 implies that under Assumption 2, an individual with income $y \in [y', y'']$ will choose to hire a generalized tax preparer. On the other hand, the proposition does not prescribe an optimal action for rich individuals directly. Instead, it only states that as the income of taxpayers increases they are more likely to hire specialized taxpayers. Corollary 1 characterizes the optimal hiring decision for the very rich:

Corollary 1. *Under Assumption 1, it is optimal for a taxpayer with income $y > y''$ to hire a specialized tax preparer, i.e., choose action $a = 2$.*

This corollary immediately follows from the second statement of Proposition 1 and a modification of the proof of the first statement. Intuitively, it is plausible to think that hiring decision increases with income. In other words, if a taxpayer does not find it optimal to self-prepare another taxpayer with higher income will also not find self-preparing optimal. Therefore, for taxpayers with income $y \geq y''$ it is better to hire a specialized preparer over a generalized tax preparer, which is better than self-preparing.

In the next two subsections, we further modify the model to provide a possible theoretical rationale for why some individuals with $y > y''$ will choose to self-prepare ($a = 0$) instead of hiring a specialized tax preparer ($a = 2$) as suggested by Corollary 1.

3.4.1.1 The effect of specialized occupation: accountants

So far we have assumed that the probability of detection can only be influenced through the action of hiring different types of tax preparers. Specifically, we assumed that all self-preparing agents are subject to the same detection probability, p_0. In reality, it is plausible that some taxpayers might be "better" at self-preparing their tax returns compared to others. For example, if some taxpayers are specialized accountants or tax preparers themselves they might be subject to the same lower probability of detection without incurring the cost of paying for services of another tax preparer.[18]

Let $\alpha \in \{0, 1\}$ be an indicator function of whether the taxpayer is a specialized accountant or not, so $\alpha = 1$ if the taxpayer is a specialized accountant and $\alpha = 0$ otherwise. Suppose the probability of detection, p, is not only a function of a, the decision of hiring a tax preparer, but also a function of α, i.e. $p(a, \alpha)$.

Suppose the ranking of probability of detection is as follows:

$$p(0, 0) < p(1, 0) < p(2, 0) = p(2, 1) = p(0, 1) = p(1, 1).$$

This relationship implies that if a taxpayer is a specialized accountant themselves, they cannot decrease the probability of detection further by hiring someone else.[19] Under this assumption, no specialized accountant will hire a tax preparer as doing so provides no benefits but results

[18] Here we are making a simplifying assumption that the self-preparation cost for an accountant is strictly less than hiring an external preparer, which would follow from the presence of any transaction cost or any asymmetric information about whether or not the tax preparer was a specialized tax preparer.

[19] Another way to interpret this relationship is to assume that these are not actual probabilities but beliefs of the taxpayer. Under this interpretation, taxpayers who are accountants do not have to be specialized, but rather believe that they are specialized and, in turn, cannot gain additional benefit by hiring a tax preparer.

in a cost. If this is true, we should observe similar benefits to hiring a paid preparer among the top 1% as we do for self-prepared returns in the top 1% where the individual was an accountant. We test this hypothesis in Section 3.4.2.

3.4.1.2 The effect of the share of income subject to third-party information

The baseline model assumes that the evasion decision of the individual is bounded by the true income y. Empirically, individuals are less likely to evade (or try to evade) income subject to third-party reporting (see, e.g., Kleven et al. (2011); Slemrod et al. (2017)). As Guyton et al. (2021) suggest, it is straightforward to extend the baseline model to include third-party reporting considerations by assuming a bound on possible evasion. Specifically, one can assume that an individual cannot evade income that is subject to third-party reporting and not shiftable to any other category. Consequently, redefining the variable y as income that can be evaded will yield the same results as the baseline model, but with different interpretations. Specifically, Proposition 1 and Corollary 1 will imply that any taxpayer with income level above y'' not subject to third-party reporting will hire a specialized tax preparer. There are several follow-up implications to this interpretation, which are summarized in Corollary 2:

Corollary 2. *Under Assumption* 1*, and the case y denotes income not subject to third-party reporting,*

- *A taxpayer can have a substantial total income, but will hire a generalized tax preparer if their income not subject to third-party reporting is greater than y' but less than y''*
- *A taxpayer can have a substantial total income but will choose to self-prepare if their income not subject to third-party reporting is less than y'*

The proof of the corollary immediately follows from reinterpreting y as income not subject to third-party reporting. This corollary predicts that individuals with a larger share of income subject to third-party reporting are more likely to self-prepare, even if their total income puts them in the top of the income distribution. We turn to this hypothesis in Section 3.4.3.

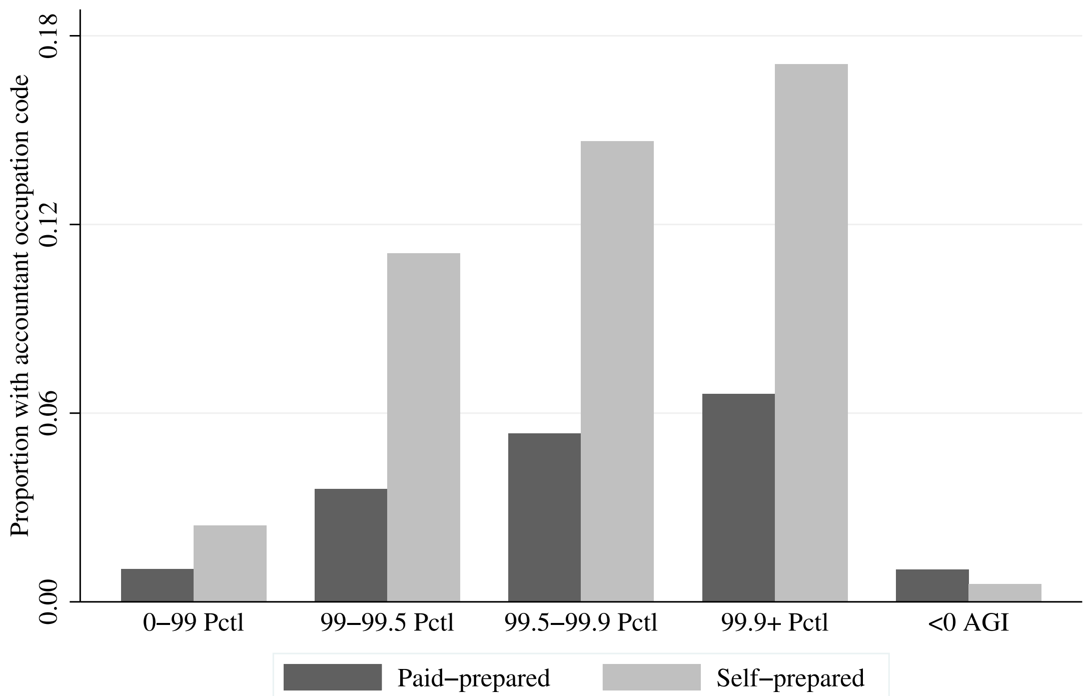
3.4.2 Testing extension 1: the top 1% who are also accountants

One explanation for why individuals in the top 1% of the income distribution would not use a paid tax preparer is because they themselves are an accountant or financial planner, and therefore expect they would be able to provide themselves with as good (or better) service for the money than paying someone else.[20]

[20] There is, of course, the question of opportunity cost–but some people find tasks like the puzzle of their tax returns compelling.

We show the proportion of individuals with an occupation code in the administrative tax data that indicates that their job concerns accountant or some type of financial planning in Figure 3.12. It is clear that, across the income distribution, individuals who are themselves accountants are more likely to self-prepare their returns. In addition, the proportion of individuals who are accountants is considerably higher in the top 1% relative to the bottom 99%. Even within the top 1%, the share of individuals who are accountants increases as corrected AGI percentile group increases, and the relative share of self-prepared returns that are prepared by accountants increases.

Figure 3.12: Occupation code indicating accountant or financial management



Notes: N = 111,603 individuals. See Table 3.2 for subtotals by corrected AGI percentile.

We can perform the same analysis more formally. In Table 3.3, we show the results of running Equation 3.4.3 over several sub-populations:

$$\text{Paid}_i = \alpha_0 + \alpha_1(\text{Accountant}_i = 1) + u_i \tag{3.4.3}$$

We find that the probability of using a paid preparer significantly decreases for individuals with an accounting occupation code, and that this decrease in probability is smaller at higher levels of the income distribution. At the same time, the baseline probability of using a paid preparer (α_0) increases at higher levels of the income distribution.

We then ask whether being an accountant in the top 1% of the income distribution has a similar effect on the size of post-NRP audit scaled corrections to various outcomes.[21] In Section 3.3.1, we found that, in general, self-prepared returns had higher corrections in the top 1% relative to paid-prepared returns. We run the regression shown in Equation 3.4.4 over a variety of outcomes y:

$$\begin{aligned} y_i = \beta_0 &+ \beta_1(\text{Paid}_i = 1) \\ &+ \beta_2(\text{Paid}_i = 1 \;\&\; \text{Top 1\%} = 1) \\ &+ \beta_3(\text{Paid}_i = 1 \;\&\; < 0 \text{ AGI} = 1) \\ &+ \beta_4(\text{Self}_i = 1 \;\&\; \text{Top 1\%} = 1 \;\&\; \text{Accountant} = 1) + \varepsilon_i \end{aligned} \tag{3.4.4}$$

We run this regression using indicators for the top 1% grouped together, and for the top 1% split into the $99 - 99.5$, $99.5 - 99.5$, and 99.9+ percentile subgroups.

Table 3.4 shows the "overall' results (total income, AGI, taxable income, and tax liability), similar to Figure 3.7). While the scaled corrections to total income, AGI, and tax liability are higher when using a paid preparer in the bottom 99%, the scaled correction is lower for the top 1%. The coefficients on the interaction terms between "Paid" and the top 1% percentile groups are statistically significant at the 0.1% level, and increasing (in magnitude) as percentile increases. In other words, using a paid preparer leads to smaller corrections among the top 1%; the higher up in the income distribution, the smaller the overall correction.

Table 3.3: Probability of using a paid preparer if have accountant occupation code

	Corrected AGI percentile					
			Top 1% subgroups			
	$0 - 99$	Top 1%	$99 - 99.5$	$99.5 - 99.9$	99.9+	< 0 AGI
Accountant	-0.201***	-0.127***	-0.189***	-0.135***	-0.0871***	0.0434
	(0.0123)	(0.0101)	(0.0270)	(0.0138)	(0.0166)	(0.0737)
Constant	0.710***	0.915***	0.884***	0.915***	0.946***	0.898***
	(0.00148)	(0.00249)	(0.00574)	(0.00345)	(0.00445)	(0.00727)
N	95,573	14,282	3,409	7,520	3,047	1,748

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

We can then compare how the coefficient on the interaction term between self-prepared, top 1%, accountant compares to the coefficients on using a paid preparer in the top 1%. These results are given in Table 3.5. Across all top 1% groupings (all grouped together, and broken into various subgroups), we cannot reject the null hypothesis that the net effect

[21] We explain how we calculate the scaled corrections in Equation 3.3.1.

Table 3.4: Size of scaled corrections, overall

	Total income		AGI		Taxable income		Tax liability	
	Top 1%	Subgroups	Top 1%	Subgroups	Top 1%	Subgroups	Top 1%	Subgroups
Paid	0.0264***	0.0264***	0.0248***	0.0248***	4.984	4.984	0.0327***	0.0327***
	(0.00141)	(0.00141)	(0.00141)	(0.00141)	(2.924)	(2.924)	(0.00321)	(0.00321)
Paid, top 1%	-0.0799***		-0.0793***		0.136		-0.197***	
	(0.00195)		(0.00195)		(4.039)		(0.00445)	
Self, top 1%, accountant	-0.0573***		-0.0582***		5.119		-0.164***	
	(0.0150)		(0.0150)		(31.12)		(0.0343)	
Paid, < 0 AGI	-0.0684***	-0.0684***	-0.0686***	-0.0686***	0	0	-0.151***	-0.151***
	(0.00514)	(0.00514)	(0.00515)	(0.00515)	(.)	(.)	(0.0117)	(0.0117)
Paid, $99 - 99.5$ Pctl		-0.0606***		-0.0597***		0.161		-0.170***
		(0.00377)		(0.00377)		(7.805)		(0.00859)
Paid, $99.5 - 99.9$ Pctl		-0.0834***		-0.0828***		0.132		-0.203***
		(0.00256)		(0.00256)		(5.303)		(0.00584)
Paid, 99.9+ Pctl		-0.0919***		-0.0915***		0.121		-0.213***
		(0.00384)		(0.00385)		(7.961)		(0.00876)
Self, $99 - 99.5$ Pctl, accountant		-0.0667*		-0.0664*		5.109		-0.172*
		(0.0294)		(0.0294)		(60.93)		(0.0671)
Self, $99.5 - 99.9$ Pctl, accountant		-0.0496*		-0.0509*		5.127		-0.155***
		(0.0199)		(0.0199)		(41.20)		(0.0453)
Self, 99.9+ Pctl, accountant		-0.0688		-0.0699		5.103		-0.180*
		(0.0362)		(0.0362)		(75.00)		(0.0826)
Constant	0.0720***	0.0720***	0.0733***	0.0733***	-5.097*	-5.097*	0.190***	0.190***
	(0.00118)	(0.00118)	(0.00118)	(0.00118)	(2.447)	(2.447)	(0.00269)	(0.00268)
N	111,603	111,603	111,603	111,603	109,855	109,855	111,603	111,603

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

Table 3.5: Size of scaled corrections, overall: testing equality of coefficients

	Top 1%	Top 1% subgroups		
		99 − 99.5	99.5 − 99.9	99.9+
Total income				
Difference	0.00381	0.03250	-0.00736	0.00336
F-stat	0.06391	1.20363	0.13541	0.00850
Prob > F	0.80042	0.27260	0.71289	0.92653
AGI				
Difference	0.00374	0.03152	-0.00697	0.00324
F-stat	0.06130	1.13027	0.12124	0.00791
Prob > F	0.80445	0.28772	0.72769	0.92915
Taxable income				
Difference	0.00143	0.03528	-0.01159	0.00110
F-stat	0.00000	0.00000	0.00000	0.00000
Prob > F	0.99996	0.99954	0.99978	0.99999
Tax liability				
Difference	-0.00088	0.03528	-0.01501	-0.00083
F-stat	0.00065	0.27292	0.10827	0.00010
Prob > F	0.97961	0.60138	0.74213	0.99203

Notes: Difference is equal to $\beta_1+\beta_2-\beta_4$, where β_1 is the coefficient on "Paid," β_2 is the coefficient on the indicator for the relevant top 1% group using a paid preparer, and β_4 is the coefficient on the indicator for the relevant top 1% group that self-prepared their returns and also have an accounting occupation code (see Equation 3.4.4). The reported F-stat and p-value are on the null hypothesis that difference is equal to 0.

on the scaled corrected value of using a paid preparer in the top 1% is equal to the effect of self-preparing as an accountant. This suggests that individuals in the top 1% who are accountants and self-prepare their tax returns experience similarly sized scaled corrections as individuals in the top 1% who use a paid preparer.

Table 3.6 shows the results when we run Equation 3.4.4 where the outcome of interest y is a variety of scaled corrections by type of income (similar to Figure 3.7). The largest scaled corrections are in self-employment income and rental income. Paid-prepared returns among the bottom 99% generally experienced larger corrections to self-employment income relative to self-prepared returns, whereas the top 1% had lower corrections to self-employment income relative to self-prepared returns. Paid-prepared returns in both the bottom 99% and top 1% had lower scaled corrections to rental income compared to self-prepared returns, but the magnitude of the decrease is considerably larger for the top 1%.

Table 3.7 gives the results of statistically testing whether the coefficients on the indicator for a paid-prepared return in the top 1% is equal to that on the indicator for a self-prepared, top 1% return where the individual is an accountant. As with the overall income measures,

Table 3.6: Size of scaled corrections, by type of income

	Wage income		Self-emp. income		Capital gains		Ordinary dividends	
	Top 1%	Subgroups	Top 1%	Subgroups	Top 1%	Subgroups	Top 1%	Subgroups
Paid	-0.000105***	-0.000105***	0.0744***	0.0744***	0.000140	0.000140	-0.000258	-0.000258
	(0.0000165)	(0.0000165)	(0.00689)	(0.00689)	(0.000271)	(0.000271)	(0.000146)	(0.000146)
Paid, top 1%	-0.000315***		-0.255***		-0.00223***		-0.00115***	
	(0.0000228)		(0.00954)		(0.000376)		(0.000202)	
Self, top 1%, accountant	-0.000371*		-0.277***		-0.00147		-0.00163	
	(0.000176)		(0.0735)		(0.00289)		(0.00155)	
Paid, < 0 AGI	-0.000286***	-0.000286***	-0.160***	-0.160***	-0.00419***	-0.00419***	-0.00151**	-0.00151**
	(0.0000603)	(0.0000603)	(0.0252)	(0.0252)	(0.000991)	(0.000990)	(0.000532)	(0.000532)
Paid, $99 - 99.5$ Pctl		-0.000262***		-0.180***		0.00163*		-0.00113**
		(0.0000441)		(0.0184)		(0.000726)		(0.000390)
Paid, $99.5 - 99.9$ Pctl		-0.000319***		-0.269***		-0.00254***		-0.000597*
		(0.0000300)		(0.0125)		(0.000493)		(0.000265)
Paid, 99.9+ Pctl		-0.000361***		-0.300***		-0.00549***		-0.00247***
		(0.0000450)		(0.0188)		(0.000740)		(0.000398)
Self, $99 - 99.5$ Pctl, accountant		-0.000253		-0.290*		-0.00410		-0.00295
		(0.000345)		(0.144)		(0.00566)		(0.00304)
Self, $99.5 - 99.9$ Pctl, accountant		-0.000506*		-0.272**		0.00200		-0.000534
		(0.000233)		(0.0973)		(0.00383)		(0.00206)
Self, 99.9+ Pctl, accountant		-0.000100		-0.272		-0.00901		-0.00325
		(0.000424)		(0.177)		(0.00697)		(0.00375)
Constant	0.000550***	0.000550***	0.276***	0.276***	0.00765***	0.00765***	0.00446***	0.00446***
	(0.0000138)	(0.0000138)	(0.00576)	(0.00576)	(0.000227)	(0.000227)	(0.000122)	(0.000122)
N	111,603	111,603	111,603	111,603	111,603	111,603	111,603	111,603

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

Table 3.6: Size of scaled corrections, by type of income (continued)

	Partnerships/S-corporations		Rental Income		Other income	
	Top 1%	Subgroups	Top 1%	Subgroups	Top 1%	Subgroups
Paid	0.00158	0.00158	-0.0847***	-0.0847***	0.00611***	0.00611***
	(0.00113)	(0.00113)	(0.0143)	(0.0143)	(0.000802)	(0.000802)
Paid, top 1%	-0.00664***		-0.297***		-0.0161***	
	(0.000909)		(0.0116)		(0.00111)	
Self, top 1%, accountant	-0.00422		-0.420***		-0.0131	
	(0.00742)		(0.0943)		(0.00856)	
Paid, < 0 AGI	-0.00818**	-0.00818**	-0.225***	-0.225***	-0.0177***	-0.0177***
	(0.00251)	(0.00251)	(0.0319)	(0.0319)	(0.00293)	(0.00293)
Paid, $99 - 99.5$ Pctl		-0.000520		-0.259***		-0.0107***
		(0.00173)		(0.0220)		(0.00215)
Paid, $99.5 - 99.9$ Pctl		-0.00813***		-0.297***		-0.0164***
		(0.00115)		(0.0146)		(0.00146)
Paid, 99.9+ Pctl		-0.00868***		-0.332***		-0.0212***
		(0.00163)		(0.0208)		(0.00219)
Self, $99 - 99.5$ Pctl, accountant		-0.0110		-0.433*		-0.0210
		(0.0158)		(0.201)		(0.0168)
Self, $99.5 - 99.9$ Pctl, accountant		-0.000522		-0.412***		-0.00632
		(0.00954)		(0.121)		(0.0113)
Self, 99.9+ Pctl, accountant		-0.00833		-0.431*		-0.0239
		(0.0172)		(0.218)		(0.0206)
Constant	0.0112***	0.0112***	0.434***	0.434***	0.0210***	0.0210***
	(0.00102)	(0.00102)	(0.0129)	(0.0129)	(0.000671)	(0.000671)
N	45,463	45,463	45,463	45,463	111,603	111,603

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

we cannot reject the null hypothesis that the size of the net effect of a paid preparer in the top 1% is the same as that for a self-prepared, top 1% accountant.

3.4.3 Testing extension 2: third-party reported income

Another explanation for why individuals in the top 1% of the income distribution would not use a paid tax preparer is that a high proportion of their income stems from sources subject to third-party reporting. As there is less grey area in reporting income subject to third-party reporting, individuals who receive more income from, e.g., wages may be less inclined to spend the money to hire a tax professional to help them prepare their tax returns. We also expect that, among those that do hire a paid-preparer, the amount of money paid for services may be less when there is more income from sources subject to third-party reporting. While this reasoning holds true for any income level, we can investigate both whether we see evidence to support these conjectures and the extent to which that evidence varies by income.

We test these hypotheses with two empirical specifications. In Section 3.4.3.1, we regress the relevant outcomes on the proportion of total income coming from different sources. In Section 3.4.3.2, we regress the relevant outcomes on an indicator for whether or not the individual received 95-100% of their income from wages.

3.4.3.1 Proportion of income from different sources

Equation 3.4.5 includes the proportion of income from seven different income sources on the right-hand side: wage income, self-employment income, capital gains, ordinary dividends, income from partnerships and S-corporations, rental income, and other income. We run Equation 3.4.5 over a variety of sub-populations on two outcomes z: an indicator of whether or not the individual used a paid preparer, and (among those that use paid-preparers), the amount of fees deducted for preparation services. We define P_j as a measure of the amount of total income from type of income j.

$$\begin{aligned} z_i = \gamma_0 + \gamma_1 P_{\text{wages}} + \gamma_2 P_{\text{self-employment}} + \gamma_3 P_{\text{capital gains}} \\ + \gamma_4 P_{\text{ord. dividends}} + \gamma_5 P_{\text{part/s-corps}} + \gamma_6 P_{\text{rental}} + \gamma_7 P_{\text{other}} + e_i \end{aligned} \tag{3.4.5}$$

It is important to acknowledge the potential reverse causality here: hiring a preparer may affect what kinds of income you report, as well as how much of each type, rather than the other way around. We expect this direction to be less problematic for sources of income that are subject to third-party reporting, such as wages. Consequently, and consistently with our theoretical analysis in Section 3.4.1.2, we focus our analysis below on the findings related to

Table 3.7: Size of scaled corrections, by type of income: testing equality of coefficients

	Top 1%	Top 1% subgroups 99 – 99.5	99.5 – 99.9	99.9+
Wage income				
Difference	-0.00005	-0.00011	0.00008	-0.00037
F-stat	0.07746	0.10767	0.12255	0.73604
Prob > F	0.78077	0.74282	0.72629	0.39093
Self-employment income				
Difference	0.09594	0.18439	0.07718	0.04659
F-stat	1.68922	1.61820	0.62170	0.06847
Prob > F	0.19371	0.20335	0.43042	0.79358
Capital gains				
Difference	-0.00062	0.00587	-0.00440	0.00366
F-stat	0.04514	1.05742	1.30586	0.27245
Prob > F	0.83174	0.30381	0.25315	0.60169
Ordinary dividends				
Difference	0.00022	0.00156	-0.00032	0.00052
F-stat	0.02000	0.25765	0.02406	0.01929
Prob > F	0.88755	0.61174	0.87673	0.88955
Partnership/S-corporation income				
Difference	-0.00084	0.01206	-0.00603	0.00123
F-stat	0.01283	0.57737	0.39933	0.00509
Prob > F	0.90981	0.44735	0.52744	0.94314
Rental income				
Difference	0.03862	0.08903	0.03125	0.01497
F-stat	0.16918	0.19482	0.06640	0.00468
Prob > F	0.68084	0.65894	0.79666	0.94548
Other income				
Difference	0.00311	0.01634	-0.00396	0.00887
F-stat	0.13095	0.93767	0.12110	0.18337
Prob > F	0.71745	0.33288	0.72785	0.66849

Notes: Difference is equal to $\beta_1+\beta_2-\beta_4$, where β_1 is the coefficient on "Paid," β_2 is the coefficient on the indicator for the relevant top 1% group using a paid preparer, and β_4 is the coefficient on the indicator for the relevant top 1% group that self-prepared their returns and also have an accounting occupation code (see Equation 3.4.4). The reported F-stat and p-value are on the null hypothesis that difference is equal to 0.

Table 3.8: Probability of using a paid preparer based on proportion of income from...

	Corrected AGI percentile					
			Top 1% subgroups			
	$0-99$	Top 1%	$99-99.5$	$99.5-99.9$	99.9+	<0 AGI
Wages	-0.0357***	-0.105***	-0.159***	-0.0892***	-0.0755**	-0.000367
	(0.00196)	(0.0141)	(0.0269)	(0.0209)	(0.0274)	(0.000619)
Self-emp. income	0.00516*	-0.0182	-0.0397	-0.000197	-0.0209	-0.000471
	(0.00236)	(0.0172)	(0.0320)	(0.0252)	(0.0369)	(0.000894)
Cap. gains	0.0130***	0.00210	-0.0124	0.00204	-0.00725	-0.00198
	(0.00210)	(0.0155)	(0.0322)	(0.0228)	(0.0292)	(0.00158)
Ord. dividends	0.0428***	0.0147	0.0213	-0.00394	0.0164	-0.00358
	(0.00533)	(0.0239)	(0.0487)	(0.0356)	(0.0413)	(0.00329)
Partnerships/S-corp.	-0.0304***	0.00811	-0.0114	0.0138	-0.00265	-0.000357
	(0.00194)	(0.0141)	(0.0268)	(0.0212)	(0.0276)	(0.000897)
Rental income	-0.0324***	0.0365	0.0453	0.0527	-0.0412	0.00113
	(0.00214)	(0.0252)	(0.0523)	(0.0351)	(0.0504)	(0.00229)
Other income	-0.0327***	-0.0290	-0.0248	-0.0434	-0.00949	-0.000314
	(0.00198)	(0.0229)	(0.0498)	(0.0317)	(0.0444)	(0.000891)
Constant	0.714***	0.949***	0.955***	0.940***	0.972***	0.897***
	(0.00210)	(0.0131)	(0.0245)	(0.0196)	(0.0258)	(0.00736)
N	101,259	14,287	3,410	7,523	3,354	1,757

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

wages. The fact that tax preparers might advise their clients to reorganize how they report their income is an important caveat to our discussion of other types of income.

Table 3.8 shows the results when the outcome of interest is whether or not the individual used a paid-preparer, and where P_j is the proportion of total income that comes from source j. Across the income distribution, we find that having a higher proportion of total income coming from wages–a type of third-party reported income–significantly reduces the likelihood of hiring a paid tax preparer. These coefficients are significant at the 0.1% level (except for the 99.9+ percentile groups, which is only significant at the 5% level). This effect is nearly twice as large among the top 1% of the income distribution compared to the bottom 99%. We test the equality of these coefficients in Table 3.9 and find that they are statistically different from each other at the 0.1% level.

We find that a larger proportion of total income from any of the other income types we consider has no statistically significant effect on the likelihood of hiring a preparer among the top 99% (and the magnitude of the estimated coefficients are considerably smaller than what we observe for wages). We do observe statistically significant coefficients for the other types of income among the bottom 99% of the income distribution, though this is driven in part by the fact that we have about seven times as many observations in the bottom 99% as we do in the top 1%. We estimate coefficients on income from partnerships and S-corporations, rental income, and other income, that are all statistically significant at the 0.1% level and comparable in sign and magnitude to the coefficient on wage income for the bottom 99%. In contrast, the coefficients on self-employment income, capital gains, and ordinary dividends

Table 3.9: Probability of using a paid preparer based on proportion of income from wages: testing equality of coefficients

		Top 1% subgroups		
	Top 1%	$99 - 99.5$	$99.5 - 99.9$	99.9+
Proportion of total income from wages				
Difference	-0.080	-0.120	-0.067	-0.044
F-stat	72.803	41.755	26.715	4.799
Prob > F	0.000	0.000	0.000	0.028

Notes: Difference is equal to $\gamma_{1,99\%} - \gamma_{1,1\%}$, where $\gamma_{1,99\%}$ is the value for the bottom 99% and $\gamma_{1,1\%}$ is the relevant value for the top 1% combined or top 1% subgroup as indicated by the table header. The reported F-stat and p-value are on the null hypothesis that difference is equal to 0.

are positive (i.e., higher proportions of total income from those types of income increase the probability of hiring a paid tax preparer), with the coefficient on self-employment income being significant only at the 5% level and an order of magnitude smaller than all of the other coefficients.

Table 3.10 shows the results when the outcome of interest is how much the taxpayer deducted in tax preparer fees, and P_j is the level of (corrected) income type j. We use the level of income divided by \$1,000, so that the interpretation of any coefficient is "the change in total fees paid when type of income j increases by \$1,000." While this does not directly address the question of why someone in the top 1% would not hire a tax preparer (because everyone included in this analysis hired a tax preparer by default), it provides additional insight into how paid-preparer use differs between the top 1% and the bottom 99%, and even within the top 1%.

We find that higher wages are associated with marginally lower fees paid to tax professionals among the bottom 99%, but with higher fees among the top 1% when considered as a whole. This is counter to our intuition that higher wages would be associated with reduced fees paid to professionals (because those returns may be less complex on average). When we consider the top 1% subgroups, we see our intuition holds in the $99 - 99.9$ percent, but again we see a

Table 3.10: Fees paid to professional preparers based on amount of income (in $1k) from...

	Corrected AGI percentile					
			Top 1% subgroups			
	$0-99$	Top 1%	$99-99.5$	$99.5-99.9$	99.9+	< 0 AGI
Wages	-1.76*	4.34***	-27.03***	-13.51***	3.03*	-31.72
	(0.73)	(0.66)	(6.53)	(2.38)	(1.23)	(28.75)
Self-emp. income	8.43***	-5.33***	-27.84***	-11.68**	-5.54*	-125.28***
	(1.34)	(1.50)	(8.29)	(4.10)	(2.68)	(5.99)
Cap. gains	-5.26***	2.77***	-27.10***	4.26	2.55***	98.08***
	(1.36)	(0.16)	(7.71)	(2.81)	(0.28)	(3.64)
Ord. dividends	198.61***	72.30***	143.05***	121.32***	67.67***	206.68***
	(2.20)	(1.84)	(10.73)	(5.31)	(3.31)	(30.09)
Partnerships/S-corp.	-25.17***	3.42***	-42.42***	-15.41***	2.71**	0.95
	(0.91)	(0.50)	(5.96)	(2.49)	(0.91)	(2.98)
Rental income	-0.89	-35.32***	-8.86	-9.16	-35.18***	-29.04
	(4.56)	(4.01)	(25.67)	(10.26)	(7.14)	(36.96)
Other income	-43.39***	-0.51	-12.72	16.91*	-1.34	1.20
	(2.08)	(1.48)	(14.43)	(6.98)	(2.62)	(1.38)
Constant	1,309***	21,873***	16,671***	24,484***	50,357***	-2,181
	(106)	(1,705)	(2,992)	(2,711)	(5,943)	(6,107)
N	17,741	7,676	1,474	3,997	2,205	336

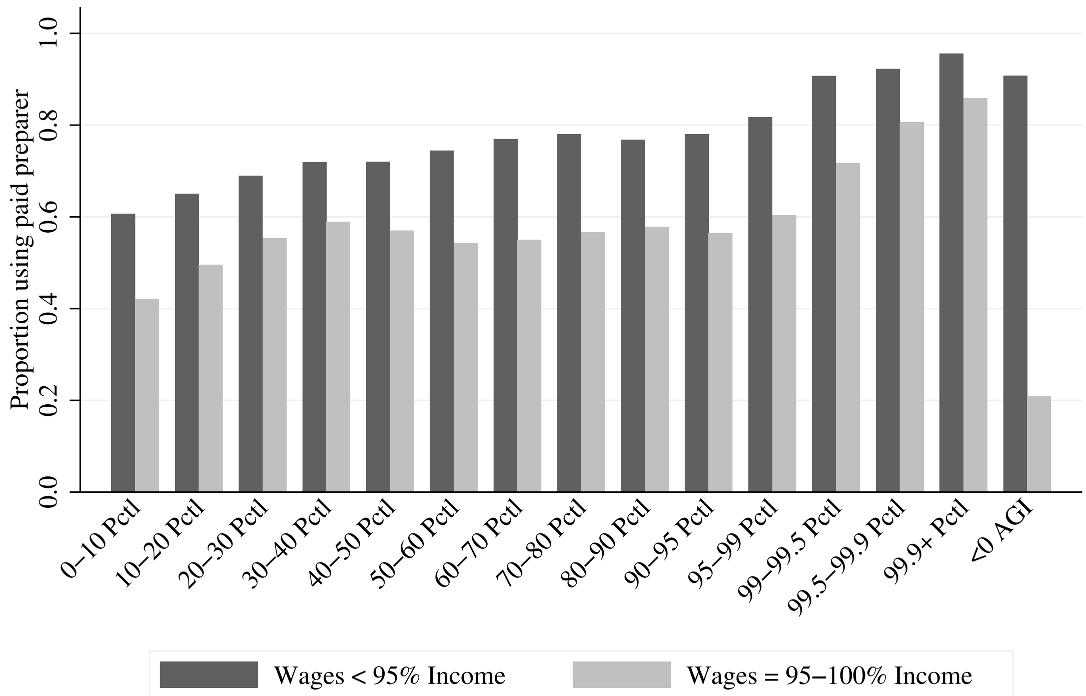
Notes: Standard errors in parentheses. Limited to returns with a paid preparer with greater than 0 estimated fees paid for professional tax services.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

positive relationship between wages and fees paid to tax professionals is positive.

Perhaps most telling is how small the magnitude of these coefficients are. Across income types and the income distribution (with the exception of the negative AGI group), the estimated γ coefficients are almost all lower than $40 and often lower than $10 in magnitude. These magnitudes are particularly small relative to the constants associated with the various income groups (e.g., the constant for the top 1% as a whole is over $21,500).

The primary exceptions to this are the coefficients on ordinary dividends. Notably, the coefficient on ordinary dividends for the bottom 99% suggests that a $1,000 increase in ordinary dividends is associated with a nearly $200 increase in fees paid to tax professionals–equivalent to a 15% increase over the constant.

Figure 3.13: 95-100% of total income from wages by preparation type



Notes: N = 117,303 individuals. See Table 3.2 for subtotals by corrected AGI percentile.

3.4.3.2 95-100% of total income from wages

In this section, we consider cases where an individual's wages are equal to 95-100% of the value of the individual's total pre-tax income.[22] We present the results graphically in Figure 3.13. Individuals with less than 95% of their income from wages are more likely to use paid tax preparation services across the income distribution. This difference generally decreases for higher income groups, reflecting the fact that higher income individuals are more likely to use paid preparers.

We capture this relationship by running Equation 3.4.6 over various income subgroups.

$$z_i = \zeta_0 + \zeta_1 \mathbb{I}\left[\frac{\text{wage income}}{\text{total income}} \in [0.95, 1.0]\right] + u_i \tag{3.4.6}$$

The results of this analysis for z_i being an indicator of using a paid tax preparer are given in Table 3.11. The results are consistent with our previous analyses: the baseline probability of using a paid preparer goes up with income group. Across all income groups, the probability of using a paid preparer goes down if 95-100% of total income is from wage income, but the

[22]Because of cases with negative income (e.g., business losses), this indicator may capture some individuals whose wage income is less than 95% of their total positive income.

Table 3.11: Probability of using a paid preparer based on indicator of wages being 95-100% of income

	Corrected AGI percentile					
			Top 1% subgroups			
	$0-99$	Top 1%	$99-99.5$	$99.5-99.9$	99.9+	<0 AGI
Wages = 95-100% of Income	-0.201***	-0.133***	-0.190***	-0.116***	-0.0973***	-0.699***
	(0.00305)	(0.00675)	(0.0148)	(0.00957)	(0.0114)	(0.0599)
Constant	0.757***	0.927***	0.907***	0.923***	0.956***	0.908***
	(0.00171)	(0.00259)	(0.00605)	(0.00359)	(0.00430)	(0.00700)
N	101,259	14,287	3,410	7,523	3,354	1,757

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

relative decrease in probability is smaller for higher income groups.

We perform the same analysis considering average fees paid to tax professionals. As in Section 3.4.3.1, setting z_i equal to fees paid to tax preparers does not itself answer the question of why someone in the top 1% of the income distribution would not hire a professional tax preparer, but it provides additional insight into how tax preparers are used by that population.

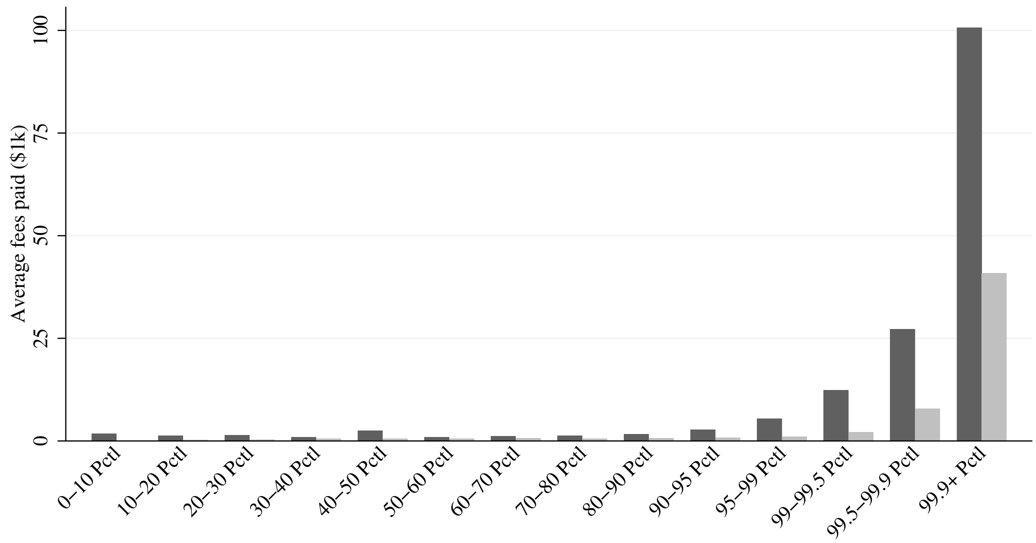
We first present the results graphically in Figure 3.14. Figure 3.14a shows the average fees paid to professional preparers by income group, split by whether or not the individuals had 95-100% of their total income come from wage income. Figure 3.14b is the same analysis, limited to fees that under $50,000. Both panels highlight that individuals who have less than 95% of their total income come from wages pay considerably more, on average, for professional tax preparation than individuals who receive nearly all of their income from wages.

The results from running Equation 3.4.6 with z_i as fees paid are given in Table 3.12. The average fees paid increase considerably when comparing the bottom 99% to the top 1%. In addition, the average decrease in fees paid is larger the higher up in the income distribution we go (in particular, the average decrease in fees paid to tax professionals when the individual receives most of their income from wages increases with each subsequent subgroup of the top 1%).

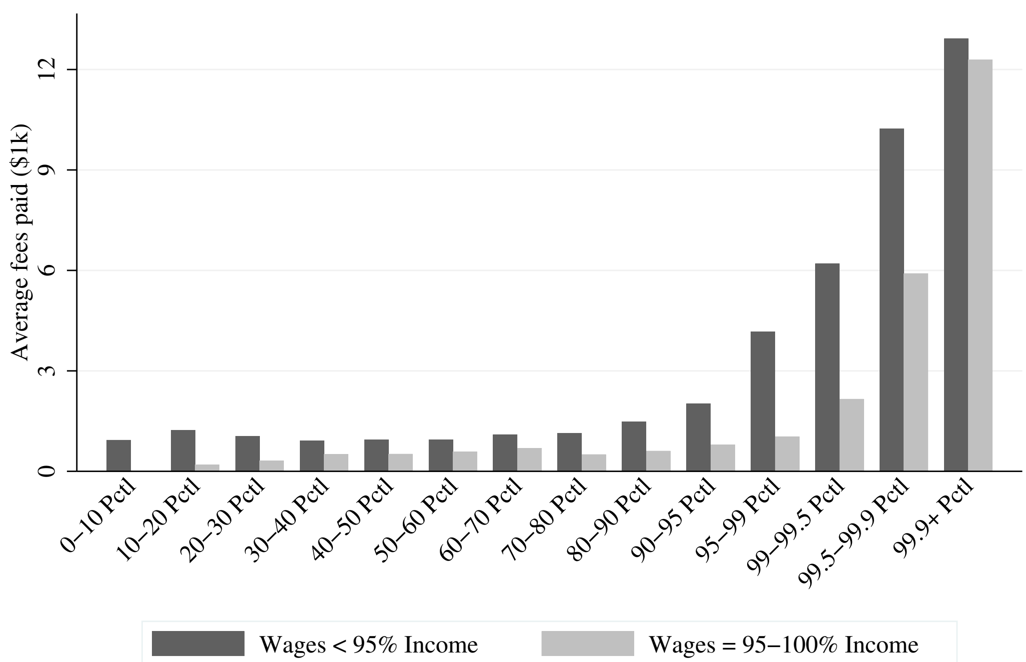
These results support the idea that having income from third-party reported sources both decreases the likelihood of using a preparer to begin with, and decreases the fees that a preparer is able to elicit for their services. These analyses are also consistent with the literature and idea that simpler returns are less likely to be prepared by a paid professional.

Figure 3.14: Tax preparation fees by indicator of wages being 95-100% of income

(a) Average tax preparation fees



(b) Average fees, limited to fees < $50k



Notes: N = 25,753 individuals. See Table 3.2 for subtotals by corrected AGI percentile.

Table 3.12: Fees paid (in $1k) to professional preparers based on indicator of wages being 95-100% of income

	Corrected AGI percentile				
			Top 1% subgroups		
	$0-99$	Top 1%	$99-99.5$	$99.5-99.9$	99.9+
Wages = 95-100% of Income	-2.333***	-29.36***	-10.14***	-19.37***	-59.82***
	(0.201)	(4.983)	(2.703)	(2.714)	(16.14)
Constant	3.080***	45.52***	12.30***	27.25***	100.7***
	(0.102)	(1.898)	(1.063)	(1.022)	(6.138)
N	17,741	7,676	1,474	3,997	2,205

Notes: Standard errors in parentheses. Limited to returns with a paid preparer with greater than 0 estimated fees paid for professional tax services.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3.5 Conclusion

This paper explores the use of professional tax preparers by the top 1% of the income distribution. Given that recent literature has suggested that the rich might be responsible for a large share of unpaid taxes, it seems important to understand the role tax preparers play in shielding their income from taxation, and what regulation might address this state of affairs. Our analysis provides crucial descriptive details, which will be key in understanding this issue in future research.

We find that trends in tax preparer usage substantially differ across the income distribution, both among individuals who use paid preparation services and the relative comparison with self-prepared returns. In particular, we find that tax preparer usage is correlated with different types of income for the top 1%, such as income from some pass-through businesses. Additionally, while tax preparer usage is associated with higher corrections upon audit in the bottom 99% of the income distribution, it is associated with smaller corrections among the top 1%.

One limitation of these results is that post-NRP audit corrections may still not be completely accurate. The auditor may fail to detect under- (or over-) reported income, or may "correct" a line item that was originally accurate. The extent the top 1% are more likely to have undetected, under-reported income, and/or more complicated tax returns, impacts our conclusions about the role of tax preparation services in this population. Future research on this topic should aim to integrate operational audit data and data on offshore accounts, to further explore these issues. Another important and not-well understood piece of the puzzle is the role of tax-preparation software, which has increased in prevalence over the last two decades among the bottom 99% and parts of the top 1% of the income distribution.

APPENDICES

APPENDIX A

Appendix to Chapter 1

A.1 Additional institutional information about IRAs

A.1.1 Intuition for IRA tax benefit

Consider an individual that wants to invest an amount a of taxable income. She faces an average income tax rate of t. She can either invest the amount in a regular savings account after remitting income tax, or in a tax-benefited savings account before remitting income tax. Both accounts face the same annual rate of return r. The growth of these accounts is outlined in Table A.1:

Table A.1: Account growth over time in simple example

			Regular account		
Period	Tax-benefited account		No tax on return		With tax on return
0	a		$a(1-t_i)$		$a(1-t_i)$
1	$a(1+r)$		$a(1-t_i)(1+r)$		$a(1-t_i)(1+r-rt_r)$
2	$a(1+r)^2$		$a(1-t_i)(1+r)^2$		$a(1-t_i)(1+r-rt_r)^2$
3	$a(1+r)^3$		$a(1-t_i)(1+r)^3$		$a(1-t_i)(1+r-rt_r)^3$
...					
Account balance in period before withdrawing					
$k-1$	$a(1+r)^{k-1}$		$a(1-t_i)(1+r)^{k-1}$		$a(1-t_i)(1+r-rt_r)^{k-1}$
Amount after withdrawal					
k	$a(1-t_i)(1+r)^k$	$=$	$a(1-t_i)(1+r)^k$	$>$	$a(1-t_i)(1+r-rt_r)^k$

If the individual withdraws from the tax-benefited account, she will owe income tax and her take-home value will be equal to the balance in the regular account. However, this amount is not taxed further. In contrast, there are two additional taxes the individual faces if she holds her money in the regular account depending on how the money is invested. She may owe tax on the interest accumulated each year she held the account. Upon withdrawal, she may owe capital gains taxes. Neither of these apply to the tax-benefited account. As a result, it is as if the returns to the tax-benefited account were not subject to additional tax beyond income tax.

A.1.2 Roth IRAs

The primary difference between traditional IRAs and Roth IRAs is the timing of when income tax is due. Contributions to Roth IRAs are made after income tax is remitted, with no tax due on qualified distributions (including on any earnings between contribution and withdrawal). Roth IRAs are not subject to Required Minimum Distributions (withdrawals are not required from Roth IRAs until after the death of the account owner). Roth IRA holders do face the early withdrawal penalty, but the penalty only applies on withdrawals that are larger than the principal investment.

There are a few other differences between traditional IRAs and Roth IRAs. All taxpayers who earn sufficient income to make a contribution to a traditional IRA are eligible to hold a traditional IRA, whereas eligibility to contribute to a Roth IRA is phased out for taxpayers with high enough income (see Appendix A.1.4). There are also differences in when and what taxes are owed upon death of the account holder (see Appendix A.1.6).

A.1.3 SEP and SIMPLE IRAs

There are two types of IRAs in addition to traditional IRAs and Roth IRAs: SIMPLE IRAs and SEP IRAs. SIMPLE IRAs allow both employee and employer to contribute to a traditional IRA set up for the employee, and are intended to be used by small employers (<100 employees) not sponsoring a retirement plan such as a 401(k). SEP (Simplified Employee Pension) IRAs are used primarily by self-employed and small business owners. Only the employer contributes, and the employee is always 100% vested. If an employer offers SEP IRAs to their employees, the employer must contribute for all employees that meet a set of requirements.[1]

Table A.2 shows the number of taxpayers that made contributions to or took distributions from each of these account types, as well as the total end-of-year fair market value of these accounts, in 2015. Because these IRAs make up a small fraction of total IRA accounts and

[1] A third additional type of IRA–the SARSEP (Savings Incentive Match Plan for Employees) IRA–are SEP IRAs established before 1997 that include a salary reduction arrangement.

face the same penalty structure as traditional IRA accounts, we follow Mortenson et al. (2019) and count contributions toward (and distributions from) a SEP or SIMPLE IRA as contributions toward (and distributions from) a traditional IRA for our analysis.

Table A.2: IRA activity by type of account, 2015

	Contributions		Distributions		End-of-year FMV	
IRA type	Taxpayers	Pct.	Taxpayers	Pct.	FMV $1,000	Pct.
Traditional	4,305,106	31.6%	17,360,396	92.4%	6,386,720	85.4%
Roth	6,363,335	46.7%	708,221	3.8%	625,077	8.4%
SEP	1,093,512	8.0%	670,990	3.6%	364,264	4.9%
SIMPLE	1,865,777	13.7%	40,459	0.2%	101,194	1.4%
Total	13,627,730	100.0%	18,780,066	100.0%	7,477,255	100.0%
Total (unique)	13,006,314	n/a	18,670,599	n/a	n/a	n/a

Source: Authors' calculations from Statistics of Incomes Tax Stats: Accumulation and Distribution of Individual Retirement Arrangements (IRA) Table 1 (2015), available at https://www.irs.gov/pub/irs-soi/15in01ira.xls. Total (unique) count is less than the total because some taxpayers may have made contributions or distributions from more than one account. Contributions do not include rollovers.

A.1.4 Contribution and deductibility limits

Contributions into IRAs are capped. In 2015, individual taxpayers could not contribute more than $5,500 in total to traditional and Roth IRA accounts. That amount was higher for individuals over age 50 ($6,500). Contributions to IRAs are limited to an individual's taxable compensation, so if that was less than $5,500 for an individual, that individual's contribution limit would be equal to their taxable compensation. Taxpayers were not allowed to make contributions to a traditional IRA after age $70\frac{1}{2}$ in 2015; contributions to a Roth IRA were allowed after age $70\frac{1}{2}$. In the event a taxpayer makes a contribution above the limit (or any contribution at age $70\frac{1}{2}$ or older), the excess amount is taxed at 6% for each year it remains in the IRA.

When money in one retirement account is moved into a different retirement account, it is called a "rollover." For example, if you leave job A, you can transfer the amount in your 401(k) from job A into the 401(k) you set up with your new job (or into your own IRA) and it would not count toward your contribution limit. Rollovers from, e.g., a 401(k) into an IRA are not counted toward the taxpayer's contribution limit.

Contributions to Roth IRAs are not allowed for taxpayers above certain income levels. Individual taxpayers with a modified adjusted gross income (MAGI) of less than $116,000 in 2015 were allowed to contribute the full $5,000 to a Roth IRA. Between a modified AGI of $116,000 and $131,000, taxpayers were eligible to contribute a reduced amount of their contribution limit to a Roth IRA. The partial amount for an individual taxpayer in 2015 can be calculated as:

$$\text{partial contribution} = \text{contribution limit} - \left(\frac{\text{MAGI} - \$116,000}{\$10,000} \cdot \text{contribution limit}\right).$$

After a modified AGI of \$131,000, taxpayers wishing to contribute to an IRA were forced to contribute to a traditional IRA. For married taxpayers filing jointly, the upper bound of modified AGI for contributing the full amount to a Roth IRA was \$183,000, and the lower bound after which IRA contributions must be made to a traditional account was \$193,000. One strategy expected to be used by some taxpayers who earn too much to contribute to a Roth IRA is to contribute to a traditional IRA and convert it to a Roth IRA later. No taxes are assessed with a rollover, but if you rollover from a traditional account into a Roth account, you must include the distributed amount as income (i.e., it will be included in your taxable income that year).

The deductibility of contributions to a traditional IRA are limited for taxpayers who are also covered by an employer-sponsored plan, such as 401(k). In 2015, only individual taxpayers with a modified AGI of less than \$61,000 (\$98,000 for taxpayers filing jointly) were able to deduct the full contribution limit of \$5,000 if the entire amount was contributed to a traditional IRA. The deductible amount phased out until a modified AGI of \$71,000 (\$188,000 for taxpayers filing jointly), at which point individual taxpayers covered by an employer-sponsored plan were not eligible to deduct any amount of a contribution to a traditional IRA.

A.1.5 Schedule for Required Minimum Distributions

After age $70\frac{1}{2}$, holders of traditional, SEP, and SIMPLE IRAs are required to take "minimum distributions" from their account. The first RMD must be taken by April 1 of the year after the calendar year in which the account holder turns $70\frac{1}{2}$; after that, RMDs are due by December 31 of each calendar year. This means that account holders who wait until the calendar year after the year in which they turn $70\frac{1}{2}$ to remit their first RMD will owe two RMD payments that year: one by April 1 and one by December 31.

The RMD amount for a given year is equal to the account balance on December 31 of the preceding year divided by a value known as a "distribution period." These values are given by the IRS's "Uniform Lifetime Table," given below in Table A.3.[2] The estimated RMD for

[2] The distribution schedule is different if the account owner's spouse is the sole beneficiary of the account and the spouse is more than 10 years younger than the account holder. In this case, the distribution period value is determined by the account holder's age and the spouse's age on their respective birthdays in that calendar year.

an IRA is reported on IRS Form 5498. If a taxpayer holds more than one IRA subject to RMDs, their total RMD is equal to the sum of the RMDs from each account. In this scenario, the taxpayer may take the total RMD from one account.[3]

Table A.3: Uniform Lifetime Table from IRA Publication 590

Age	Distribution period	Age	Distribution period	Age	Distribution period	Age	Distribution period
70	27.4	82	17.1	94	9.1	106	4.2
71	26.5	83	16.3	95	8.6	107	3.9
72	25.6	84	15.5	96	8.1	108	3.7
73	24.7	85	14.8	97	7.6	109	3.4
74	23.8	86	14.1	98	7.1	110	3.1
75	22.9	87	13.4	99	6.7	111	2.9
76	22.0	88	12.7	100	6.3	112	2.6
77	21.2	89	12.0	101	5.9	113	2.4
78	20.3	90	11.4	102	5.5	114	2.1
79	19.5	91	10.8	103	5.2	115+	1.9
80	18.7	92	10.2	104	4.9		
81	17.9	93	9.6	105	4.5		

Source: IRA Required Minimum Distribution Worksheet, https://www.irs.gov/pub/irs-tege/uniform_rmd_wksht.pdf. This schedule has been in place since 2002. The schedule from 1999-2001 required slightly higher withdrawals than the schedule presented here.

A.1.6 Penalties and the death of the account holder

What happens to a traditional IRA after the account owner dies depends on who inherits the account and the age of the original account owner upon death. If a spouse inherits the account, the spouse is eligible to designate herself as the IRA account owner or keep the account in the decedent's name. In the former scenario, the surviving spouse is allowed to make contributions, take distributions, rollover the assets into a different account, and even name a new beneficiary for the account. The surviving spouse is also subject to RMD rules based on their own age. If the surviving spouse keeps the account in the decedent's name, he or she must start taking distributions when the descendent would have turned $70\frac{1}{2}$ (or continue taking distributions if the decedent had already started taking RMDs), but based on his or her own life expectancy.

Non-spouse beneficiaries of traditional IRAs do not owe taxes on the assets in the account until they begin taking distributions. If the original account owner died after RMDs began, the beneficiary must continue to take distributions. The amount of the distribution is based off of the longer of the original account owner's and the beneficiary's life calculated expectancy.

[3]The same is not true for 401(k)s: RMDs from 401(k)s must be taken from each account separately.

If the owner died before they were subject to RMDs, the beneficiary must either take RMDs based off of their own life expectancy starting the year after the original account owner's death, or elect to fully distribute the account under the "5-year rule."[4] The 5-year rule requires the IRA beneficiaries to withdraw all of the assets of the IRA by December 31 of the calendar year containing the fifth anniversary of the original account owner's death, but does not require any distributions before that time. Failing to fully distribute the assets after electing to distribute according to the 5-year rule results in a 50% excise tax on the remaining amount. The 5-year rule always applies in cases where the beneficiary isn't an individual (i.e., is a trust or an organization).

When the holder of a Roth IRA dies and the beneficiary is the spouse, the spouse becomes the new owner of the account. The spouse may choose not to take ownership of the account, in which case he or she faces the same options as a non-spouse beneficiary. A non-spouse beneficiary may either distribute the entirety of the Roth IRA by the end of the calendar year containing the fifth anniversary of the account holder's death, or over his or her own life expectancy. A surviving spouse is allowed to delay distributions until the deceased spouse would have turned $70\frac{1}{2}$. The ability of a non-spouse beneficiary to take distributions based on their life expectancy was curtailed by the SECURE Act, which was signed into law in December 2019. The SECURE Act includes a provision that, going forward, requires non-spouse beneficiaries to distribute the funds of inherited IRAs within 10 years. This is meant to prevent "Stretch IRAs," in which IRA accounts are passed down, e.g. to grandchildren, who are then able to take distributions over their expected lifespan and enjoy tax-free growth along the way.

A.2 Data description

This appendix serves as a guide to the data and terminology used in this project. Appendix A.2.1 defines relevant terms and, when applicable, provides additional information about how key variables were constructed. Appendix A.2.2 discusses our sample construction and the IRS forms from which we extracted the data.

A.2.1 Glossary of terms and variable construction

Contribution We get contribution information from Form 5498. We assume that anything labeled as a contribution that was larger than 110% of the taxpayer's contribution limit in

[4]If there are multiple non-spouse beneficiaries, the required distribution amounts are based on the age of the oldest beneficiary. If the beneficiaries split the original account into separate accounts, each beneficiary will be subject to RMD rules based on their own age.

that year was a data error. In this case, we coded the contribution as a rollover.

Defined-benefit (DB) plan Employer-sponsored savings plan that pays pre-determined benefits from a collective trust fund rather than an individualized account. Commonly referred to as pensions.

Defined-contribution (DC) plan Employer-sponsored savings plan with an individual account for each participant. When an individual retires, the designated account is used to provide benefits to that individual.

Distribution The word used by the IRS and financial industry to talk about withdrawing money from a tax-deferred retirement plan. We calculate distributions from IRA accounts using data from Form 1099-R. We pull all Form 1099-Rs associated with our random sample. We follow a standard algorithm for determining which Form 1099-Rs indicated distributions from traditional, Roth, and SIMPLE IRAs, and which Form 1099-Rs indicated rollovers from Roth accounts into traditional accounts. We exclude distributions from accounts other than IRAs, as well as the following types of IRA distributions:

- IRA distributions due to death of the account holder
- IRA distributions due to disability
- Recharacterizations of this or prior year's IRA distribution
- Corrective distributions due to IRA contributions or deferrals above the annual limit

These exclusions eliminate 5.7% from our sample of IRA-related Form 1099-Rs, the majority of which come from distributions due to death of the account owner (4.8%).

For each taxpayer, we determine the taxable distributions and gross distributions for each type of IRA, as well as gross distributions excluding any distributions not subject to the early withdrawal penalty. Out of 22,192,136 Form 1099-Rs associated with our sample, there are 388 cases where the taxable amount is greater than the gross amount. In these cases, we assume the taxable amount is correct and replace the gross amount with the taxable amount, unless the difference between the two is greater than $1,000$ and the value of the gross amount is greater than 500. When this occurs, we add the value of the taxable amount to the gross amount.

Early withdrawal penalty The penalty owed to the IRS when the owner of an individual retirement account takes a distribution before the age of $59\frac{1}{2}$. The penalty is equal to 10% of the amount withdrawn. Taxpayers may receive an exemption if they take an early withdrawal from an IRA to pay for qualified medical or higher education expenses. First-time homebuyers may take a distribution of up to $10,000 from an IRA without penalty.

Employer-sponsored plan (ESP) Can be either defined-contribution (DC) or defined-benefit (DB). Figure A.1 presents evidence that DC employer-sponsored accounts such as 401(k)s are increasingly popular over DB accounts such as pensions. In each of the panels, DB plans are shown by the solid line and DC plans by the dashed line. Figure A.1a shows the total number of plans offered (in thousands), Figure A.1b shows the total number of active participants (in millions), and A.1c shows the total assets in these plans (in trillions). For all three measures, it's clear that DC plans are growing faster than DB plans (and, in some cases, that DB plans are shrinking).

Figure A.1: Trends in employer-sponsored DB vs. DC plans

(a) Number of plans (b) Active participants (c) Total assets

Source: Tables E1, E7, and E10 of *Private Pension Plan Bulletin Historical Tables and Graphs 1975-2017*, respectively, published by the U.S. Department of Labor. Defined-contribution plans are shown by the dashed line; defined-benefit plans are shown by the solid line. In 2004, the definition of who was participating was revised.

Excess accumulation penalty The penalty owed to the IRS when the owner or beneficiary of an individual retirement account fails to take a required minimum distribution. The penalty is equal to 50% of the required amount not withdrawn. The penalty may be waived when the excess accumulation was the result of a reasonable error and the account holder took steps to correct the insufficient distribution.

Fair-market value (FMV) The value of an individual retirement account if all of the assets were to be sold as reported on IRS Form 5498. For our purposes, equivalent to the account balance.

Individual Retirement Account (IRA) Also known as an Individual Retirement Arrangement. IRAs are tax-benefited retirement savings accounts. We define anyone in our sample that we observe making a contribution to an IRA (including rollovers), taking a distribution from an IRA, or having a positive IRA balance as an "IRA account holder."

Lump-sum distribution The distribution of the entire balance of a taxpayer's ESP(s) from a single employer.

Qualified charitable distribution (QCD) Direct transfer from an IRA account to a qualified charity. These distributions can be counted toward RMDs. Conditions: must be $70\frac{1}{2}$ or older at the date of distribution, total amount contributed not to exceed the amount that would otherwise be taxed as ordinary income up to the maximum amount ($100,000), which applies to the sum of all QCDs made. A distribution made to an account holder that is then donated to charity does not count as a QCD.

Required Minimum Distribution (RMD) Starting in the tax year that an account holder turns $70\frac{1}{2}$, holders of traditional 401(k) and IRA accounts are required to make a minimum distribution every year. Late RMD payments are taxed at 50%. Amount is based on account balance on Dec. 31 of previous tax year and life expectancy tables. First payment is due by April 1 of the year after you turn $70\frac{1}{2}$; subsequent payments must be made by Dec. 31 of each calendar year. As a result, taxpayers may have two RMDs included in their taxable income in the first calendar year they make RMD payments. In December 2019, the Setting Every Community Up for Retirement Enhancement (SECURE) Act was signed into law. Section 114 increases the age at which taxpayers are subject to excess accumulation penalties for failing to take required distributions from $70\frac{1}{2}$ to 72. Because we consider a time period before 2019, we use $70\frac{1}{2}$ as the age threshold throughout the paper.

Rollover The word used when money in one retirement account is moved into a different retirement account. For example, if you leave job A, you can transfer the amount in your 401(k) from job A into the 401(k) you set up with your new job (or into your own IRA). No taxes are assessed with a rollover unless it is a Roth conversion. Rollovers are not counted toward contribution limits.

When a taxpayer rolls over funds from, e.g., a 401(k) or another IRA into an IRA, a Form 1099-R will be issued to mark the distribution. If the rollover is a Roth conversion, the distributions will be marked as a rollover on Form 1099-R and a Roth conversion on Form 5498. The Form 5498 is issued to mark the contribution to the IRA.

Roth conversion A type of rollover when funds in a traditional IRA are rolled over into a Roth IRA. For this type of rollover, income tax is due on the distribution (i.e., you must include the distributed amount as income (i.e., it will be included in your taxable income that year).

Roth IRA A type of IRA in which income tax is remitted before funds are moved into the account. Income taxes are not due on qualified distributions, including on returns to the principal.

Savings (initial distribution of non-tax-benefited) IRS Form 1099-INT is used to report interest income to the IRS. We add together the values of Box 1 ("Interest income") and Box 3 ("Interest on U.S. Savings Bonds and Treasury obligations"), both winsorized at the 1^{st} and 99^{th} percentiles, as reported on IRS Form 1099-INT for the individuals in our sample at age 40. We estimate assets in regular savings by dividing that amount by 0.01. The average estimated value of regular savings using this method is $57,539. This is well within the ballpark of the average holdings in transaction accounts, Certificates of Deposit, and savings bonds estimated by the Federal Reserve, which in 2016 was estimated to be $64,800 for families with a head of household between 35 and 45. Source: authors' calculations using the U.S. Federal Reserve Board, Historic Tables and Charts - Tables Based on Internal Data - Estimates Inflation Adjusted to 2016 Dollars, Table 6, available at https://www.federalreserve.gov/econres/scfindex.htm.

Total distribution When all of the contents of an individual retirement account are withdrawn.

Traditional IRA A type of IRA in which funds are contributed before income tax is remitted. Income tax is then due on all qualified distributions, including on returns to the principal.

Withdrawal See distribution.

A.2.2 Sample construction

We take advantage of administrative tax data in this project. Our initial panel is a 5% random sample of individuals aged 18 or older in 1999. We follow these taxpayers through 2015. The 17-year panel is balanced apart from exit due to death and emigration.

Our initial sample includes 14,606,095 individual taxpayers. We make several cuts to this initial sample. We exclude taxpayers who died before 2000 and therefore are never alive during our sample period. We also exclude taxpayers who were older than 90 (born before

1909) in 1999. We exclude any taxpayer whose taxpayer identification number (TIN) was ever associated with an invalid TIN flag or multiple Form 1040 filings.[5] These last two exclusions in particular are to deal with issues of identity theft. Table A.4 summarizes the impact these exclusions have to our sample size. Our final sample includes 11,950,600 individual taxpayers, 32.7% of which are IRA account holders. We identify taxpayers as "IRA account holders" if we ever observe them making a contribution to an IRA (including a rollover), taking a distribution from an IRA, or having an outstanding IRA balance reported on Form 5498.

Table A.4: Data exclusions

Exclusion	Unique taxpayers	Percent of initial sample
Original sample	14,606,095	100.0%
Died before 2000	456,240	3.1%
Born before 1909	1,072,032	7.3%
Invalid TIN	1,126,383	7.7%
Multiple F-1040s	840	0.0%
Final sample	11,950,600	81.8%
IRA holders	3,913,401	
		Percent of IRA sample
Traditional IRA holders	2,790,313	72.2%
Roth IRA holders	1,071,964	27.8%
Both types	616,020	15.7%

We focus on IRAs because we can cleanly identify IRA contributions and distributions in the administrative tax data. Form 5498 is intended for contributions to IRAs only, and separately identifies contributions to traditional and Roth IRAs. Form 5498 is filed by the individual or organization who made contributions on behalf of the recipient. For example, the bank that manages Joe's IRA would submit Form 5498 for Joe if Joe made a contribution to his IRA. In addition to contribution information, Form 5498 includes the end-of-year fair-market values (FMV) of an IRA. Financial institutions are required to submit Form 5498 each year for an existing IRA even if no contribution is made. There is some debate about whether or not financial institutions actually do this, but our data suggest that most financial institutions do submit Form 5498 even if the account holder did not make a contribution that year.

Line 7 of Form 1099-R includes a checkbox to indicate that the distribution comes from an IRA. We use the codes entered in Line 7 to then distinguish between distributions from traditional and Roth IRAs. Notably, we can separately identify normal distributions, early

[5]We understand that this necessarily excludes individuals choosing to file Married Filing Separately.

distributions without a known exception, and early distributions with a known exception. We are also able to identify rollovers to traditional and Roth IRAs. In contrast, neither Form 1099-R nor Form 1040 separately identify distributions from pensions and annuities from distributions from 401(k)s and other employer-sponsored defined-contribution accounts. Form 1099-R is filed by the institution who issued the distribution. If Joe took a distribution from his IRA, the bank that manages the IRA would be responsible for submitting Form 1099-R to the IRS.

IRA distributions are also reported on Lines 15a and 15b of Form 1040. We prefer to use the amounts from the third-party reported Form 1099-R. Using Form 1099-R has two primary advantages. First, we are able to capture distributions from non-filers. Second, Lines 15a and 15b on Form 1040 aggregates distribution amounts for spouses for married taxpayers filing jointly. We observe that, for single taxpayers, the total amount distributed from traditional IRAs as reported on Form 1099-R and the taxable amount reported on Line 15b of Form 1040 are within 1% of each other in 78% of cases, and within \$100 of each other in 94% of cases. This is in line with what is reported in Brady and Bass (2020).

We also use data from individual income tax returns. In particular, we use data from Forms 1040 and 5329. Form 1040 includes information such as total wages and adjusted gross income. Schedules A and C for Form 1040 include charitable giving (for itemizers) and self-employment income, respectively. Form 5329 is used to declare additional taxes owed due to early withdrawal, non-qualified distributions, or failing to receive a minimum required distribution. Finally, we use data from Form 1099-INT to estimate the initial distribution of non-IRA savings in the model.

A.3 Additional evidence of bunching

A.3.1 Diagnostic tests

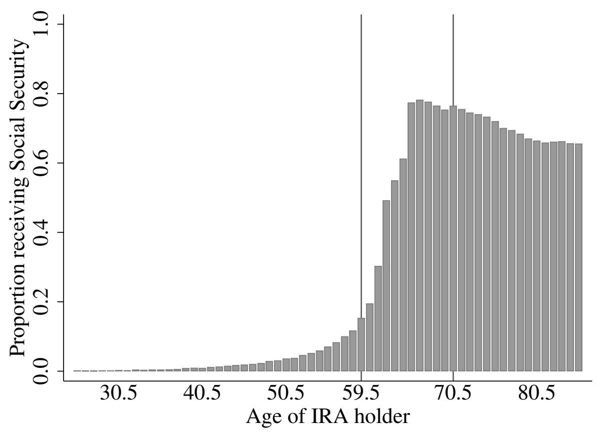
We conduct two diagnostic tests to confirm that our observed bunching is a result of the age thresholds for penalties related to IRA withdrawals. In Appendix A.3.1.1, we plot the proportion of taxpayers receiving two placebo sources of income (wages and Social Security) at each age in 2005. In Appendix A.3.1.2, we separately plot the proportion of taxpayers taking a distribution from an IRA for taxpayers whose half-birthday is in the first or second half of the year. The results of both of these diagnostic tests suggest that taxpayers are changing their behavior in response to the penalties at ages $59\frac{1}{2}$ and $70\frac{1}{2}$ (i.e., the bunching we presented in Section 1.4 was not a result of other policies related to timing of income).

A.3.1.1 Placebo outcomes

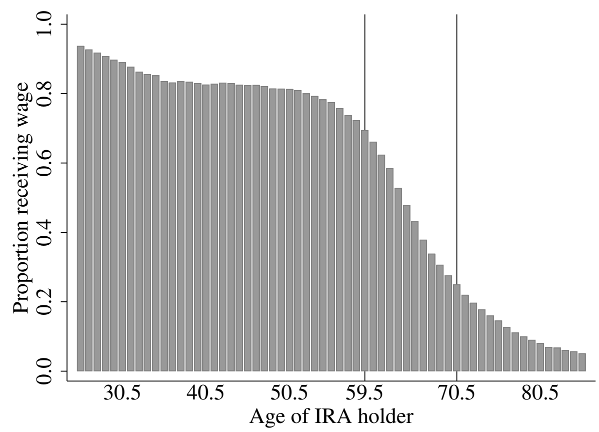
We create the same figure as Figure 1.3a for two placebo outcomes: Social Security payments and wages. Figure A.2a shows the proportion of taxpayers in our sample of IRA account holders who receive a social security payment at that age. Similarly, Figure A.2b shows the proportion of taxpayers in our sample of IRA account holders who receive a wage at that age. In both cases, there is no bunching the proportion of taxpayers at either $59\frac{1}{2}$ or $70\frac{1}{2}$. This implies that the jumps we are seeing at $59\frac{1}{2}$ and $70\frac{1}{2}$ are not confounded with other changes in financial circumstances during retirement or pre-retirement.

Figure A.2: Fraction of traditional IRA holders receiving Social Security or wages in 2005, by age

(a) Social Security benefits



(b) Wages



Notes: N = 1,449,868 taxpayers. Limited to taxpayers with a positive traditional IRA balance in 2004 or 2005.

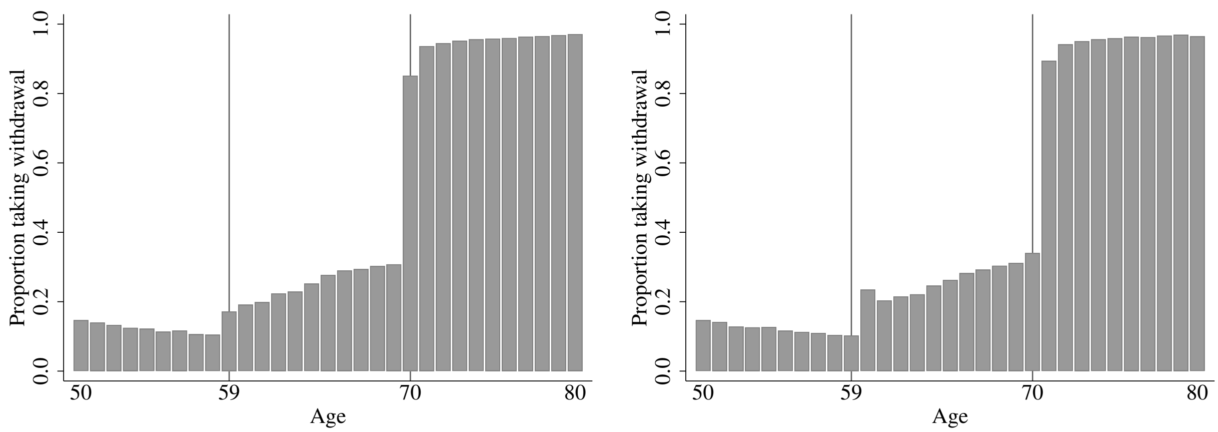
A.3.1.2 Timing of birthday

In our second diagnostic, we take advantage of the fact that the age thresholds kick in at taxpayers' half birthdays. We compare traditional IRA holders who turn age X and X.5 in the same calendar year versus traditional IRA holders who turn age X and X.5 in different calendar years. Individuals who turn 59 and $59\frac{1}{2}$ in the same year are eligible to take withdrawals penalty-free in that year, whereas individuals who turn 59 and $59\frac{1}{2}$ in different years are not eligible to take distributions penalty-free in the year they turn 59. If taxpayers are delaying withdrawals as a result of the age threshold, we should see a spike in the proportion of taxpayers withdrawing at age 59 for taxpayers that turned 59 and $59\frac{1}{2}$ in the same calendar year, but at age 60 for taxpayers that turned $59\frac{1}{2}$ in the calendar year after the year in which they turned 59.

Figure A.3 shows that this is exactly what happens. Figure A.3a shows taxpayers whose half-birthday is in the same calendar year as their birthday. For these taxpayers, we see increases in withdrawal rates at ages 59 and 70. Figure A.3b shows taxpayers whose half-birthday is in a different calendar year than their birthday. For these taxpayers, we see increases in withdrawal rates at ages 60 and 71. This is consistent with taxpayers changing their behavior as a result of exactly when these specific age thresholds apply.

Figure A.3: Fraction of traditional IRA holders taking a distribution in 2005, by timing of half-birthday

(a) Half-birthday in same calendar year (b) Half-birthday in next calendar year



Notes: Figure A.3a: N = 490,904 unique taxpayers. Figure A.3b: N = 517,808 unique taxpayers. Limited to taxpayers with a positive traditional IRA balance in 2004 or 2005.

A.3.2 Roth IRAs

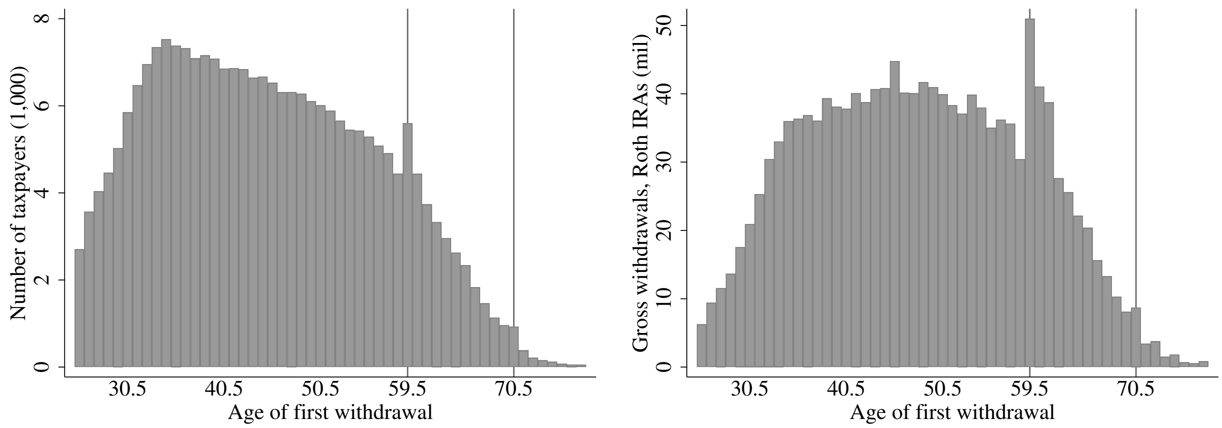
We created the same figures for Roth IRAs as we do to demonstrate bunching for traditional IRAs. We do not observe the same clear bunching for Roth IRAs as we do for traditional IRAs.

Figure A.4 shows the same empirical distribution as Figure 1.1 in Section 1.4.2, but for Roth IRA account holders. Figure A.4a shows the number of Roth IRA holders taking their first distribution from a Roth IRA at each age; Figure A.4b shows total amount taken at those distributions.

There are a few key things to note about this figure. First, the scale of these graphs is considerably different than that for traditional IRAs. Figure A.4a goes through 8 thousand, whereas Figure 1.1a is scaled through 150 thousand. Similarly, Figure A.4b is scaled through \$50 million, whereas Figure 1.1b is scaled through \$1.8 billion. We also observe small spikes at age $59\frac{1}{2}$, but not at $70\frac{1}{2}$. This is consistent with the fact that the early withdrawal penalty applies to Roth IRAs, but the excess accumulation penalty does not.

Figure A.4: Age of first distribution from a Roth IRA

(a) Number of Roth IRA holders (b) Gross distribution amount ($milion)



Notes: N = 236,062 unique taxpayers. Excludes taxpayers older than age 65 at the beginning of our sample period, and all data from 2009. Figure A.4a shows the number of taxpayers in our sample we observe taking their first distribution from a Roth IRA at each age, excluding early distributions (before age $59\frac{1}{2}$) with a known, qualifying exception. Figure A.4b shows the total amount withdrawn in the first distribution by Roth IRA account holders who took their first withdrawal at each age. Distribution amounts are adjusted for 2015 dollar values. Because taxpayers may hold both types of accounts, some taxpayers may appear in both Figure 1.1 (showing traditional IRAs in the main text) and this figure.

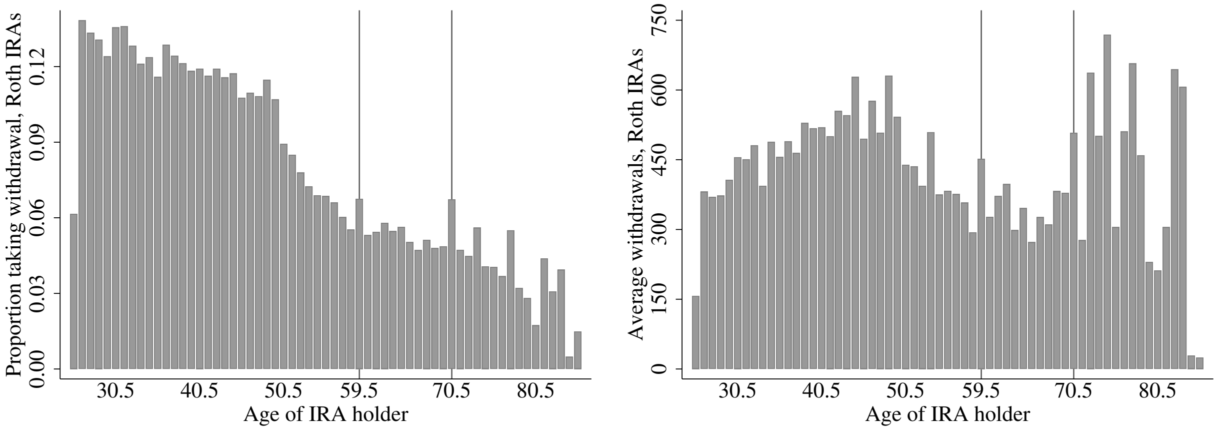
Figure A.5 shows the same empirical distribution as Figure 1.3 in Section 1.4.3, but for Roth IRA holders. The first panel shows the proportion of Roth IRA account holders taking a distribution in 2005; the second panel shows the average value of those distributions.

There are two things to note about these figures. First, we do not observe the same striking bunching at age $59\frac{1}{2}$ that we observe for traditional accounts. Second, there is an unexpected uptick in the number of Roth account holders who take their first distribution from a Roth IRA at age $70\frac{1}{2}$ (A.4a), and in the proportion of Roth account holders taking a distribution from a Roth IRA at age $70\frac{1}{2}$ (A.5a). This observation is unexpected because Roth IRA holders are not bound to the RMD rules. Figure A.6 splits Roth IRA account holders into those that also hold a traditional IRA (solid line) and those that only hold a Roth IRA (dashed line). If it's true that IRA account holders with both types of accounts are more likely to take non-rollover distributions from a Roth account at age $70\frac{1}{2}$ than IRA account holders who only have a Roth account, it would suggests that these taxpayers may not realize the Required Minimum Distribution rules do not apply to Roth IRAs.

Figure A.6a shows the number of taxpayers taking their first distribution by age. We see that there is an increase in the number of taxpayers taking their first distribution at age $70\frac{1}{2}$ for Roth IRA account holders that also have a traditional IRA, but not for Roth account holders who do not have a traditional IRA. Figure A.6b shows the proportion of Roth IRA holders taking a distribution at each age. While we observe an uptick at $70\frac{1}{2}$ for both groups,

Figure A.5: Distribution behavior in 2005: Roth IRAs

(a) Proportion of Roth IRA holders (b) Average distribution amount



Notes: *Notes*: $N = 269{,}353$ unique taxpayers. Excludes early distributions (before age $59\frac{1}{2}$) with a known, qualifying exception. Excludes taxpayers older than age 65 at the beginning of our sample period. Limited to taxpayers with a positive Roth IRA balance in 2004 or 2005. Figure A.5a shows the proportion of Roth IRA account holders taking a distribution in 2005, while Figure A.5b show the total amount distributed from Roth IRA by age in 2005. Distribution amounts are inflated to 2015 values. Because taxpayers may hold both types of accounts, some taxpayers may appear in both Figure 1.3 (showing traditional IRAs in the main text) and this figure.

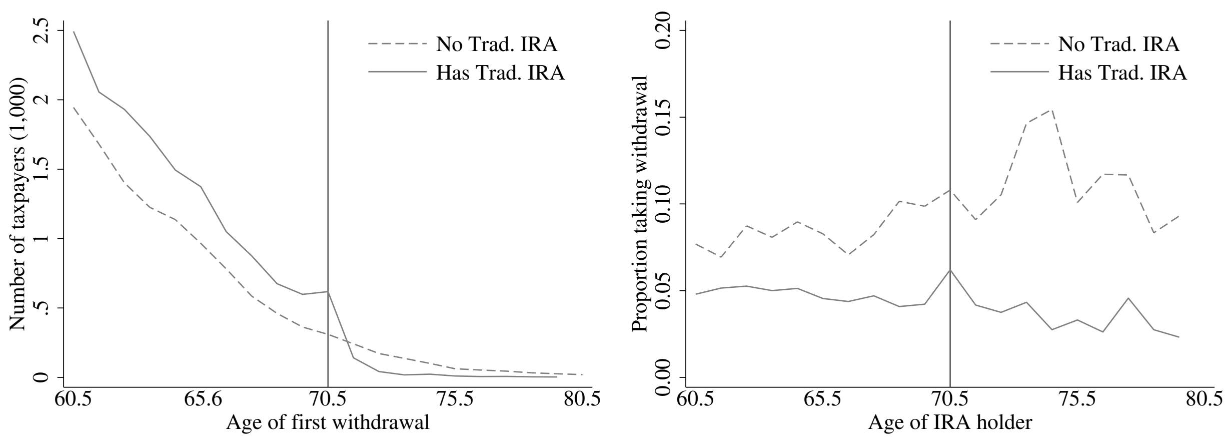
the trend for Roth IRA account holders who do not also have a traditional IRA is generally more erratic than that of Roth IRA account holders who do show evidence of also having a traditional IRA. These figures suggest that it may be true that taxpayers that hold both types of accounts are more likely to take distributions from their Roth IRA account at age $70\frac{1}{2}$ even though they are not required to.

A.4 Comparing different values of $\mathbf{a_{59.5,-}}$ and $\mathbf{a_{59.5,+}}$

We estimated $\widehat{B}_{59.5}$ using the values of $a_{59.5,-}$ and $a_{59.5,-}$ that minimized the difference between $\widehat{B}_{59.5}$ and $\widehat{M}_{59.5}$. We find that, for the number of taxpayers impacted, $a_{59.5,-} = 54.5$ and $a_{59.5,+} = 61.5$. Table A.5 shows the absolute value of $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for all possible combinations of $a_{59.5,-}$ and $a_{59.5,+}$ for the number of taxpayers that change when they take their first distribution. Table A.6 shows the same calculation, but for amount of money taken out at those first distributions. Similarly, Table A.7 shows the absolute value of $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for all possible combinations of $a_{59.5,-}$ and $a_{59.5,+}$ for the proportion of taxpayers taking a withdrawal by age in 2005, while Table A.8 shows the same calculation for the average withdrawal amount by age in 2005.

Figure A.6: Distribution behavior of Roth IRA account holders by whether or not they also hold a traditional IRA

(a) Age of first distribution from Roth IRA (b) Proportion taking distributions



Notes: Figure A.6a: N = 26,904 taxpayers; excludes taxpayers older than age 65 at the beginning of our sample period. Figure A.6b: N = 50,756; limited to taxpayers with a positive Roth IRA balance in 2004 or 2005. The dotted line represents Roth IRA account holders who do not also hold a traditional IRA. The solid line represents Roth IRA account holders who do show evidence of holding a traditional IRA.

Table A.5: Robustness check: $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for reduced-form analysis: number of taxpayers, first distribution (1000s)

	$a_{59.5,+}$										
$a_{59.5,-}$	59.5	60.5	61.5	62.5	63.5	64.5	65.5	66.5	67.5	68.5	69.5
25.5	448.9	797.5	773.3	691.4	603692.6	2163406.3	2038868.1	270894.3	1302.2	1325.4	9631493.1
26.5	9794.1	521.8	12571.6	44365.1	153170.0	312354.3	226137.8	65355.0	1283.0	1295.2	350244.7
27.5	987.1	99.4	1174.7	1924.7	4593.9	5207.5	1044.2	1221.9	1269.2	1268.5	162897.1
28.5	1002.2	222.8	1166.9	1658.5	3018.9	2761.9	642.0	961.7	917.0	28.7	9216.3
29.5	1037.6	579.1	1158.8	1365.9	1816.9	1590.2	197.0	274.9	22.1	1363.8	7727.6
30.5	1115.8	898.7	1214.0	1315.9	1527.6	1392.0	585.9	301.6	511.3	1689.6	7380.5
31.5	1133.5	1013.1	1205.5	1267.6	1405.3	1301.0	668.3	436.0	627.1	1717.0	7014.0
32.5	1123.2	1041.3	1180.2	1225.8	1334.0	1242.1	683.3	474.4	653.8	1678.2	6569.6
33.5	1090.1	1024.6	1136.4	1172.5	1264.0	1176.9	660.4	464.8	631.2	1584.6	5918.8
34.5	1049.3	991.4	1086.8	1115.8	1194.9	1108.8	622.6	435.8	585.3	1448.0	5010.9
35.5	1003.4	949.1	1032.1	1054.2	1121.3	1032.9	571.4	389.8	514.8	1257.0	3876.2
36.5	949.9	896.4	967.9	981.5	1034.2	938.9	497.9	316.0	405.0	992.4	2663.2
37.5	894.4	840.3	899.9	903.3	938.9	832.8	411.6	228.0	276.1	700.4	1640.0
38.5	839.8	783.8	830.7	822.5	838.5	720.2	322.2	141.0	154.0	439.0	929.5
39.5	786.1	727.4	761.0	740.0	735.6	606.4	238.9	68.1	58.8	244.5	499.3
40.5	729.9	667.8	686.8	652.3	627.4	490.5	159.4	5.0	16.5	103.5	239.5
41.5	672.4	606.4	610.8	563.8	521.7	383.4	94.1	39.4	63.9	16.9	95.8
42.5	612.7	543.0	533.7	476.6	422.0	288.6	42.1	69.9	93.3	36.1	14.1
43.5	553.2	481.0	460.2	397.1	336.3	213.4	7.4	84.6	104.8	61.2	25.3
44.5	494.5	421.3	392.2	326.9	264.8	155.2	15.7	91.3	108.1	73.1	45.0
45.5	434.4	361.3	326.1	261.1	200.3	103.3	40.1	103.8	118.9	90.5	69.0
46.5	377.0	305.7	267.3	204.8	147.4	61.9	59.6	114.4	128.2	104.5	87.6
47.5	322.9	254.5	214.8	156.0	102.6	27.1	77.8	125.9	138.9	119.1	106.0
48.5	273.1	208.4	168.9	114.1	64.8	2.6	94.7	137.8	150.3	133.8	124.1
49.5	229.1	168.4	130.1	79.4	33.9	26.8	108.8	148.0	160.1	146.2	139.5
50.5	189.4	132.7	95.9	48.9	6.7	48.6	122.8	159.1	171.1	159.7	156.0
51.5	156.3	103.6	68.6	25.0	14.1	64.7	132.3	165.9	177.6	168.0	166.3
52.5	126.5	77.3	44.0	3.4	33.2	80.0	142.3	173.8	185.3	177.6	178.3
53.5	100.0	53.9	22.2	15.9	50.4	94.0	151.7	181.4	192.8	187.1	190.0
54.5	75.2	31.8	1.4	34.7	67.3	108.4	162.2	190.5	202.1	198.6	204.3
55.5	51.5	10.3	19.1	53.4	84.8	123.7	174.2	201.4	213.6	212.5	221.5
56.5	30.6	8.6	36.8	69.6	99.5	136.3	183.6	209.5	221.8	222.5	233.2
57.5	10.5	26.9	54.2	85.5	114.2	149.2	193.5	218.2	230.6	233.1	245.4
58.5	9.5	45.3	71.8	102.0	129.6	163.0	204.6	228.4	241.1	245.5	259.3

Notes: Shows (in thousands) the absolute value of $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for each possible combination of $a_{59.5,-}$ and $a_{59.5,+}$ for our estimates of the number of taxpayers who changed when they took their first IRA distribution in response to the age thresholds.

Table A.6: Robustness check: $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for reduced-form analysis: gross distributions, first distribution ($bil)

	$a_{59.5,+}$										
$a_{59.5,-}$	59.5	60.5	61.5	62.5	63.5	64.5	65.5	66.5	67.5	68.5	69.5
25.5	7485.5	10.0	8.0	7.8	8061.2	21959.4	6154.0	21.9	23.0	23.5	110388.1
26.5	383.4	7.7	3.2	1.4	677.4	2002.8	17.0	21.2	21.8	16.9	3948.8
27.5	28.6	3.8	12.5	18.6	47.8	70.4	12.5	19.0	18.1	61.8	1813.2
28.5	28.7	6.3	16.1	22.9	40.9	48.1	5.2	1.2	18.9	48.8	89.8
29.5	28.9	13.4	23.1	27.8	35.1	36.8	21.2	20.3	30.7	46.9	73.5
30.5	28.6	20.4	26.3	28.8	32.4	33.1	24.0	23.5	31.3	44.9	68.3
31.5	28.2	23.6	27.2	28.8	31.2	31.7	24.5	24.2	31.0	43.5	65.0
32.5	27.3	24.1	26.8	28.0	29.8	30.1	23.8	23.4	29.7	41.3	60.3
33.5	26.4	23.9	26.1	27.0	28.6	28.8	22.8	22.4	28.2	38.7	54.1
34.5	25.5	23.4	25.2	26.0	27.3	27.4	21.8	21.2	26.4	35.6	45.6
35.5	24.4	22.4	24.0	24.6	25.7	25.6	20.1	19.3	23.5	30.7	33.6
36.5	23.3	21.4	22.7	23.2	24.0	23.6	18.2	17.0	20.1	24.8	21.6
37.5	22.1	20.2	21.4	21.6	22.1	21.4	15.9	14.3	16.1	18.3	11.4
38.5	20.9	19.0	19.9	19.9	20.0	18.8	13.4	11.3	11.8	12.1	4.2
39.5	19.6	17.7	18.3	17.9	17.6	16.0	10.6	8.1	7.7	6.7	0.6
40.5	18.1	16.2	16.5	15.8	15.0	12.9	7.8	5.1	4.1	2.6	3.7
41.5	16.8	14.8	14.7	13.7	12.5	10.3	5.5	3.0	1.9	0.3	4.7
42.5	15.2	13.2	12.8	11.5	10.1	7.7	3.5	1.1	0.0	1.5	5.6
43.5	13.7	11.6	11.0	9.5	7.9	5.6	2.0	0.1	1.1	2.4	5.8
44.5	12.1	10.0	9.2	7.6	6.0	3.9	0.7	1.0	2.0	3.1	6.1
45.5	10.6	8.5	7.6	6.0	4.4	2.5	0.2	1.7	2.6	3.6	6.2
46.5	9.0	7.1	6.0	4.5	3.0	1.2	1.1	2.5	3.3	4.2	6.6
47.5	7.6	5.7	4.6	3.2	1.8	0.2	1.8	3.0	3.8	4.7	6.8
48.5	6.3	4.6	3.5	2.2	0.9	0.6	2.4	3.5	4.2	5.0	7.0
49.5	5.2	3.6	2.6	1.3	0.1	1.1	2.8	3.8	4.4	5.2	7.0
50.5	4.3	2.7	1.7	0.6	0.5	1.7	3.2	4.1	4.7	5.4	7.2
51.5	3.5	2.1	1.1	0.1	0.9	2.0	3.3	4.2	4.8	5.5	7.0
52.5	2.8	1.5	0.6	0.4	1.3	2.3	3.5	4.3	4.9	5.5	7.0
53.5	2.2	1.0	0.2	0.7	1.6	2.5	3.7	4.4	4.9	5.5	6.8
54.5	1.6	0.5	0.3	1.1	1.9	2.8	3.9	4.6	5.1	5.6	6.8
55.5	1.1	0.0	0.7	1.5	2.3	3.1	4.1	4.8	5.3	5.8	6.9
56.5	0.6	0.4	1.1	1.8	2.6	3.3	4.3	4.9	5.3	5.8	6.8
57.5	0.2	0.8	1.4	2.2	2.9	3.6	4.5	5.1	5.5	6.0	6.9
58.5	0.3	1.2	1.8	2.5	3.2	3.9	4.7	5.2	5.7	6.1	6.9

Notes: Shows (in billions) the absolute value of $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for each possible combination of $a_{59.5,-}$ and $a_{59.5,+}$ for our estimates of the value of the gross distributions taken by taxpayers who changed when they took their first IRA distribution in response to the age thresholds.

Table A.7: Robustness check: $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for reduced-form analysis: number of taxpayers, single year (proportion)

	$a_{59.5,+}$										
$a_{59.5,-}$	59.5	60.5	61.5	62.5	63.5	64.5	65.5	66.5	67.5	68.5	69.5
25.5	3.6234	3.6369	3.5574	3.3918	3.0002	2.4033	2.5100	3.2003	4.2430	6.0148	10.4817
26.5	0.1523	0.3600	0.5712	0.7627	0.9371	1.0700	1.1355	1.1857	1.2524	1.3824	10.4817
27.5	0.8126	0.9828	1.1586	1.3227	1.4772	1.6005	1.6673	1.7219	1.7942	1.9335	48.8059
28.5	1.0516	1.2062	1.3677	1.5207	1.6671	1.7862	1.8531	1.9092	1.9842	2.1297	25.1869
29.5	1.1796	1.3254	1.4788	1.6255	1.7670	1.8835	1.9503	2.0074	2.0843	2.2349	15.7760
30.5	1.2365	1.3760	1.5234	1.6651	1.8026	1.9166	1.9826	2.0395	2.1166	2.2684	9.9137
31.5	1.2662	1.4006	1.5432	1.6809	1.8150	1.9265	1.9914	2.0476	2.1239	2.2740	6.6321
32.5	1.2727	1.4028	1.5411	1.6751	1.8057	1.9146	1.9778	2.0323	2.1060	2.2496	4.3943
33.5	1.2753	1.4017	1.5363	1.6668	1.7943	1.9005	1.9619	2.0145	2.0851	2.2204	3.1488
34.5	1.2681	1.3909	1.5220	1.6491	1.7733	1.8766	1.9356	1.9854	2.0512	2.1733	2.2295
35.5	1.2607	1.3803	1.5079	1.6318	1.7527	1.8530	1.9096	1.9564	2.0169	2.1243	1.6922
36.5	1.2568	1.3734	1.4979	1.6187	1.7365	1.8340	1.8882	1.9321	1.9875	2.0808	1.4344
37.5	1.2398	1.3531	1.4741	1.5913	1.7055	1.7992	1.8500	1.8891	1.9358	2.0057	1.0613
38.5	1.2228	1.3329	1.4504	1.5641	1.6745	1.7644	1.8114	1.8454	1.8829	1.9281	0.8241
39.5	1.2137	1.3209	1.4353	1.5457	1.6525	1.7389	1.7827	1.8125	1.8426	1.8684	0.7739
40.5	1.2001	1.3043	1.4153	1.5221	1.6252	1.7076	1.7476	1.7722	1.7930	1.7957	0.6955
41.5	1.1781	1.2789	1.3861	1.4890	1.5875	1.6652	1.7001	1.7179	1.7265	1.6995	0.5596
42.5	1.1523	1.2496	1.3529	1.4514	1.5451	1.6174	1.6469	1.6569	1.6520	1.5936	0.4340
43.5	1.1232	1.2169	1.3159	1.4099	1.4983	1.5650	1.5883	1.5899	1.5705	1.4803	0.3215
44.5	1.0951	1.1851	1.2799	1.3692	1.4525	1.5134	1.5307	1.5241	1.4911	1.3732	0.2520
45.5	1.0685	1.1548	1.2455	1.3302	1.4082	1.4635	1.4750	1.4606	1.4154	1.2750	0.2167
46.5	1.0400	1.1226	1.2088	1.2888	1.3614	1.4108	1.4163	1.3942	1.3370	1.1766	0.1838
47.5	1.0095	1.0882	1.1700	1.2450	1.3120	1.3555	1.3549	1.3250	1.2563	1.0787	0.1525
48.5	0.9724	1.0469	1.1238	1.1935	1.2544	1.2913	1.2839	1.2453	1.1643	0.9692	0.0976
49.5	0.9327	1.0029	1.0749	1.1390	1.1938	1.2239	1.2097	1.1629	1.0705	0.8613	0.0459
50.5	0.8672	0.9319	0.9976	1.0547	1.1013	1.1223	1.0981	1.0385	0.9281	0.6935	0.1064
51.5	0.7967	0.8559	0.9151	0.9650	1.0032	1.0150	0.9810	0.9092	0.7825	0.5274	0.2472
52.5	0.7117	0.7648	0.8169	0.8588	0.8880	0.8896	0.8447	0.7596	0.6155	0.3396	0.4151
53.5	0.6208	0.6676	0.7125	0.7464	0.7664	0.7581	0.7027	0.6051	0.4457	0.1542	0.5734
54.5	0.5261	0.5666	0.6044	0.6304	0.6416	0.6237	0.5586	0.4500	0.2782	0.0232	0.7172
55.5	0.4244	0.4586	0.4891	0.5071	0.5095	0.4822	0.4078	0.2891	0.1067	0.2012	0.8606
56.5	0.3180	0.3459	0.3693	0.3795	0.3734	0.3372	0.2543	0.1268	0.0638	0.3742	0.9974
57.5	0.2146	0.2365	0.2533	0.2564	0.2426	0.1988	0.1091	0.0246	0.2198	0.5270	1.1095
58.5	0.1059	0.1219	0.1321	0.1283	0.1072	0.0562	0.0397	0.1787	0.3769	0.6794	1.2236

Notes: Shows the absolute value of $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for each possible combination of $a_{59.5,-}$ and $a_{59.5,+}$ for our estimates of the value of the proportion of taxpayers who changed when they took a distribution from a traditional IRA in response to the age thresholds.

Table A.8: Robustness check: $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for reduced-form analysis: gross distributions, single year ($mil)

	$a_{59.5,+}$										
$a_{59.5,-}$	59.5	60.5	61.5	62.5	63.5	64.5	65.5	66.5	67.5	68.5	69.5
25.5	19.474	28.539	28.933	30.573	77.559	123.478	151.917	252.767	253.322	79.845	2.047
26.5	24.004	24.751	24.675	24.646	25.926	26.008	24.918	24.041	21.156	18.426	2.047
27.5	25.321	25.909	25.874	25.873	26.957	27.044	26.082	25.283	22.542	19.906	145.757
28.5	25.300	25.817	25.784	25.782	26.784	26.863	25.941	25.166	22.485	19.890	49.319
29.5	25.369	25.850	25.820	25.818	26.775	26.851	25.956	25.200	22.570	20.021	37.169
30.5	24.854	25.300	25.260	25.248	26.161	26.222	25.334	24.574	21.960	19.385	6.531
31.5	24.323	24.741	24.691	24.668	25.543	25.588	24.700	23.930	21.321	18.694	5.961
32.5	23.757	24.149	24.088	24.051	24.891	24.917	24.024	23.236	20.619	17.904	12.496
33.5	23.192	23.560	23.487	23.436	24.240	24.245	23.342	22.530	19.895	17.064	15.374
34.5	22.618	22.963	22.877	22.811	23.578	23.559	22.642	21.799	19.134	16.157	16.697
35.5	22.081	22.403	22.304	22.223	22.954	22.909	21.977	21.100	18.401	15.272	16.248
36.5	21.642	21.945	21.835	21.741	22.438	22.371	21.425	20.520	17.800	14.556	13.695
37.5	21.090	21.370	21.245	21.132	21.789	21.691	20.721	19.769	16.998	13.560	13.516
38.5	20.717	20.979	20.844	20.718	21.342	21.222	20.241	19.262	16.481	12.970	10.446
39.5	20.308	20.550	20.404	20.263	20.851	20.703	19.707	18.694	15.894	12.291	8.505
40.5	19.935	20.160	20.002	19.848	20.400	20.227	19.218	18.175	15.370	11.717	6.452
41.5	19.471	19.674	19.501	19.328	19.837	19.631	18.597	17.509	14.672	10.911	5.714
42.5	18.871	19.047	18.853	18.654	19.110	18.859	17.784	16.625	13.717	9.759	6.229
43.5	18.454	18.609	18.401	18.185	18.600	18.319	17.227	16.035	13.130	9.173	4.886
44.5	17.803	17.928	17.695	17.449	17.806	17.472	16.335	15.064	12.088	7.963	5.494
45.5	17.430	17.536	17.293	17.032	17.351	16.993	15.849	14.565	11.635	7.631	3.882
46.5	16.914	16.995	16.733	16.451	16.720	16.324	15.157	13.833	10.909	6.932	3.422
47.5	16.267	16.318	16.032	15.719	15.930	15.484	14.279	12.893	9.944	5.933	3.706
48.5	15.625	15.646	15.336	14.995	15.148	14.656	13.419	11.983	9.030	5.040	3.813
49.5	15.110	15.108	14.782	14.422	14.529	14.007	12.761	11.312	8.422	4.586	3.208
50.5	13.921	13.866	13.489	13.067	13.076	12.456	11.112	9.519	6.479	2.386	5.651
51.5	12.900	12.800	12.383	11.913	11.842	11.150	9.742	8.062	4.970	0.825	6.897
52.5	11.524	11.364	10.890	10.353	10.178	9.383	7.875	6.056	2.838	1.486	9.229
53.5	10.218	10.005	9.481	8.885	8.618	7.738	6.154	4.238	0.960	3.410	10.885
54.5	8.715	8.442	7.860	7.197	6.828	5.854	4.183	2.156	1.196	5.631	12.914
55.5	6.977	6.636	5.987	5.247	4.768	3.687	1.916	0.234	3.678	8.199	15.357
56.5	5.270	4.864	4.156	3.347	2.768	1.597	0.251	2.490	5.972	10.488	17.377
57.5	3.497	3.028	2.260	1.385	0.711	0.544	2.460	4.772	8.268	12.745	19.335
58.5	1.560	1.024	0.192	0.754	1.526	2.867	4.853	7.237	10.750	15.189	21.520

Notes: Shows (in millions) the absolute value of $\widehat{B}_{59.5} - \widehat{M}_{59.5}$ for each possible combination of $a_{59.5,-}$ and $a_{59.5,+}$ for our estimates of change of the value of the average distribution from traditional IRAs in response to the age thresholds.

A.5 Intuition from a stylized model

In this section, we provide intuition for why we would expect taxpayers to bunch at the age thresholds in order to avoid the early withdrawal penalty. We consider a simple, highly-stylized model that allows us to graphically demonstrate our anticipated bunching. We also discuss how we can use bunching in this setting to identify the elasticity of intertemporal substitution.

A.5.1 A simple model with no penalties

Consider a taxpayer with assets A in an account earning a rate of return $R = 1 + r$. These assets are the only source of wealth available to the taxpayer, and the taxpayer must withdraw all of the assets at once. Once the assets are withdrawn, they stop earning the rate of return and the taxpayer's lifetime consumption is then equal to the value of the account. The taxpayer chooses when to withdraw the assets and their subsequent consumption stream in order to maximize her lifetime utility from consumption, as shown in Equation A.1:

$$\max_{\{c_t\},t^*} \sum_{t=0}^{T} \beta^t u(c_t) \tag{A.1}$$
$$\text{subject to}$$
$$\sum_{t=t^*}^{T} c_t = R^{t^*} A$$

The benefit of withdrawing in an earlier time period is the taxpayer receives utility from consumption earlier (the discount factor does not have as much of an impact on lifetime utility). The benefit of waiting to withdraw is the account grows in value and total lifetime consumption is higher. The taxpayer will pick t^* to equalize the marginal benefit of these two options, given their initial level of assets A and the rate of return R.

A.5.2 Anticipated bunching due to an early withdrawal penalty

Now introduce a penalty ρ on withdrawing before $t = t_p$. This reduces the value of lifetime consumption for withdrawing before $t = t_p$ by ρ times the values of the assets at the point of withdrawal. The taxpayer's new problem is shown in Equation A.2:

$$\max_{t^*} \sum_{t=0}^{T} \beta^t u(c_t) \tag{A.2}$$

$$\text{subject to}$$

$$\sum_{t=t^*}^{T} c_t = R^{t^*} A - \rho R^{t^*} A \mathbb{I}(t^* < t_p)$$

Some taxpayers who otherwise would have chosen $t^* < t_p$ will wait to withdraw until after $t = t_p$.

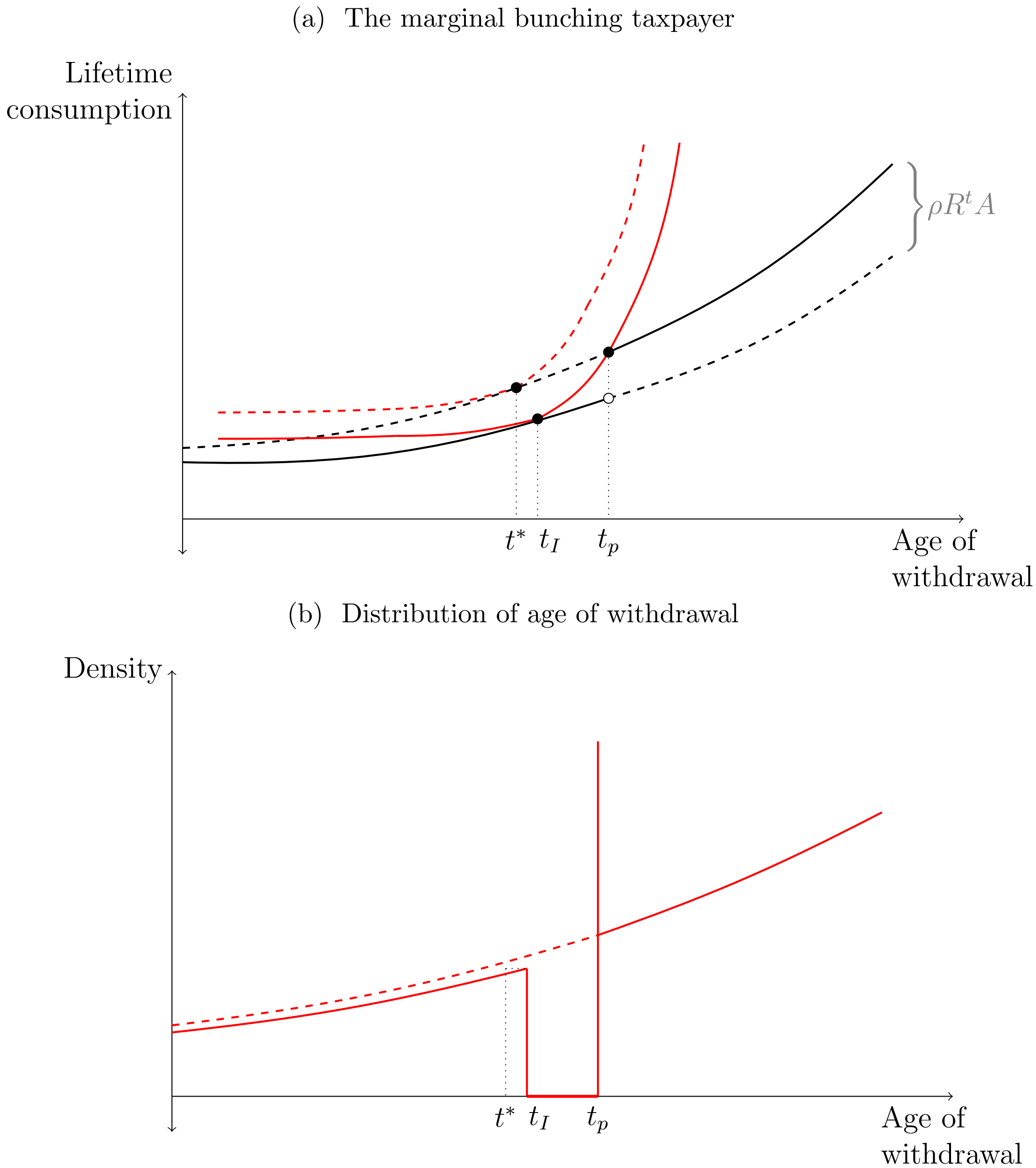
Figure A.7 shows the choice of a taxpayer deciding when to withdraw assets from an account when faced with an early withdrawal penalty and the subsequent bunching in the distribution of chosen ages for withdrawal. Figure A.7a illustrates how the taxpayer's choice of when to withdraw the assets from the account translates directly into total lifetime consumption. The solid black line represents the budget constraint when there is a penalty for withdrawing the assets before age t_p. The slope of the budget constraint is equal to $tR^{t-1}A$ above t_p and $(1-\rho)tR^{t-1}A$ below. The discrete change in the budget constraint at $t = t_p$ is due to the lack of early withdrawal penalty above that age. The solid red line represents the indifference curve for the marginal buncher. The marginal buncher is indifferent to withdrawing at ages t_I and t_p. The dashed red line represents the indifference curve for the marginal buncher in the absence of the early withdrawal penalty. In this scenario, the taxpayer would have chosen t^*.

Figure A.7b demonstrates how the early withdrawal penalty might impact the distribution of withdrawal ages chosen. We've drawn a potential, arbitrary distribution as a dotted red line. The solid red line illustrates what happens when we introduce the early withdrawal penalty. Taxpayers for whom the optimal age of withdrawal would have been between t_I and t_p now prefer to wait to withdraw until t_p. This results in bunching in the distribution of withdrawal ages at t_p.

A.5.3 Identifying the elasticity of intertemporal substitution

The elasticity of intertemporal substitution (EIS) measures the relative response of consumption in period 1 versus consumption in period 2 when the rate of return changes. When the real rate of return increases, two things happen. First, the taxpayer's lifetime budget increases. This suggests an increase in consumption in all periods (an income effect). Second, consumption in period 1 becomes more expensive, because it means the taxpayer saves less (and therefore benefits less from the increased interest rate). This suggests a

Figure A.7: Bunching due to early withdrawal penalty in a highly stylized model

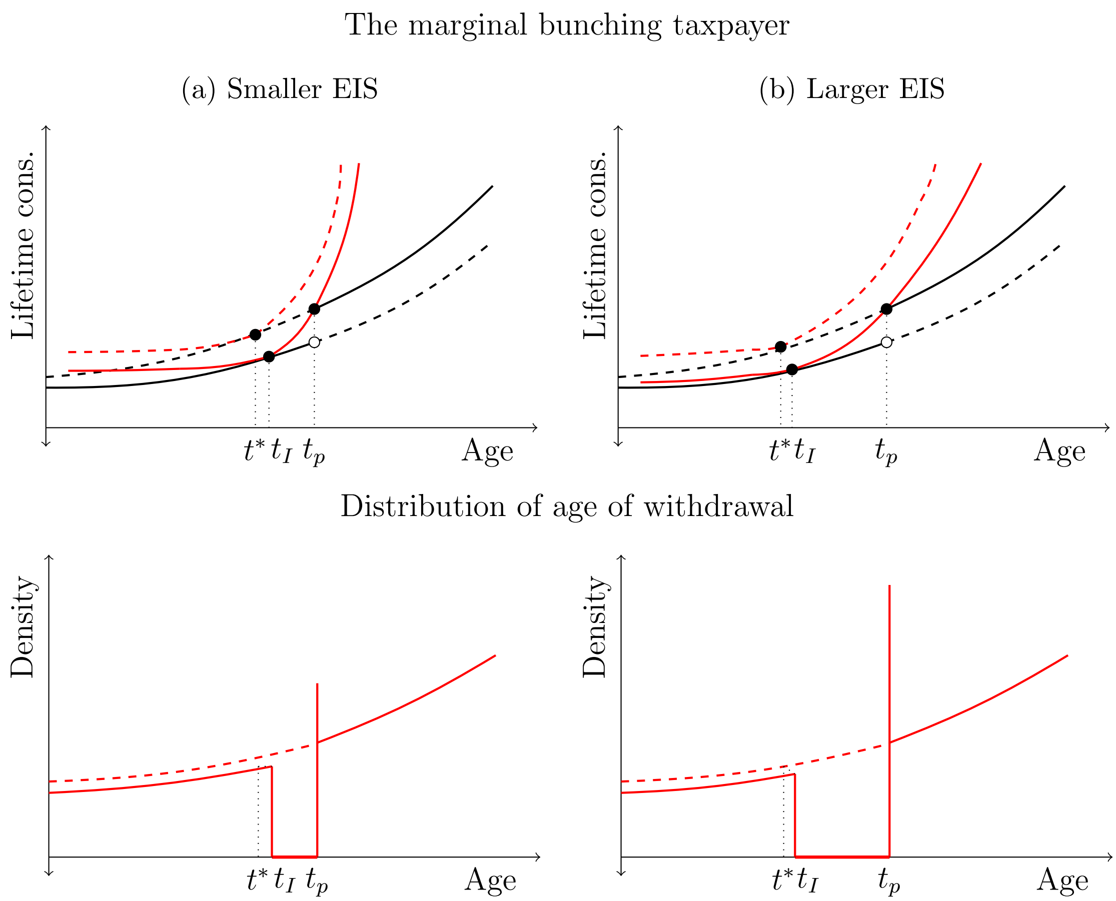


Notes: The distribution of "age of withdrawal" is arbitrary and used for illustrative purposes only.

decrease of consumption in period 1 in favor of consumption in period 2 (a substitution effect). The net effect of these two forces is represented by the EIS.

The value of the EIS directly impacts the curvature of the indifference curve. Greater values of the EIS equal less curved indifference curves. The more elastic the taxpayer's preferences are between periods 1 and periods 2, the less period 2 consumption the taxpayer needs to be compensated with to maintain their utility level if their period 1 consumption decreases. We can take advantage of the relationship between the EIS and the curvature of the indifference curve to identify the EIS in our setting. Figure A.8 shows the same setting as in Figure A.7, but where the indifference curves are drawn using different values of the EIS. The taxpayers represented in the left-hand panel have more strongly curved indifference curves (i.e., a lower EIS) than the taxpayers in the right-hand panel. This causes the space between t_I and t_p to be narrower, resulting in fewer taxpayers bunching at t_p. The amount of bunching at t_p, then, is directly related to the value of the EIS.

Figure A.8: Identifying the EIS using a bunching response



Notes: The distribution of "age of withdrawal" is arbitrary and used for illustrative purposes only.

The previous argument assumed that we knew the value of the discount factor. In the model presented above, present versus future consumption decisions are dictated by two parameters: the EIS and the discount factor. If we only had one notch, we would not separately be able to identify either parameters. However, there are two age notches in our setting: the notch used in the simple model above (equivalent to the notch created by the early withdrawal penalty), and the notch generated by the onset of required minimum distributions. Because we have two bunching conditions, we are able to separately identify the two relevant parameters.

A.6 Parameters used in the structural model

Table A.9 defines the parameters used in the life-cycle model described in Section 1.5.

Table A.9: Structural model parameters

Preference parameters	
β	Discount factor
σ	Elasticity of intertemporal substitution
A	Bequest weight
α	Bequest elasticity
State variables	
Ω	Set of all state variables
S_t	Stock of assets in standard savings account
A_t	Stock of assets in tax-benefited savings account
θ	Labor income shock
z_t	Total income subject to income tax ($= y_t + P \cdot \mathbb{I}[t \geq t_P] + q(S_t \cdot r)$)
t	period (equivalent to age)
P	Value of annual pension
t_P	Age at which individual claims pension
Choice variables	
S_{t+1}	Next period stock of assets in standard savings account
A_{t+1}	Next period stock of assets in tax-benefited savings account
c_t	Consumption
Labor income	
y_t	Labor income
$\{\rho_i\}_{i=0}^{3}$	Deterministic part of labor income
ε_{it}	Stochastic part of labor income
η	Autocorrelation of stochastic part of labor income
σ_ζ^2	Variance of stochastic term
Tax-benefited savings account parameters	
$\overline{a}_t$	Contribution limit for tax-benefited savings account
t_e	Age before which early withdrawal penalty applies
p_e	Rate of early withdrawal penalty
t_{rmd}	Age after which excess accumulation penalty applies
p_{rmd}	Rate of excess accumulation penalty
$a_{t,rmd}$	Value of RMD (function of age and A_t)
Taxes	
τ_y	Income tax function
τ_e	Amount of early withdrawal penalty
τ_{rmd}	Amount of excess accumulation penalty
Additional parameters	
r_A	Rate of return on regular savings
r_S	Rate of return on tax-benefited savings
π_t	The probability of living to period t conditional on surviving to period $t-1$
W_{t+1}	Bequeathed wealth if die at end of period t ($= (1 + r_S)S_{t+1} + (1 + r_A)A_{t+1}$)

A.7 Numerical solution

We numerically solve the dynamic optimization problem presented in Section 1.5 using backwards iteration. Two of our state variables are already discrete: age, and age at which the individual claims Social Security. We solve the model for three values of claiming Social Security age: 62, 67, and 70. These ages respectively correspond to the earliest age an individual is eligible to claim Social Security, the full retirement age, and the age after which delayed retirement credits stop permanently increasing the monthly (and therefore annual) Social Security benefit for individuals born in 1940 or later.

We have three continuous state variables that we need to discretize: labor income shocks and the stock of both regular savings and tax-benefited savings. Labor income shocks are placed on a 5-point grid using the Rouwenhorst method (Rouwenhorst, 1995).

Discretizing the stocks of tax-benefited saving is finicky in our setting. For each period in our model, we need at least one non-zero contribution option, and a "no contribution" option. Ensuring that both of these options were available at each period required that we make some non-standard choices. We space the grid at the oldest age individuals in our model can make a contribution (70) by the contribution limit at age 69. We then create age-specific grids working backwards, where grid point $A_{j,t} = A_{j,t+1}/(1 + r_A)$. Because savings accounts will be depleted as the simulated individuals near the final period, we also create age-specific grids from 70 forward such that $A_{j,t} = A_{j,t-1}/(1 + r_A)$.

This methodology ensures that, for any grid point $A_{j,t}$ with $t < 70$, the model can choose not to contribute by choosing point $A_{j,t+1}$.[6] This procedure also ensures that the model always has at least one gridpoint to choose that is within the contribution limit, because the grid points are spaced within the contribution limit at the oldest age one can make a contribution and become closer together as age decreases. We set the standard savings grid to be equal to the grid used for tax-benefited savings so that, when solving the model, it is never beneficial to take a distribution from the tax-benefited account in order to save in the regular savings account. In other words, the next-period options for both accounts are the same.

In order to not severely truncate the maximum level of assets that can be saved in these accounts, we use 90-point grids for both regular and tax-benefited savings. This is considerably larger than is often used in these models (e.g., Choukhmane (2021) uses 20 grid points for liquid assets and 16 grid points for retirement wealth). Solving the model for ages 40-85 with two 90-point asset grids, a 5-point labor income shock grid, and 3 possible ages for

[6]This is because choosing not to contribute is equivalent to picking a next-period stock equal to the current stock multiplied by the return.

claiming Social Security means that we are finding the solution for 5,589,000 combinations of the state variables, each of which faces 8,100 possible combinations of the choice variables.[7] Even with these large grids, the top 3.7% of IRA balances in our simulated sample are constrained by our highest grid value at age 40, and the top 1.6% are constrained at age 41.

We minimize the score ($Z(\Phi, n)$, see Equation 1.6.4 using Nelder-Mead optimization. Because Nelder-Mead can get stuck at local minima, we pick our starting point for the SMM algorithm as follows: we generate 4,000 points defined by a Sobol sequence over a wide range of possible parameter values: between 0.1 and 2.0 for the EIS (σ) and bequest elasticity (α), between 0.85 and 1.0 for the discount factor (β), and between 0.25 and 5.0 for the bequest weight (A). We then calculate the score for all 4,000 preference parameter combinations. We use the parameter combination that minimizes the score as our initial guess for the SMM procedure.

Estimating the model is computationally intensive. The code is written in Python. All functions are JIT-compiled in "nopython" mode using Numba v0.49.1 with parallel = True, cache = True, and all object types defined.

[7]We solve for optimal next-period stocks in both assets using a grid search, which implies post-tax consumption. Not every state space has this many options because of the contribution limit to the tax-benefited savings account and the no borrowing constraint.

A.8 Counterfactual policy analysis

A.8.1 Changing the excess accumulation penalty

Table A.10 shows the results of changing the excess accumulation penalty. We compare the values of four outcomes to the base policy with an age threshold = $70\frac{1}{2}$ and penalty rate = 50%. For all outcomes considered, a positive value indicates that our simulated individuals are, on average, better off in the counterfactual world, while a negative value indicates that our simulated individuals are, on average, better off in the base policy.

Table A.10a shows a measure of equivalent variation: the average amount of income that our simulated individuals would need to receive (give up) in the base policy to reach the present discounted value of lifetime utility for each counterfactual policy. We show the value in levels, and as a percentage of average income at age $40\frac{1}{2}$ (i.e., the first period in our model). Table A.10b shows the change in the present discounted value of lifetime total tax remittances relative to the base policy, where total tax remittances include income taxes plus any tax penalties paid. Table A.10c shows the change in the present discounted value of of lifetime income tax remittances relative to the base policy. Table A.10d shows the change in the bequeathed IRA balance relative to the base policy. For each panel, we show the amount in levels and as a percentage of the value in the base policy.

Table A.10: Results of changing the excess accumulation penalty

(a) Change in income at age $40\frac{1}{2}$ under base policy to reach counterfactual PDV lifetime utility

Age	40% penalty		50% penalty		60% penalty		70% penalty	
threshold	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$68\frac{1}{2}$	-761	-0.0037%	-811	-0.0040%	-851	-0.0042%	-878	-0.0043%
$69\frac{1}{2}$	-335	-0.0016%	-362	-0.0018%	-395	-0.0019%	-411	-0.0020%
$70\frac{1}{2}$	22	0.0001%	0	0.0000%	-38	-0.0002%	-53	-0.0003%
$71\frac{1}{2}$	204	0.0010%	186	0.0009%	168	0.0008%	151	0.0007%
$72\frac{1}{2}$	344	0.0017%	339	0.0017%	325	0.0016%	309	0.0015%
$73\frac{1}{2}$	459	0.0023%	454	0.0022%	451	0.0022%	441	0.0022%
$74\frac{1}{2}$	533	0.0026%	531	0.0026%	530	0.0026%	525	0.0026%
$75\frac{1}{2}$	565	0.0028%	570	0.0028%	570	0.0028%	571	0.0028%
$76\frac{1}{2}$	579	0.0029%	588	0.0029%	595	0.0029%	597	0.0029%
$77\frac{1}{2}$	568	0.0028%	586	0.0029%	597	0.0029%	604	0.0030%
$78\frac{1}{2}$	550	0.0027%	569	0.0028%	579	0.0028%	584	0.0029%

Notes: Each iteration includes 10,000 unique simulated individuals. This table compares the base policy for the excess accumulation policy (age threshold = $70\frac{1}{2}$, penalty rate = 50%) and a range of counterfactual policies. Table A.10a shows the average amount of income that our simulated individuals would need to receive (give up) in the base policy to reach the present discounted value of lifetime utility for each counterfactual policy. Table A.10b shows the change in the present discounted value of lifetime total tax remittances. Table A.10c shows the change in the present discounted value of lifetime income tax remittances. Table A.10d shows the change in the bequeathed IRA balance. For each panel, we show the amount in levels and as a percentage of the value in the base policy.

Table A.10: Results of changing the excess accumulation penalty, continued

(b) Change in PDV lifetime total tax remittances relative to base policy

Age	40% penalty		50% penalty		60% penalty		70% penalty	
threshold	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$68\frac{1}{2}$	-57	-0.0178%	8	0.0026%	53	0.0165%	77	0.0243%
$69\frac{1}{2}$	-27	-0.0086%	59	0.0187%	122	0.0385%	140	0.0442%
$70\frac{1}{2}$	-94	-0.0296%	0	0.0000%	64	0.0201%	88	0.0276%
$71\frac{1}{2}$	-124	-0.0391%	-38	-0.0120%	25	0.0079%	67	0.0211%
$72\frac{1}{2}$	-167	-0.0526%	-81	-0.0254%	-16	-0.0052%	20	0.0062%
$73\frac{1}{2}$	-225	-0.0707%	-143	-0.0449%	-75	-0.0237%	-32	-0.0102%
$74\frac{1}{2}$	-288	-0.0905%	-206	-0.0649%	-140	-0.0442%	-103	-0.0323%
$75\frac{1}{2}$	-373	-0.1175%	-295	-0.0928%	-236	-0.0743%	-195	-0.0614%
$76\frac{1}{2}$	-478	-0.1504%	-404	-0.1273%	-346	-0.1089%	-311	-0.0980%
$77\frac{1}{2}$	-601	-0.1890%	-534	-0.1682%	-480	-0.1509%	-445	-0.1402%
$78\frac{1}{2}$	-731	-0.2299%	-672	-0.2115%	-624	-0.1964%	-596	-0.1875%

(c) Change in PDV lifetime income tax remittances relative to base policy

Age	40% penalty		50% penalty		60% penalty		70% penalty	
threshold	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$68\frac{1}{2}$	-205	-0.0648%	-73	-0.0232%	50	0.0157%	156	0.0495%
$69\frac{1}{2}$	-109	-0.0344%	13	0.0040%	149	0.0473%	250	0.0792%
$70\frac{1}{2}$	-123	-0.0388%	0	0.0000%	137	0.0434%	240	0.0761%
$71\frac{1}{2}$	-98	-0.0310%	33	0.0104%	161	0.0509%	240	0.0758%
$72\frac{1}{2}$	-95	-0.0300%	27	0.0086%	142	0.0451%	234	0.0740%
$73\frac{1}{2}$	-97	-0.0307%	25	0.0078%	128	0.0405%	205	0.0649%
$74\frac{1}{2}$	-125	-0.0395%	-7	-0.0021%	95	0.0301%	167	0.0528%
$75\frac{1}{2}$	-159	-0.0504%	-56	-0.0176%	39	0.0124%	98	0.0311%
$76\frac{1}{2}$	-212	-0.0670%	-125	-0.0396%	-44	-0.0138%	16	0.0050%
$77\frac{1}{2}$	-285	-0.0901%	-208	-0.0659%	-133	-0.0421%	-84	-0.0265%
$78\frac{1}{2}$	-377	-0.1192%	-309	-0.0978%	-239	-0.0757%	-198	-0.0627%

(d) Bequeathed IRA balance relative to base policy

Age	40% penalty		50% penalty		60% penalty		70% penalty	
threshold	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$68\frac{1}{2}$	-5,580	-8.6%	-7,286	-11.3%	-8,910	-13.8%	-10,176	-15.8%
$69\frac{1}{2}$	-1,937	-3.0%	-3,527	-5.5%	-5,159	-8.0%	-6,462	-10.0%
$70\frac{1}{2}$	1,441	2.2%	0	0.0%	-1,539	-2.4%	-2,863	-4.4%
$71\frac{1}{2}$	3,554	5.5%	1,992	3.1%	516	0.8%	-352	-0.5%
$72\frac{1}{2}$	5,724	8.9%	4,283	6.6%	2,946	4.6%	1,976	3.1%
$73\frac{1}{2}$	7,504	11.6%	6,198	9.6%	5,045	7.8%	4,297	6.7%
$74\frac{1}{2}$	9,360	14.5%	8,130	12.6%	7,041	10.9%	6,330	9.8%
$75\frac{1}{2}$	11,053	17.1%	10,004	15.5%	9,022	14.0%	8,397	13.0%
$76\frac{1}{2}$	12,642	19.6%	11,782	18.3%	10,926	16.9%	10,309	16.0%
$77\frac{1}{2}$	14,246	22.1%	13,454	20.9%	12,657	19.6%	12,142	18.8%
$78\frac{1}{2}$	15,788	24.5%	15,106	23.4%	14,417	22.3%	14,033	21.7%

Notes: Each iteration includes 10,000 unique simulated individuals. This table compares the base policy for the excess accumulation policy (age threshold = $70\frac{1}{2}$, penalty rate = 50%) and a range of counterfactual policies. Table A.10a shows the average amount of income that our simulated individuals would need to receive (give up) in the base policy to reach the present discounted value of lifetime utility for each counterfactual policy. Table A.10b shows the change in the present discounted value of lifetime total tax remittances. Table A.10c shows the change in the present discounted value of lifetime income tax remittances. Table A.10d shows the change in the bequeathed IRA balance. For each panel, we show the amount in levels and as a percentage of the value in the base policy.

A.8.2 Changing the early withdrawal penalty

Table A.11 shows the results of changing the early withdrawal penalty. We compare the values of four outcomes to the base policy with an age threshold = $59\frac{1}{2}$ and penalty rate = 10%. For all outcomes considered, a positive value indicates that our simulated individuals are, on average, better off in the counterfactual world, while a negative value indicates that our simulated individuals are, on average, better off in the base policy.

Table A.11a shows a measure of equivalent variation: the average amount of income that our simulated individuals would need to receive (give up) in the base policy to reach the present discounted value of lifetime utility for each counterfactual policy. We show the value in levels, and as a percentage of average income at age $40\frac{1}{2}$ (i.e., the first period in our model). Table A.11b shows the change in the present discounted value of lifetime total tax remittances relative to the base policy, where total tax remittances include income taxes plus any tax penalties paid. Table A.11c shows the change in the present discounted value of of lifetime income tax remittances relative to the base policy. Table A.11d shows the change in IRA balance at age $65\frac{1}{2}$ relative to the base policy. For each panel, we show the amount in levels and as a percentage of the value in the base policy.

Table A.11: Results of changing the early withdrawal penalty

(a) Change in income at age $40\frac{1}{2}$ under base policy to reach counterfactual PDV lifetime utility

Age	5% penalty		10% penalty		20% penalty		30% penalty	
threshold	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$55\frac{1}{2}$	3,427	0.0169%	1,408	0.0069%	-1,600	-0.0079%	-3,806	-0.0187%
$56\frac{1}{2}$	3,246	0.0160%	1,092	0.0054%	-2,140	-0.0105%	-4,472	-0.0220%
$57\frac{1}{2}$	3,076	0.0151%	749	0.0037%	-2,685	-0.0132%	-5,205	-0.0256%
$58\frac{1}{2}$	2,884	0.0142%	390	0.0019%	-3,302	-0.0162%	-5,964	-0.0293%
$59\frac{1}{2}$	2,672	0.0131%	0	0.0000%	-3,892	-0.0191%	-6,753	-0.0332%
$60\frac{1}{2}$	2,462	0.0121%	-365	-0.0018%	-4,521	-0.0222%	-7,547	-0.0371%
$61\frac{1}{2}$	2,237	0.0110%	-724	-0.0036%	-5,205	-0.0256%	-8,404	-0.0413%
$62\frac{1}{2}$	2,004	0.0099%	-1,164	-0.0057%	-5,921	-0.0291%	-9,353	-0.0460%
$63\frac{1}{2}$	1,862	0.0092%	-1,488	-0.0073%	-6,445	-0.0317%	-10,034	-0.0494%
$64\frac{1}{2}$	1,684	0.0083%	-1,817	-0.0089%	-7,026	-0.0346%	-10,716	-0.0527%
$65\frac{1}{2}$	1,486	0.0073%	-2,148	-0.0106%	-7,665	-0.0377%	-11,519	-0.0567%

Notes: Each iteration includes 10,000 unique simulated individuals. This table compares the base policy for the early withdrawal policy (age threshold = $59\frac{1}{2}$, penalty rate = 10%) and a range of counterfactual policies. Table A.11a shows the average amount of income that our simulated individuals would need to receive (give up) in the base policy to reach the present discounted value of lifetime utility for each counterfactual policy. Table A.11b shows the change in the present discounted value of lifetime total tax remittances. Table A.11c shows the change in the present discounted value of lifetime income tax remittances. Table A.11d shows the change in IRA balances at age $65\frac{1}{2}$. For each panel, we show the amount in levels and as a percentage of the value in the base policy.

Table A.11: Results of changing the early withdrawal penalty, continued

(b) Change in PDV lifetime total tax remittances relative to base policy

Age threshold	5% penalty		10% penalty		20% penalty		30% penalty	
	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$55\frac{1}{2}$	-385	-0.1210%	-167	-0.0525%	20	0.0062%	144	0.0453%
$56\frac{1}{2}$	-368	-0.1159%	-160	-0.0505%	-7	-0.0021%	94	0.0297%
$57\frac{1}{2}$	-341	-0.1072%	-132	-0.0415%	-20	-0.0062%	33	0.0105%
$58\frac{1}{2}$	-301	-0.0948%	-73	-0.0231%	4	0.0014%	-4	-0.0012%
$59\frac{1}{2}$	-250	-0.0787%	0	0.0000%	78	0.0245%	-3	-0.0009%
$60\frac{1}{2}$	-179	-0.0563%	100	0.0316%	199	0.0625%	70	0.0221%
$61\frac{1}{2}$	-84	-0.0264%	244	0.0769%	369	0.1162%	225	0.0709%
$62\frac{1}{2}$	41	0.0130%	432	0.1360%	620	0.1951%	449	0.1413%
$63\frac{1}{2}$	144	0.0453%	597	0.1880%	844	0.2657%	669	0.2104%
$64\frac{1}{2}$	268	0.0845%	801	0.2520%	1,138	0.3580%	937	0.2948%
$65\frac{1}{2}$	412	0.1295%	1,057	0.3326%	1,512	0.4756%	1,300	0.4092%

(c) Change in PDV lifetime income tax remittances relative to base policy

Age threshold	5% penalty		10% penalty		20% penalty		30% penalty	
	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$55\frac{1}{2}$	521	0.1649%	379	0.1199%	129	0.0407%	28	0.0088%
$56\frac{1}{2}$	470	0.1489%	274	0.0867%	-51	-0.0161%	-188	-0.0594%
$57\frac{1}{2}$	424	0.1342%	178	0.0564%	-236	-0.0747%	-429	-0.1358%
$58\frac{1}{2}$	379	0.1199%	89	0.0280%	-422	-0.1334%	-684	-0.2167%
$59\frac{1}{2}$	332	0.1052%	0	0.0000%	-586	-0.1856%	-933	-0.2952%
$60\frac{1}{2}$	292	0.0924%	-90	-0.0284%	-746	-0.2363%	-1,155	-0.3655%
$61\frac{1}{2}$	254	0.0804%	-175	-0.0554%	-903	-0.2859%	-1,363	-0.4316%
$62\frac{1}{2}$	216	0.0683%	-262	-0.0830%	-1,064	-0.3367%	-1,579	-0.4997%
$63\frac{1}{2}$	182	0.0576%	-331	-0.1048%	-1,192	-0.3772%	-1,738	-0.5501%
$64\frac{1}{2}$	150	0.0476%	-408	-0.1293%	-1,325	-0.4196%	-1,939	-0.6139%
$65\frac{1}{2}$	108	0.0343%	-486	-0.1539%	-1,464	-0.4635%	-2,153	-0.6814%

(d) IRA balance at age $65\frac{1}{2}$ relative to base policy

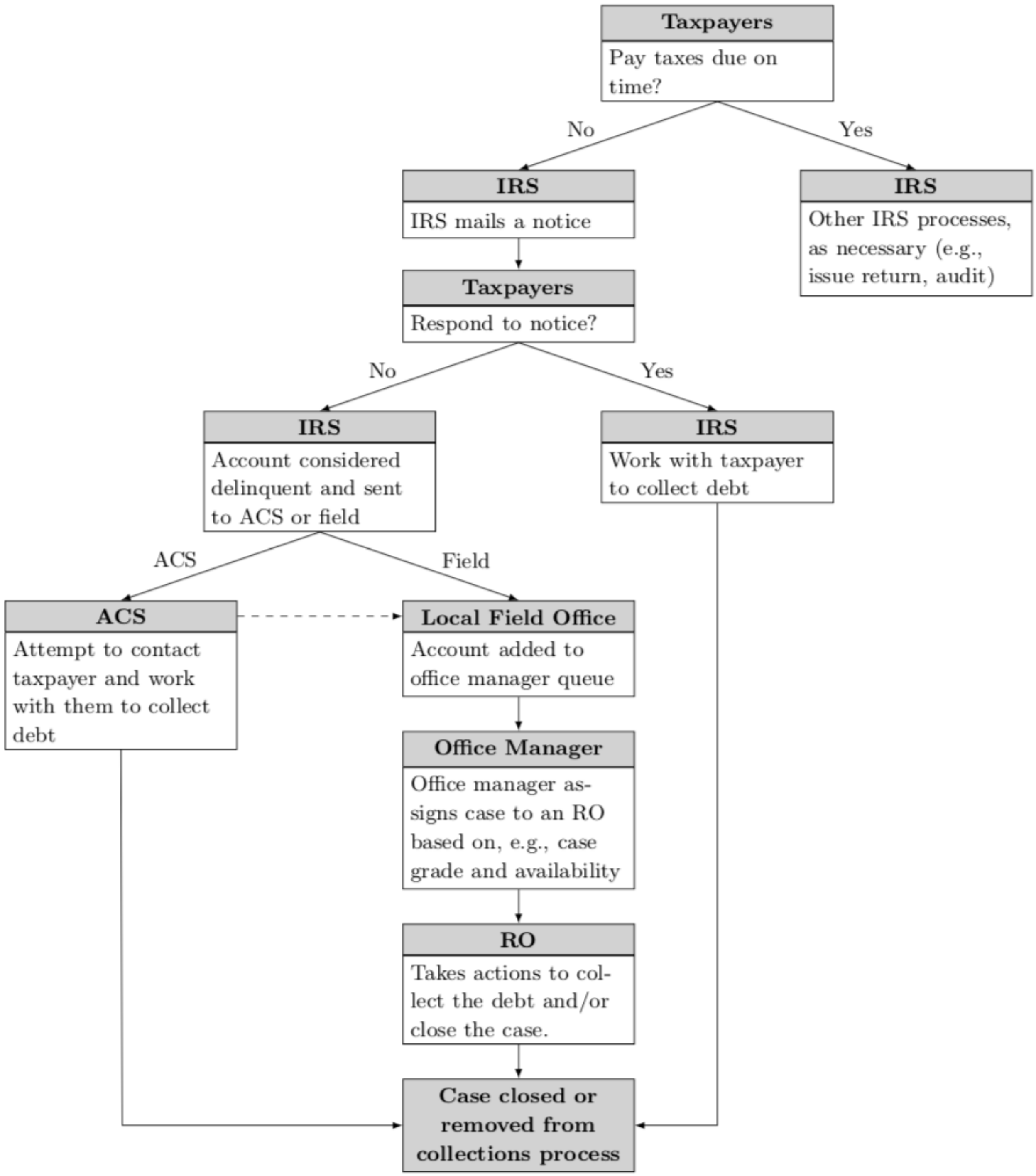
Age threshold	5% penalty		10% penalty		20% penalty		30% penalty	
	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent	Levels ($)	Percent
$55\frac{1}{2}$	-186	-0.1912%	-40	-0.0413%	31	0.0322%	183	0.1875%
$56\frac{1}{2}$	-209	-0.2146%	-106	-0.1084%	-83	-0.0847%	68	0.0695%
$57\frac{1}{2}$	-221	-0.2271%	-121	-0.1247%	-177	-0.1816%	-41	-0.0426%
$58\frac{1}{2}$	-202	-0.2072%	-106	-0.1084%	-192	-0.1969%	-122	-0.1257%
$59\frac{1}{2}$	-138	-0.1421%	0	0.0000%	-121	-0.1246%	-104	-0.1071%
$60\frac{1}{2}$	-39	-0.0396%	143	0.1463%	106	0.1090%	161	0.1653%
$61\frac{1}{2}$	135	0.1386%	384	0.3942%	520	0.5335%	655	0.6722%
$62\frac{1}{2}$	322	0.3304%	748	0.7674%	1,148	1.1781%	1,498	1.5375%
$63\frac{1}{2}$	484	0.4969%	1,126	1.1561%	1,804	1.8519%	2,476	2.5419%
$64\frac{1}{2}$	712	0.7311%	1,572	1.6139%	2,666	2.7370%	3,646	3.7424%
$65\frac{1}{2}$	1,096	1.1249%	2,279	2.3391%	3,931	4.0346%	5,309	5.4494%

Notes: Each iteration includes 10,000 unique simulated individuals. This table compares the base policy for the early withdrawal policy (age threshold = $59\frac{1}{2}$, penalty rate = 10%) and a range of counterfactual policies. Table A.11a shows the average amount of income that our simulated individuals would need to receive (give up) in the base policy to reach the present discounted value of lifetime utility for each counterfactual policy. Table A.11b shows the change in the present discounted value of lifetime total tax remittances. Table A.11c shows the change in the present discounted value of lifetime income tax remittances. Table A.11d shows the change in IRA balances at age $65\frac{1}{2}$. For each panel, we show the amount in levels and as a percentage of the value in the base policy.

APPENDIX B

Appendix to Chapter 2

B.1 Overview of the collections process



B.2 Data dictionary and sample statistics

B.2.1 Acronyms

ALE Allowable living expenses
ACS Automated Collection System
AGI Adjusted gross income
CNC Currently not collectible
EAP Estimated ability to pay
GS General schedule
IRM Internal Revenue Manual
IRS Internal Revenue Service
TPI Total positive income

B.2.2 Variable definitions

B.2.2.1 Variables included in the residualization process

Allowable living expenses The allowable living expenses data is provided quarterly at the county level by the IRS. We used the "housing and utilities" standard. We used the value for a "Family of 1" unless we had evidence that the taxpayer was married, in which case we used the value for a "Family of 2." When a zip code covered multiple counties, we took the average of the allowable living expenses given for those counties. Because we conduct our analysis at the annual level, we use the values from quarter 4. The zip-to-county conversion was done using data provided by the U.S. Department of Housing and Urban Development. We matched this to the administrative tax data by zip code.

Case grade Provided in the field data. Cases have a grade of 9, 11, 12, or 13. The assigned grade reflects the expected difficulty of closing the case.

Estimated ability to pay We develop an "estimated ability to pay" metric which is equal to a taxpayers average AGI for the three years before their case was assigned to a Revenue Officer, multiplied by the lesser of 10 and the number of years before the taxpayer turns 65, divided by the outstanding balance on the account when the taxpayer is assigned to a Revenue Officer. When AGI is missing for some year, we impute the filing threshold. Larger values suggest that the taxpayer would have greater income and therefore be more able to resolve their outstanding debt.

Group Provided in the Revenue Officer data. This is the group to which the Revenue Officer belongs. A Revenue Officer is assigned cases by the group manager.

High priority indicator Cases are assigned priority codes. The relationship between priority code and priority level is described in IRM 1.4.50.8.4 1. High priority cases include priority codes 99-108, with priority 99 and 100 cases being the highest priority cases. Medium priority cases include priority codes 201-208. Low priority cases include priority codes 301-303. This indicator is equal to 1 for cases with priority code 99 or 100.

Oldest debt more than 12 months old indicator This indicator is equal to 1 if the oldest debt on the case when it is assigned to field is older than 12 months old.

Oldest debt more than 36 months old indicator This indicator is equal to 1 if the oldest debt on the case when it is assigned to field is older than 36 months old.

Previously assigned to field indicator This indicator is equal to 1 if the taxpayer had modules assigned to the field before the case considered in the project. We consider debt starting in 2009 when constructing this variable.

Revenue Officer GS grade Provided in the Revenue Officer data. Revenue Officers may have a GS grade of 4, 5, 7, 9, 11, 12, or 13. A grade of 4 indicates a group manager. Grades 5 and 7 are training grades. As of 2017, IRS employees with a GS grade of 9 could no longer serve as Revenue Officers. Because our analysis focuses on cases closed before 2017, we observe Revenue Officers with GS grades 9, 11, 12, and 13.

Urban indicator The urban dummy is based on data provided by the U.S. Department of Agriculture Economic Research Service (USDA ERS). The data include the Rural Urban Continuum Code (RUCC) by county. We designate a location as not urban if the RUCC for the zip code is 7 or 9, which includes areas with populations less than 20,000 that are non-adjacent to metro areas. We match the county to zip code using the 2014 Q4 zip-to-county data provided by the U.S. Department of Housing and Urban Development. For a few zip codes not included in the data from USDA ERS we use population-by-zip code data from the 2010 Census. We categorize these zip codes as urban if the population is greater than 20,000. We matched this to the administrative tax data by zip code.

Year of birth The year of birth of the taxpayer. We censor at both ends: we set the year of birth equal to 1930 for taxpayers with year of birth older than 1930, and we set the year of birth equal to 1997 for taxpayers with year of birth earlier than 1997.

B.2.2.2 Outcome variables

Adjusted gross income We use the value of Adjusted Gross Income provided on a taxpayer's Form 1040 when available. When this value is blank, we use the sum of income from all third party information reporting sources. When this second value is not available, we use the sum of all unique W-2s received for the taxpayer.

Filing conditional on Form W-2 (indicator) This indicator is equal to 1 if the taxpayer filed a return conditional on whether or not the IRS received a W-2 for the taxpayer.

Fully paid new tax liability (indicator) This indicator is equal to 1 if the taxpayer fully paid any new tax liability in the tax year after their case was closed.

Payments to the IRS Calculated as the sum of all payments remitted to the IRS against tax debt from the current tax year, plus all payments remitted to the IRS during the current tax year against debt from previous tax years.

Self employed income The total amount of reported Schedule C income (self-employed business) income (or loss) from all Schedule C (or C-EZ) attached to the taxpayer's tax return.

W-2 wages We use the sum of all unique W-2s received for the taxpayer. We present this variable at the household, individual, and spouse level.

B.2.2.3 Balance variables

Model score An estimate generated by the IRS of the probability of repayment, with lower scores indicating a lower estimated probability of collection. This variable has 25^{th}, 50^{th}, and 75^{th} percentile values of approximately 0.05, 0.09, and 0.2, respectively.

Average W-2 wages The average of wages reported through W-2s for the three years prior to assignment to a Revenue Officer.

Average AGI The average of Adjusted Gross Income (as defined above) for the three years prior to assignment to a Revenue Officer.

AGI before assignment The value of Adjusted Gross Income (as defined above) for the year prior to assignment to a Revenue Officer.

Filed before assignment This indicator is equal to 1 if the taxpayer filed a tax return in the year prior to assignment to a Revenue Officer.

B.2.3 Sample statistics

Table B.1 shows the mean and standard deviation of the variables used in our analysis in Year 0 by CNC designation.

Table B.1: Sample statistics (Year 0)

	Non-CNC	CNC	Total
Observations	86,083	37,313	123,396
Residualization variables			
High priority indicator	0.216	0.219	0.217
	(0.411)	(0.414)	(0.412)
Case grade = 9	0.142	0.142	0.142
	(0.349)	(0.349)	(0.349)
Case grade = 11	0.480	0.538	0.497
	(0.500)	(0.499)	(0.500)
Case grade = 12	0.334	0.306	0.325
	(0.471)	(0.461)	(0.468)
Case grade = 13	0.045	0.014	0.036
	(0.207)	(0.118)	(0.185)
Year of birth	1963	1961	1962
	(11)	(12)	(11)
Estimated ability to pay	54.566	15.733	42.824
	(148.502)	(65.649)	(130.406)
Previously assigned	0.413	0.425	0.417
	(0.818)	(0.806)	(0.814)
Oldest debt > 12 mo.	0.713	0.755	0.726
	(0.452)	(0.430)	(0.446)
Oldest debt > 36 mo.	0.598	0.701	0.629
	(0.490)	(0.458)	(0.483)
Allowable living expenses	1,860	1,731	1,820
	(585)	(524)	(570)
Urban indicator	0.978	0.973	0.976
	(0.147)	(0.163)	(0.152)
RO Inventory	54.330	54.881	54.497
	(16.518)	(15.883)	(16.331)
RO GS grade = 9	0.150	0.137	0.146
	(0.357)	(0.344)	(0.353)
RO GS grade = 11	0.429	0.482	0.445
	(0.495)	(0.500)	(0.497)
RO GS grade = 12	0.390	0.368	0.383
	(0.488)	(0.482)	(0.486)
RO GS grade = 13	0.031	0.014	0.026
	(0.174)	(0.116)	(0.159)

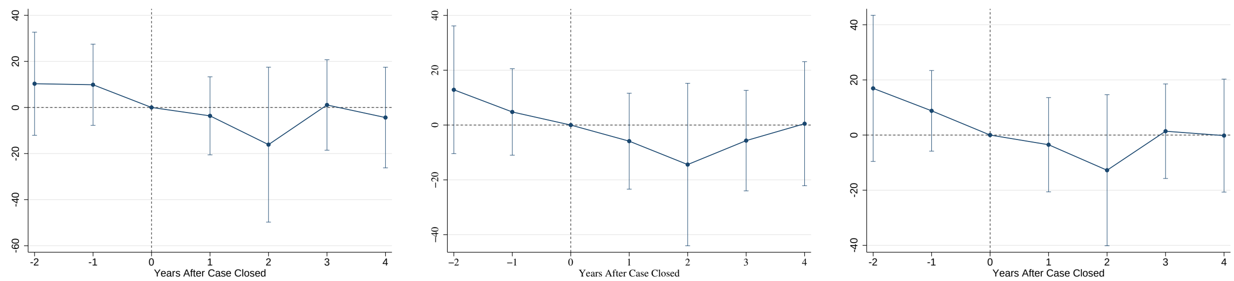
Table B.1: Sample statistics (Year 0, continued)

	Non-CNC	CNC	Total
Observations	86,083	37,313	123,396
Outcome variables			
Payments, HH (total debt)	14,047	4,573	11,182
	(146,792)	(40,091)	(124,648)
W-2 withholding, HH	6,994	1,642	5,376
	(133,826)	(9,955)	(111,936)
QES payments, HH	276	41	205
	(20,040)	(1,810)	(16,768)
Remittances, HH	-0	0	-0
	(4)	(0)	(3)
Fully paid new debt	0.507	0.475	0.499
	(0.500)	(0.499)	(0.500)
Filed | W-2s	0.793	0.668	0.755
	(0.405)	(0.471)	(0.430)
AGI, HH	114,976	59,872	98,314
	(183,863)	(105,126)	(166,029)
W-2 wages, HH	63,925	36,810	57,061
	(90,438)	(55,207)	(83,780)
W-2 wages	81,609	48,491	73,269
	(99,555)	(62,130)	(92,704)
W-2 wages, spouse	11,801	6,047	10,165
	(30,817)	(19,626)	(28,212)
Sch. C profit/loss | filed	28,437	23,103	27,010
	(66,492)	(49,872)	(62,523)
Balance variables			
Model score	0.170	0.105	0.150
	(0.162)	(0.098)	(0.148)
Pre-Ave. W-2 wages, HH	62,430	22,431	50,335
	(485,752)	(51,761)	(407,128)
Pre-Ave. AGI, HH	170,507	54,989	135,577
	(2,699,416)	(1,365,700)	(2,377,013)
AGI year before assignment, HH	209,803	50,231	161,551
	(5,323,174)	(330,781)	(4,450,411)
Filed before assignment	0.831	0.671	0.783
	(0.375)	(0.470)	(0.412)
Instruments			
Simple LOO instrument	0.290	0.331	0.302
	(0.108)	(0.115)	(0.112)
Residualized instrument	-0.010	0.022	0.000
	(0.079)	(0.084)	(0.082)

Notes: Includes all cases that meet our sample restriction criteria between November 2014 and December 2018. Limited to cases worked by Revenue Officers who closed at least 20 cases that met our sample restriction criteria between November 2014 and December 2018. All monetary values adjusted for inflation to 2017 values.

Figure B.1: Effect of a CNC designation on payments toward outstanding tax debt ($1,000), IV with residualized instrument

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. Payment values adjusted for inflation to 2017 values.

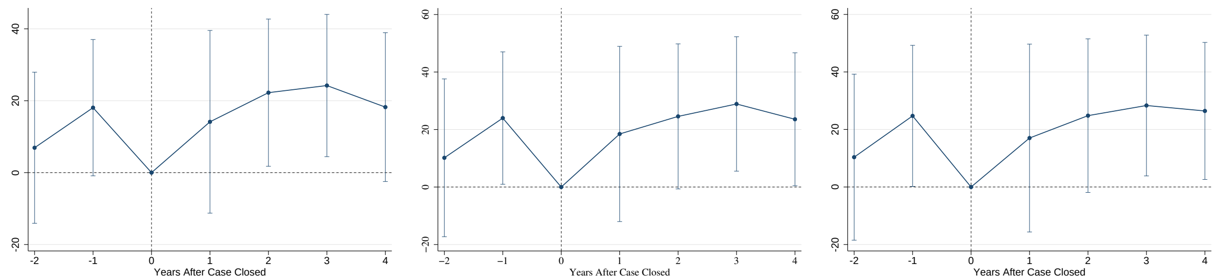
B.3 Robustness to choice of case count cutoff

Figures B.1 through B.8 show our residualized IV results using the alternative Revenue Officer case count cut-offs of 10, 20, and 30.

Figure B.2: Effect of a CNC designation on payments toward current tax liability ($1,000), IV with residualized instrument

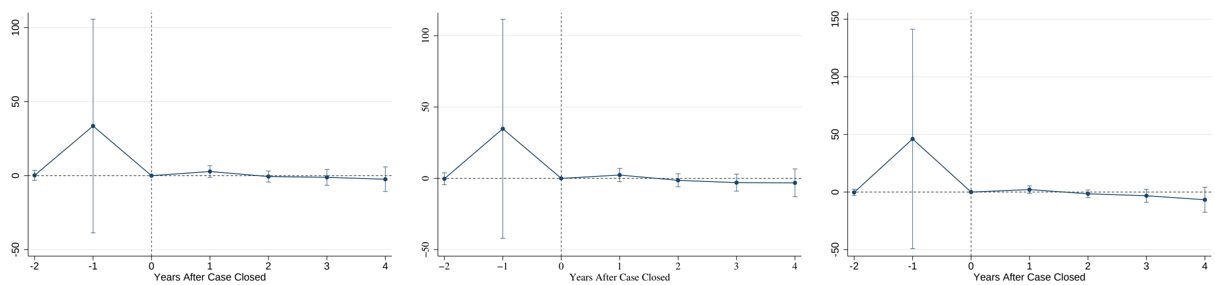
W-2 Withholding

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



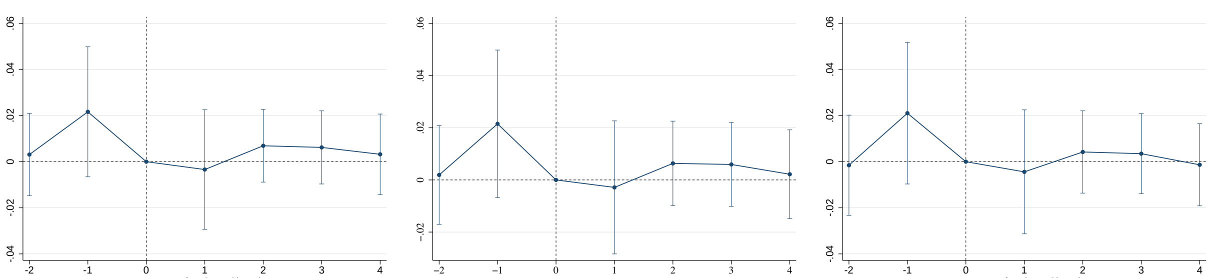
Quarterly Estimated Tax Payments

(d) 10+ Cases (e) 20+ Cases (f) 30+ Cases



Remittances with Returns

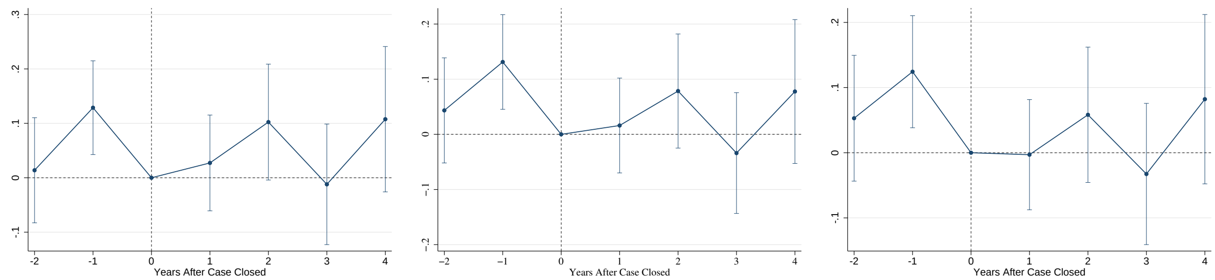
(g) 10+ Cases (h) 20+ Cases (i) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. All payment variables adjusted for inflation to 2017 values.

Figure B.3: Effect of a CNC designation on whether or not the taxpayer remitted their entire current tax liability, IV with residualized instrument

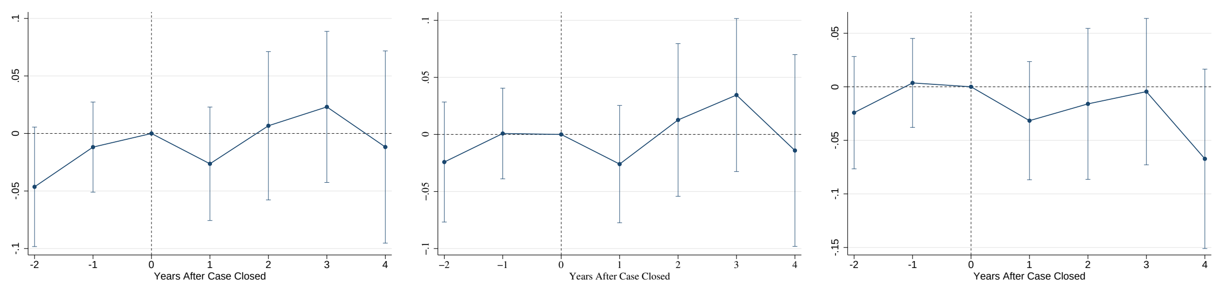
(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

Figure B.4: Effect of a CNC designation on filing a tax return, IV with residualized instrument

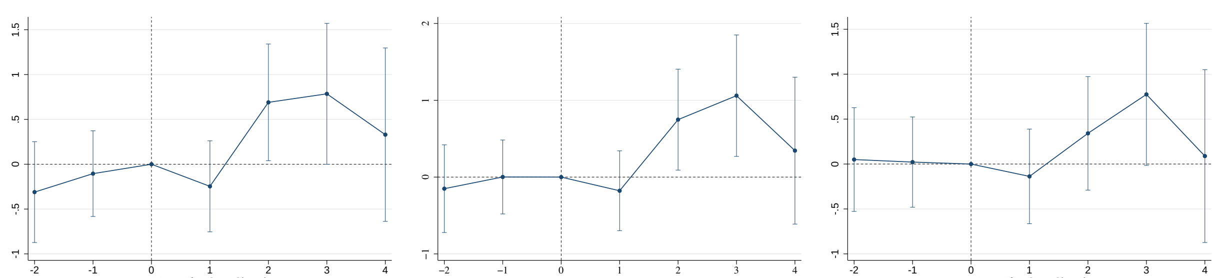
(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

Figure B.5: Effect of a CNC designation on adjusted gross income (IHS), IV with residualized instrument

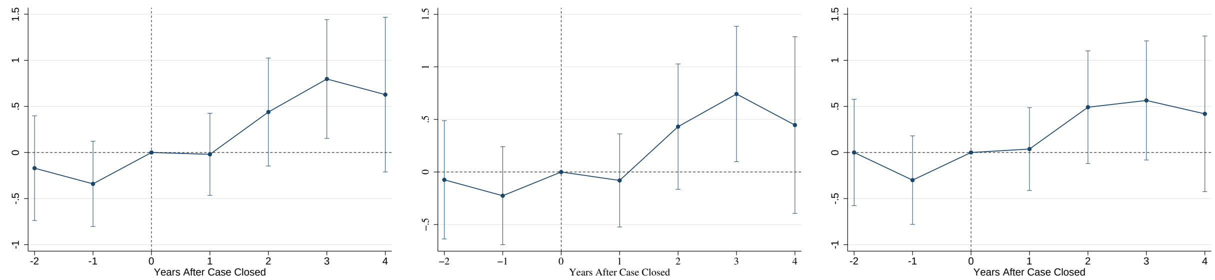
(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Adjusted gross income adjusted for inflation to 2017 values.

Figure B.6: Effect of a CNC designation on W-2 wages (IHS), IV with residualized instrument

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases

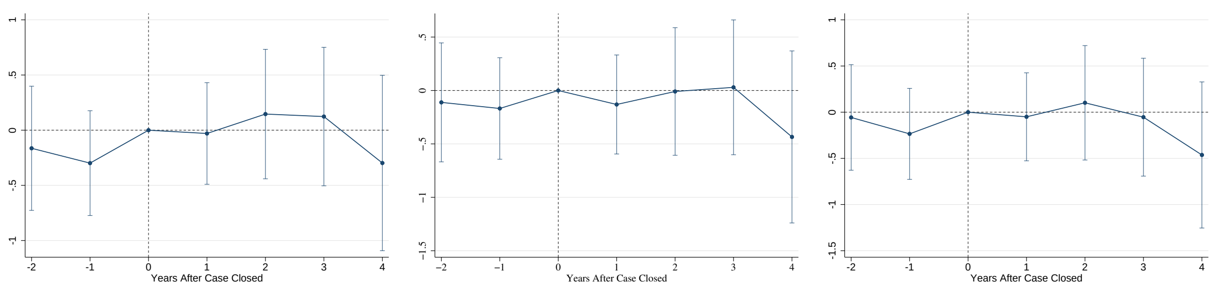


Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. W-2 income adjusted for inflation to 2017 values.

Figure B.7: Effect of a CNC designation on W-2 Wages (IHS): Taxpayer vs. Spouse

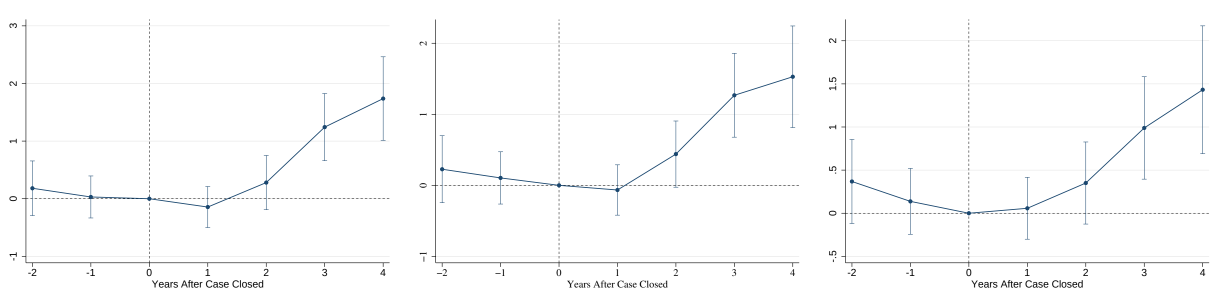
Taxpayer whose case was designated CNC

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Spouse

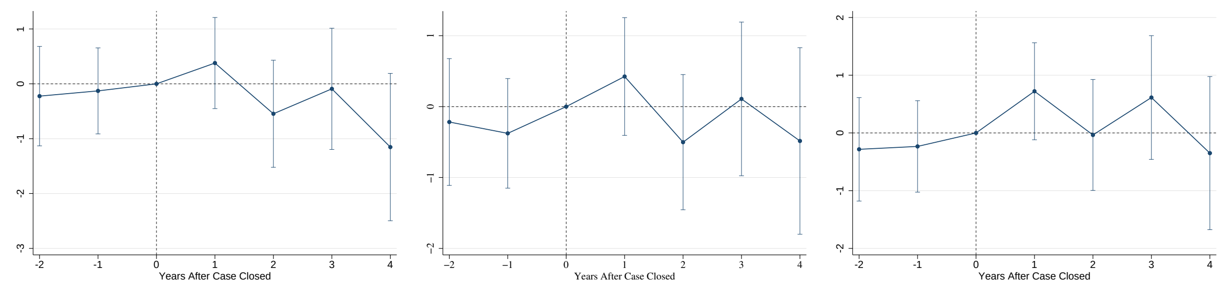
(d) 10+ Cases (e) 20+ Cases (f) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Spouses include individuals married to taxpayers in the year in which their case was closed. W-2 income adjusted for inflation to 2017 values.

Figure B.8: Effect of a CNC designation on self-employment income (IHS), IV with residualized instrument

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Self employment income adjusted for inflation to 2017 values.

APPENDIX C

Appendix to Chapter 3

C.1 Glossary

This glossary serves as a guide to the acronyms and terminology used in this project.

C.1.1 Acronyms

AGI Adjusted Gross Income

CPA Certified Public Accountant

IRS Internal Revenue Service

NRP National Research Program

PTIN Preparer Tax Identification Number

SRTP Supervised Registered Tax Preparer

VITA program Volunteer Income Tax Assistance program

C.1.2 Terminology

Enrolled Agent An Enrolled Agent is a tax advisor who has been federally authorized by the U.S. Department of the Treasury. Unlike CPAs and attorneys, which are licensed to practice on a state-by-state basis, the "Enrolled Agent" credential is recognized in all 50 states.

NRP audit A random, full-return audit performed as part of the IRS's National Research Program.

Operational audit A non-random audit of a specific portion of a tax return.

Paid-prepared return A return that was prepared by a paid tax professional. A taxpayer might get assistance from four types of tax preparers: Certified Public Accountants (CPAs), Enrolled Agents, tax attorneys, and non-credentialed preparers. While the IRS does not currently require tax preparers to be licensed or receive continuing education, preparers are required to register with the IRS and obtain a Preparer Tax Identification Number (PTIN) to receive payment for assisting with tax returns. Throughout the paper, a "paid-prepared" return refers to returns completed by the professionals who have a PTIN.

Self-prepared return A return that was not prepared by a paid tax professional or through the VITA program. Could have been prepared by hand or using tax-preparation software.

Supervised Registered Tax Preparer SRTPs do not themselves sign tax returns (nor are they required to), but work for an organization that is owned at least 80% by CPAs, attorneys, or Enrolled Agents and who is supervised by a CPA, attorney, or Enrolled Agent who has a PTIN.

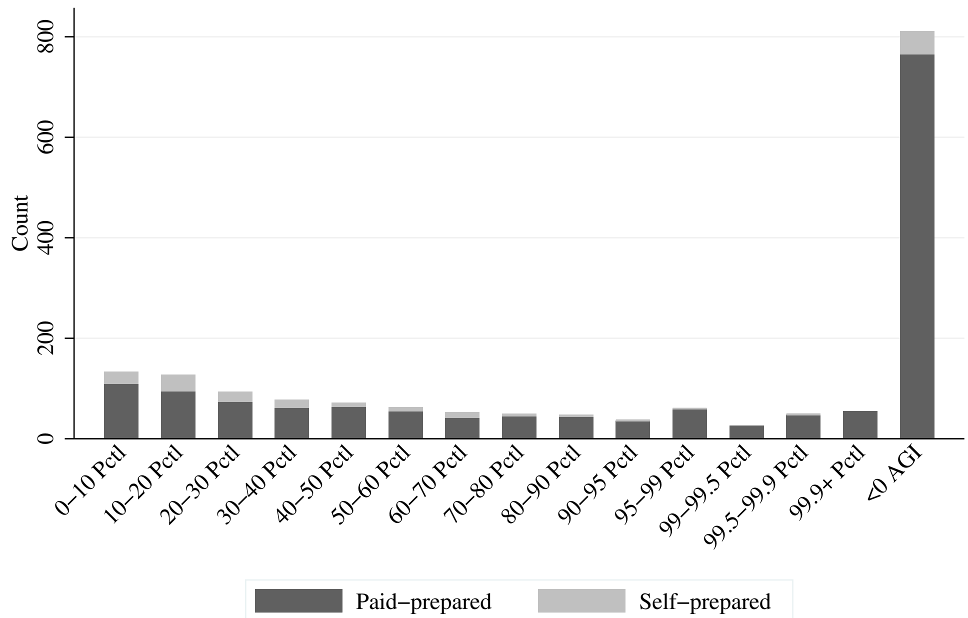
C.2 Additional and alternative figures

C.2.1 Individuals with negative AGI

In this section, we undertake one exercise with the individuals in our negative AGI group in an attempt to tease out the extent to which these individuals looked like other parts of the income distribution. We ask the question: if we took these individuals' AGI for our sample period (2006-2014), scaled everything up to 2014 dollars, and computed the average 2014-adjusted AGI for each individual, what AGI percentile would they fall into?

The results of this exercise are given in Figure C.1. We observe that the overwhelming majority of these individuals have a negative average AGI over the sample period, with the remaining individuals scattered across the rest of the income distribution.

Figure C.1: AGI percentile for average AGI if had negative corrected AGI in NRP



Notes: N = 1,757. See Table 3.2 for counts by subgroup.

C.2.2 Tax return preparation methods over time

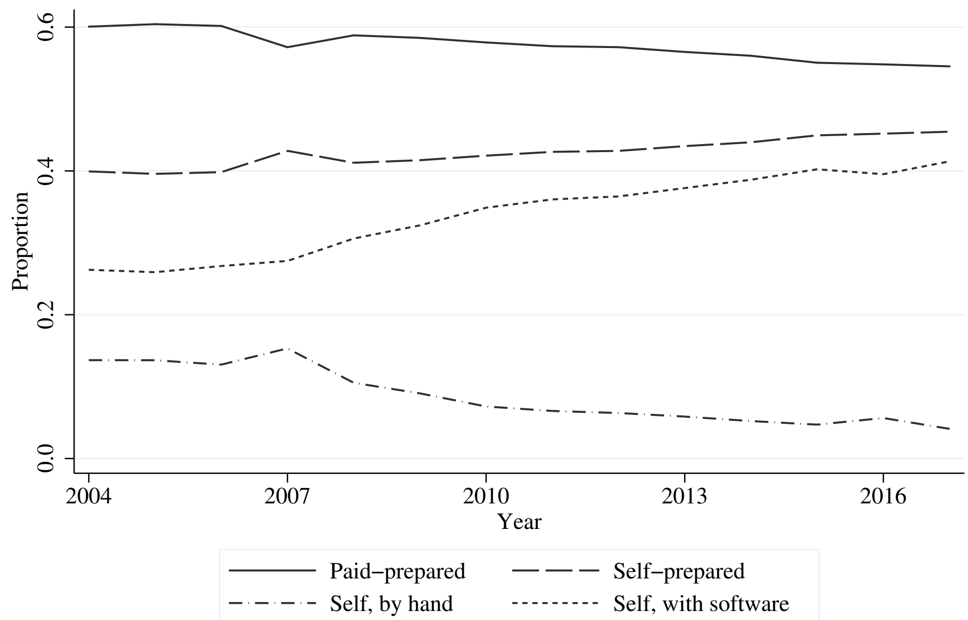
Figure C.2 shows the proportion of returns (excluding VITA-prepared returns) that were prepared by a paid professional or self-prepared over the period from 2004-2017. The figure also breaks down self-prepared returns by those done by hand or using tax preparation software. Over this time period, the proportion of individuals using a paid preparer has slowly declined, while the proportion of self-prepared returns has slowly increased.

What's more striking is the share of self-prepared returns that are prepared by hand or with tax software. At the beginning of this time period, about one-third of self-prepared returns were completed by hand and about two-thirds using tax software. By the end of this period, over 90% of self-prepared returns were prepared using tax preparation software.

C.2.3 Tax preparation fees: Figure 3.9d with $0 - 10$ percentile group included

Figure C.3 is the same as Figure 3.9d in the text, but includes the $0 - 10$ corrected AGI percentile group. This figure suggests that, among taxpayers in the $0 - 10$ corrected AGI

Figure C.2: Tax return preparation methods over time (2004-2017)



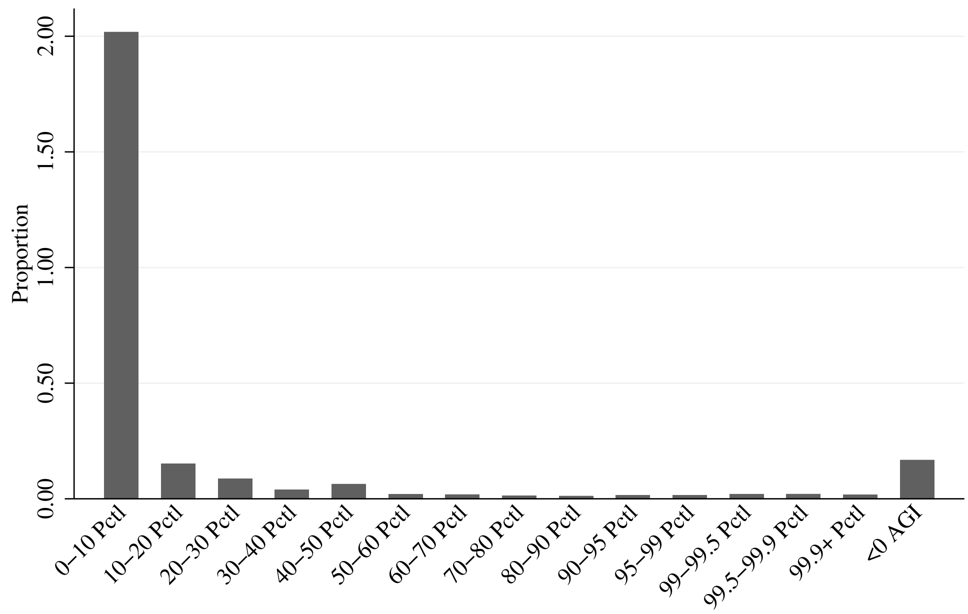
Notes: N = 1,989,052,208 individual-years. Covers 2004 - 2017.

percentile group who itemized their deductions and deducted tax preparation fees, those fees were, on average, almost two times the value of the adjusted gross income reported by those taxpayers. This statistic leads us to believe that some of the individuals included in the $0-10$ corrected AGI percentile group are actually more similar to those in the negative AGI group (in that they are likely actually high income and/or high wealth individuals), but happened to have low but non-negative AGI during the study year.

C.2.4 Number of returns by preparer: figures with alternative return count cut-offs

Figure C.4 is the same as Figure 3.10 in the body of the text, but with alternative return count cut-offs for which tax preparers were included. In the body of the text, we use a cut-off of 30,000 cases. Below, we show the figure for cut-offs of 1,000, 5,000, and 10,000 cases. The trends in these figures mirror what we observe in Figure 3.10: as corrected AGI percentile group increases, the average number of returns prepared by the preparers used by that percentile group decreases.

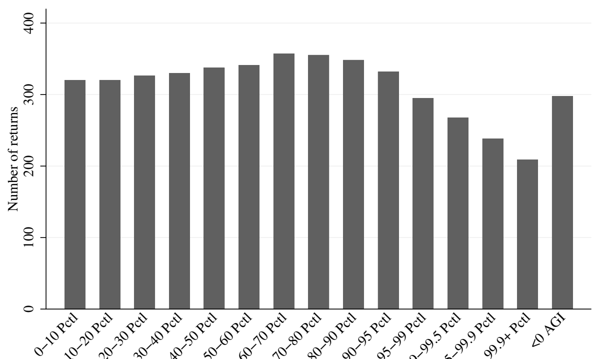
Figure C.3: Tax preparation feels as a proportion of abs(AGI)



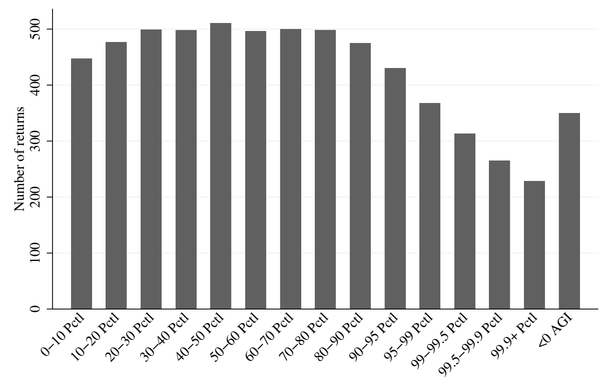
Notes: N = 25,753. See Table 3.2 for subtotals by corrected AGI percentile.

Figure C.4: Average number of returns prepared: alternative return count cut-offs

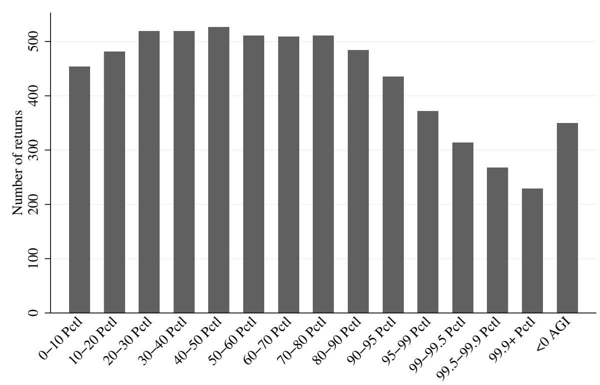
(a) Cut-off = 1,000 returns



(b) Cut-off = 5,000 returns



(c) Cut-off = 10,000 returns



Notes: Figure C.4a $N = 75{,}969$; Figure C.4b $N = 83{,}137$; Figure C.4c $N = 83{,}259$. See Table 3.2 for subtotals by corrected AGI percentile.

C.3 Proofs

Lemma 3. *Under Assumption 1, as y becomes arbitrarily large:*

- $g_1(y, p_1) - g_0(y, p_1)$ *converges to* 0,
- $g_2(y, p_2) - g_1(y, p_2)$ *converges to* 0.

Proof. We direct the reader to Guyton et al. (2021) for the proof of the first statement, which is the same as Lemma 2 in that paper.[1] To prove the second statement, we replicate the proof of Lemma 2 in Guyton et al. (2021) using our notation and with the revised Assumption 1.

The optimization problem in Equation 3.4.1 can be restated with (g, a) as the choice variables instead of (e, a). After fixing a hiring decision, a, the first-order condition with respect to g in Equation 3.4.1 is as follows:

$$\frac{u'(c_D)}{u'(c_N)} = \frac{1-p}{p\theta} \tag{C.1}$$

where $c_N = (1 - \tau + \tau g - \tilde{\kappa}_a)y$ and $c_D = (1 - \tau - \tau\theta g - \tilde{\kappa}_a)y$ denote consumption in the case of no detection and detection respectively, and $\tilde{\kappa}_a = \kappa_a/y$ denote the fixed cost to income ratio, given a.

As we are interested in $g_2(y, p_2) - g_1(y, p_2)$, we substitute the relevant consumption function into the FOC to get the condition given in Equation C.2:

$$\frac{u'\big((1 - \tau - \tau\theta g_2(y, p_2) - \tilde{\kappa}_2)y\big)}{u'\big((1 - \tau + \tau g_2(y, p_2) - \tilde{\kappa}_2)y\big)} = \frac{u'\big((1 - \tau - \tau\theta g_1(y, p_2) - \tilde{\kappa}_1)y\big)}{u'\big((1 - \tau + \tau g_1(y, p_2) - \tilde{\kappa}_1)y\big)} = \frac{1-p_2}{p_2\theta} \tag{C.2}$$

Because $u'' < 0$, u'^{-1} is uniquely defined, both sides of the first equality are invertible in e. Assumption 1 guarantees that for sufficiently large y, $g_1(y, p_2)$ approaches a strictly positive constant, whereas $\tilde{\kappa}_a$ converges to 0 as y becomes arbitrarily large.

Next, let $p_2 = p$ be fixed across g_1 and g_2 and restate the g functions as $g(y, \tilde{\kappa}_1) = g_1(y, p_2)$ and $g(y, \tilde{\kappa}_2) = g_2(y, p_2)$:

$$g(y, \tilde{\kappa}_2) = \underset{g\in[0,1]}{\arg\max}(1-p)u\big((1 - \tau + \tau g - \tilde{\kappa}_2)y\big) + pu\big((1 - \tau - \tau\theta g - \tilde{\kappa}_2)y\big) \tag{C.3}$$

$$g(y, \tilde{\kappa}_1) = \underset{g\in[0,1]}{\arg\max}(1-p)u\big((1 - \tau + \tau g - \tilde{\kappa}_1)y\big) + pu\big((1 - \tau - \tau\theta g - \tilde{\kappa}_1)y\big) \tag{C.4}$$

Because both κ_1 and κ_2 are constants, $lim_{y\to\infty}\tilde{\kappa}_a = 0$ for $a \in \{1, 2\}$, which yields $lim_{y\to\infty}(\tilde{\kappa}_2 -$

[1] Alternatively, the first statement can be proven by the following substitutions in the proof of the second statement, given below: p_1 instead of p_2, p_0 instead of p_1, κ_2 instead of κ_1 and 0 instead of κ_1.

$\tilde{\kappa}_1) = 0$. By definition of the limit, $\forall \delta > 0, \exists c \in \mathbb{R}$, such that

$$y > c \implies |\tilde{\kappa}_2 - \tilde{\kappa}_1| < \delta. \tag{C.5}$$

Continuity of g on $\mathbb{R}_{++}$ implies that for a given $\varepsilon > 0$, $\exists c \in \mathbb{R}$, such that

$$y > c \implies |\tilde{\kappa}_2 - \tilde{\kappa}_1| < \delta \implies |g(y, \tilde{\kappa}_2) - g(y, \tilde{\kappa}_1)|. \tag{C.6}$$

$g_2(y, p_2)$ converges to $g_1(y, p_2)$ as y becomes arbitrarily large. Furthermore, Assumption 1 guarantees that, as y becomes arbitrarily large, $g_1(y, p_2)$ will be arbitrarily close to a non-zero constant.

□

Proposition 2. *Under Assumption 1,*

- *There is a cutoff income level, y', such that all else equal, agents with income higher than this cutoff will prefer to hire a generalized tax preparer over self-preparation,*
- *There is a cutoff income level, y'', such that all else equal, agents with income higher than this cutoff will prefer to hire a specialized tax preparer over generalized tax preparer.*
- *$y'' > y'$.*

Proof. As Guyton et al. (2021) show, proving that for a sufficiently large, y, the difference in expected utility between $a = 1$ and $a = 0$ given optimal g_1 and g_0 is enough to prove the first item. The fact that there is a third option ($a = 2$) does not change the trade-off between actions $a = 0$ and $a = 1$. As a result, the proof of Proposition 1 in Guyton et al. (2021) can be applied directly to show the first claim.

To prove the second claim, we follow the proof in Guyton et al. (2021), replacing κ with κ_2, 0 with κ_1, p_0 with p_1, and p_1 with p_2. We want to show that for high enough y, the difference in the indirect utility between $a = 2$ and $a = 1$ given the respective optimal choices is positive. In other words, we want to show the following:

$$\Delta_{21} U = U(p_2, \kappa_2, y) - U(p_1, \kappa_1, y) > 0, \tag{C.7}$$

where the indirect utility as a function of a and g_a is defined as follows:

$$U(p, \kappa_i, y) = (1 - p)u\big((1 - \tau + \tau g(p, \kappa_i, y) - \tilde{\kappa}_i)y\big) + pu\big((1 - \tau - \tau\theta g(p, y, \kappa_i) - \tilde{\kappa}_i)y\big), \tag{C.8}$$

where $\tilde{\kappa}_i = \frac{\kappa_i}{y}$ denotes the related fixed cost as a share of income.

By leveraging Lemma 2, we can decompose the difference in Equation C.7 into the difference due to increased cost, and the difference due to decreased probability as shown in Equation C.9:

$$\Delta_{21}U = [U(p_2, \kappa_2, y) - U(p_2, \kappa_1, y)] + \{U(p_2, \kappa_1, y) - U(p_1, \kappa_1, y)\}. \tag{C.9}$$

Note that the term in square brackets is negative because, all else equal, utility decreases in κ_i and $\kappa_1 < \kappa_2$, whereas the term in the curly brackets is positive because, all else equal, expected utility decreases in p and $p_2 < p_1$.

We next show that, given the second part of Assumption 1 (i.e., that $g_1(y, p_2)$ diminishes as y becomes larger), the expression in the square brackets is dominated by the expression in the curly brackets in Equation C.9.

The term in the curly brackets can be expressed as the following by using second fundamental theorem of calculus:

$$U(p_2, \kappa_1, y) - U(p_1, \kappa_1, y) = -\int_{p_2}^{p_1} U_p(p, \kappa_1, y)dp, \tag{C.10}$$

where $U_p = \frac{\partial U}{\partial p}$. Invoking the envelope theorem gives us

$$\begin{gathered} U(p_2, \kappa_1, y) - U(p_1, \kappa_1, y) = \\ \int_{p_2}^{p_1} [u\big((1 - \tau + \tau g(p, \kappa_1, y) - \tilde{\kappa_1})y\big) - u\big((1 - \tau - \tau\theta g(p, \kappa_1, y) - \tilde{\kappa_1})y\big)]dp \end{gathered} \tag{C.11}$$

Given that a risk-averse agent will choose $g(p, \kappa_1, y) > 0$ for $p \in [0, 1)$ and $u' > 0$, the expression in Equation C.11 is positive.

Next, let $c_D(p, \kappa_i, y)$ denote the consumption in the case of detection and $c_N(p, \kappa_i, y)$ denote the consumption in case of no detection, given the optimal behavior $g(p, \kappa_i, y)$. By substituting Equation C.11, we can rewrite Equation C.9 as the following:

$$\Delta_{21}U = \big[U(p_2, \kappa_2, y) - U(p_2, \kappa_1, y)\big] + \int_{p_2}^{p_1} \big[u\big((c_N(p, \kappa_1, y)\big) - u\big(c_D(p, \kappa_1, y)\big)\big]dp \tag{C.12}$$

Substituting the definition of U as given in Equation C.8 into Equation C.12 yields:

$$\begin{aligned} \Delta_{21}U = (1 - p_2)\big[u\big(c_N(p_2, \kappa_2, y)\big) - u\big(c_N(p_2, \kappa_1, y)\big)\big] \\ + p_2\big[u\big(c_D(p_2, \kappa_2, y)\big) - u\big(c_D(p_2, \kappa_1, y)\big)\big] \\ + \int_{p_2}^{p_1} \big[u\big((c_N(p, \kappa_1, y)\big) - u\big(c_D(p, \kappa_1, y)\big)\big]dp \end{aligned} \tag{C.13}$$

Applying the second fundamental theorem of calculus again for the terms in square brackets, we get,

$$
\begin{aligned}
\Delta_{21}U = -(1-p_2)&\int_{c_N(p_2,\kappa_2,y)}^{c_N(p_2,\kappa_1,y)} u'(c)dc \\
-p_2&\int_{c_D(p_2,\kappa_2,y)}^{c_D(p_2,\kappa_1,y)} u'(c)dc \\
+\int_{p_2}^{p_1}&\int_{c_D(p,\kappa_1,y)}^{c_N(p,\kappa_1,y)} u'(c)dcdp
\end{aligned}
\tag{C.14}
$$

Note that, both $c_N(p,\kappa_i,y)$ and $c_D(p,\kappa_i,y)$ decreases in κ_i and $c_N(.) > c_D(.)$, therefore, all integrals in the previous expression are well defined.

Next, we construct a lower bound for $\Delta_{21}U$ by evaluating the integrals in the previous equation:

$$
\begin{aligned}
\Delta_{21}U > &-(1-p_2)\big[c_N(p_2,\kappa_1,y) - c_N(p_2,\kappa_2,y)\big]u'(c_N(p_2,\kappa_2,y)) \\
&- p_2\big[c_D(p_2,\kappa_1,y) - c_D(p_2,\kappa_2,y)\big]u'(c_D(p_2,\kappa_2,y)) \\
&+ \int_{p_2}^{p_1}\big[c_N(p,\kappa_1,y) - c_N(p,\kappa_1,y)\big]u'(c_D(p,\kappa_1,y))dp
\end{aligned}
\tag{C.15}
$$

The first and second terms in Equation C.15 are the integrals in the first and second terms in Equation C.14 respectively, evaluated at their lower limits of integration. Similarly, the third term in Equation C.15 is the integral in the last term in Equation C.14 evaluated at its upper limit. Because $u'' > 0$, u' decreases with consumption, the inequality in Equation C.15 holds strictly.

To further simplify the lower bound, we substitute the following first-order condition from the agent's problem:

$$
u'(c_D(p_2,\kappa_2,y)) = u'(c_N(p_2,\kappa_2,y))\frac{1-p_2}{\theta p_2}
\tag{C.16}
$$

Substituting this into the right-hand-side of inequality C.15 yields the following:

$$
\begin{aligned}
&\Delta_{21}U > \\
&-(1-p_2)u'\big(c_N(p_2,\kappa_2,y)\big)\left(c_N(p_2,\kappa_1,y) - c_N(p_2,\kappa_2,y) + \frac{1}{\theta}\big[c_D(p_2,\kappa_1,y) - c_D(p_2,\kappa_2,y)\big]\right) \\
&+ \int_{p_2}^{p_1}\big[c_N(p,\kappa_1,y) - c_N(p,\kappa_1,y)\big]u'\big(c_N(p,\kappa_1,y)\big)dp \equiv \bar{L}(y)
\end{aligned}
\tag{C.17}
$$

As c_N is decreasing in p and $u'' < 0$, evaluating the last term with a constant marginal utility,

$u'\big(c_N(p_2,\kappa_1,y)\big)$, gives us the following simplified version of the lower bound:

$$\begin{aligned}
&\bar{L}(y) > \\
&-(1-p_2)u'\big(c_N(p_2,\kappa_2,y)\big)\left(c_N(p_2,\kappa_1,y)-c_N(p_2,\kappa_2,y)+\frac{1}{\theta}\big[c_D(p_2,\kappa_1,y)-c_D(p_2,\kappa_2,y)\big]\right)\\
&+u'\big(c_N(p_2,\kappa_1,y)\big)\int_{p_2}^{p_1}\big[c_N(p,\kappa_1,y)-c_D(p,\kappa_1,y)\big]dp \equiv f_{21}(y)
\end{aligned} \tag{C.18}$$

Lemma 1 implies that $c_N(p,\kappa_1,y)-c_D(p,\kappa_1,y)=\tau(1+\theta)g(p,\kappa_1,y)y$ is decreasing in p. We can therefore shrink the expression in the previous equation further by evaluating it at the upper limit of integration. This final simplification provides us with the following inequality, where we define the final bound as follows:

$$\begin{aligned}
\Delta_{21}U > f_{21}(y) = &-(1-p_2)u'\big(c_N(p_2,\kappa_2,y)\big)\Big(c_N(p_2,\kappa_1,y)-c_N(p_2,\kappa_2,y)\\
&+\frac{1}{\theta}\big[c_D(p_2,\kappa_1,y)-c_D(p_2,\kappa_2,y)\big]\Big)\\
&+u'\big(c_N(p_2,\kappa_1,y)\big)(p_1-p_2)\tau(1+\theta)g(p_1,\kappa_1,y)y
\end{aligned} \tag{C.19}$$

Because $u'>0$, dividing everywhere by $u'\big(c_N(p_2,\kappa_2,y)\big)$ gives us

$$\begin{aligned}
f_{21}(y)>0 \iff &-(1-p_2)\left(c_N(p_2,\kappa_1,y)-c_N(p_2,\kappa_2,y)+\frac{1}{\theta}\big[c_D(p_2,\kappa_1,y)-c_D(p_2,\kappa_2,y)\big]\right)\\
&+\frac{u'\big(c_N(p_2,\kappa_1,y)\big)}{u'\big(c_N(p_2,\kappa_2,y)\big)}(p_1-p_2)\tau(1+\theta)g(p_1,\kappa_1,y)y>0
\end{aligned} \tag{C.20}$$

Now consider high levels of income, y. Recall that the second part of Lemma 2 implies that as income becomes arbitrarily large, the difference in fixed cost $\kappa_2-\kappa_1$ becomes negligible, which in turn implies that both $c_N(p_2,\kappa_1,y)-c_N(p_2,\kappa_2,y)$ and $c_D(p_2,\kappa_1,y)-c_D(p_2,\kappa_2,y)$ also become negligible. Therefore, the term in the top row of Equation C.20 can be made arbitrarily small. Another implication of Lemma 2 is that the marginal utility ratio $\frac{u'(c_N(p_2,\kappa_1,y))}{u'(c_N(p_2,\kappa_2,y))}$ converges to unity as y grows. Finally, the second part of Assumption 1 guarantees that $\tau(1+\theta)g(p,\kappa_1,y)y$ becomes arbitrarily large as y grows. Hence, $f_{21}(y)>0$ for sufficiently high y, which in turn implies that $\Delta_{21}U>0$ also for sufficiently large y.

To prove the third statement of Proposition 1, it is enough to show that at least one individual chooses $a=1$. We start with deriving the conditions under which $U(p_1,\kappa_1,y) > U(p_2,\kappa_2,y)$. We follow similar steps as in the previous statement's proof. Let $\Delta_{12} =$

$U(p_1, \kappa_1, y) - U(p_2, \kappa_2, y)$ denote the maximized utility difference between the cases where agent chooses $a = 1$ and $a = 2$. Using Lemma 2, we decompose this difference as follows:

$$\Delta_{12}U = \big[U(p_1, \kappa_1, y) - U(p_1, \kappa_2, y)\big] + \{U(p_1, \kappa_2, y) - U(p_2, \kappa_2, y)\} \tag{C.21}$$

Note that the item in the square brackets is positive, whereas the item in the curly brackets is negative because, all else constant, utility decreases in κ_i and p respectively.

Next, using the second fundamental theorem of calculus and the envelope theorem, we get the following expression for the item in curly parentheses:

$$\begin{aligned} &U(p_1, \kappa_2, y) - U(p_2, \kappa_2, y) = \\ &\int_{p_2}^{p_1} \big[u\big((1 - \tau + \tau g(p, \kappa_2, y) - \tilde{\kappa_2})y\big) - u\big((1 - \tau - \tau\theta g(p, \kappa_2, y) - \tilde{\kappa_2})y\big)\big]\, dp \end{aligned} \tag{C.22}$$

Substituting the definition of U given in Equation C.8 for the square brackets in Equation C.21 yields:

$$\begin{aligned} U(p_1, \kappa_1, y) - U(p_1, \kappa_2, y) = (1 - p_1)&\big[u\big(c_N(p_1, \kappa_1, y)\big) - u\big(c_N(p_1, \kappa_2, y)\big)\big] \\ + p_2&\big[u\big(c_D(p_1, \kappa_2, y)\big) - u\big(c_D(p_1, \kappa_2, y)\big)\big] \end{aligned} \tag{C.23}$$

Combining Equations C.22 and C.23 gives a new expression for the difference equation:

$$\begin{aligned} \Delta_{12}U = (1 - p_1)\big[u\big(c_N(p_1, \kappa_1, y)\big) - u\big(c_N(p_1, \kappa_2, y)\big)\big] + p_2\big[u\big(c_D(p_1, \kappa_2, y)\big) - u\big(c_D(p_1, \kappa_2, y)\big)\big] \\ - \int_{p_2}^{p_1} \big[u\big((1 - \tau + \tau g(p, \kappa_2, y) - \tilde{\kappa_2})y\big) - u\big((1 - \tau - \tau\theta g(p, \kappa_2, y) - \tilde{\kappa_2})y\big)\big]\, dp \end{aligned} \tag{C.24}$$

Applying the second fundamental theorem of calculus again to the brackets in the previous equation we get:

$$\Delta_{12}U = (1-p_1)\int_{c_N(p_1,\kappa_2,y)}^{c_N(p_1,\kappa_1,y)} u'(c)dc + p_1 \int_{c_D(p_1,\kappa_2,y)}^{c_D(p_1,\kappa_1,y)} u'(c)dc - \int_{p_2}^{p_1}\int_{c_D(p,\kappa_2,y)}^{c_N(p,\kappa_2,y)} u'(c)dcdp \tag{C.25}$$

Note that the upper bounds of each integral are arranged such that the integrals are well-defined.

We next want to find a minimum bound for this expression. Due to the diminishing marginal utility of U, the first two integrals are at their minimum when evaluated at their upper bounds, whereas the third term is largest when evaluated at its lower bound. Evaluating the integrals so $\Delta_{12}U$ is at its minimum, gives us the following lower bound for the utility difference equation:

$$
\begin{aligned}
\Delta_{12}U >& (1-p_1)\big[c_N(p_1,\kappa_1,y)-c_N(p_1,\kappa_2,y)\big]u'\big(c_N(p_1,\kappa_1,y)\big) \\
&+ p_1\big[c_D(p_1,\kappa_1,y)-c_D(p_1,\kappa_2,y)\big]u'\big(c_D(p_1,\kappa_1,y)\big) \\
&- \int_{p_2}^{p_1}[c_N(p,\kappa_2,y)-c_N(p,\kappa_2,y)]u'\big(c_D(p,\kappa_2,y)\big)dp
\end{aligned}
\tag{C.26}
$$

We use the FOC from the agent's problem, given in Equation C.1, to further simplify Equation C.26:

$$
\begin{aligned}
\Delta_{12}U >& p_1u'\big(c_D(p_1,\kappa_1,y)\big)\Big(c_D(p_1,\kappa_1,y)-c_D(p_1,\kappa_2,y)+\theta\big(c_N(p_1,\kappa_1,y)-c_N(p_1,\kappa_2,y)\big)\Big) \\
&- \int_{p_2}^{p_1}\big[c_N(p,\kappa_2,y)-c_D(p,\kappa_2,y)\big]u'\big(c_D(p,\kappa_2,y)\big)dp
\end{aligned}
\tag{C.27}
$$

Because c_D increases in p, $u'\big(c_D(p,\kappa_2,y\big)$ decreases in p, evaluating this marginal utility as constant at $p=p_2$ will simplify the previous equation without violating the inequality:

$$
\begin{aligned}
\Delta_{12}U >& p_1u'\big(c_D(p_1,\kappa_1,y)\big)\Big(c_D(p_1,\kappa_1,y)-c_D(p_1,\kappa_2,y)+\theta\big(c_N(p_1,\kappa_1,y)-c_N(p_1,\kappa_2,y)\big)\Big) \\
&- u'\big(c_D(p_2,\kappa_2,y)\big)\int_{p_2}^{p_1}\big[c_N(p,\kappa_2,y)-c_D(p,\kappa_2,y)\big]dp
\end{aligned}
\tag{C.28}
$$

Substituting the definitions of c_D and c_N with the relevant parameters we get:

$$
c_D(p_1,\kappa_1,y)-c_D(p_1,\kappa_2,y)=y\Big(\tilde{\kappa}_2-\tilde{\kappa}_1-\tau\theta\big(g(p_1,\kappa_1,y)-g(p_1,\kappa_2,y)\big)\Big) \tag{C.29}
$$

$$
c_N(p_1,\kappa_1,y)-c_N(p_1,\kappa_2,y)=y\Big(\tilde{\kappa}_2-\tilde{\kappa}_1+\tau\big(g(p_1,\kappa_1,y)-g(p_1,\kappa_2,y)\big)\Big) \tag{C.30}
$$

The expression in the big parenthesis boils down to the following:

$$
\begin{aligned}
\Big(c_D(p_1,\kappa_1,y)-c_D(p_1,\kappa_2,y)+\theta\big(c_N(p_1,\kappa_1,y)-c_N(p_1,\kappa_2,y)\big)\Big) &= y(1+\theta)(\tilde{\kappa}_2-\tilde{\kappa}_1) \\
&= (1+\theta)(\kappa_2-\kappa_1)
\end{aligned}
$$

Substituting this back into the lower bound in Equation C.28 gives

$$
\begin{aligned}
\Delta_{12}U >& p_1\Big(u'\big(c_D(p_1,\kappa_1,y)\big)(1+\theta)(\kappa_2-\kappa_1)\Big) \\
&- u'\big(c_D(p_2,\kappa_2,y)\big)\int_{p_2}^{p_1}\big[c_N(p,\kappa_2,y)-c_D(p,\kappa_2,y)\big]dp
\end{aligned}
\tag{C.31}
$$

Note that $c_N(p,\kappa_2,y) - c_D(p,\kappa_2,y) = \tau(1+\theta)g(p,\kappa_2,y)y$ which decreases in p as $g(p,\kappa_2,y)$ decreases in p for a risk-averse agent by Lemma 1. Therefore, evaluating the integral at p_2, does not violate the above inequality. We define $f(y_{12})$ as a lower bound for $\Delta_{12}U$:

$$\Delta_{12}U > f(y_{12}) = p_1\Big(u'\big(c_D(p_1,\kappa_1,y)\big)(1+\theta)(\kappa_2-\kappa_1)\Big) \\ - u'\big(c_D(p_2,\kappa_2,y)\big)(p_1-p_2)\tau(1+\theta)g(p_2,\kappa_2,y)y \tag{C.32}$$

rearranging the terms gives us the following condition

$$f(y_{12}) > 0 \iff \kappa_2 - \kappa_1 > \frac{u'\big(c_D(p_1,\kappa_1,y)\big)}{u'\big(c_D(p_2,\kappa_2,y)\big)}\frac{(p_1-p_2)}{p_1}\tau g(p_2,\kappa_2,y)y \tag{C.33}$$

We next need to show that the agents with income levels that satisfy this inequality will not prefer self-preparation over the action $a=1$. Following the same steps as the first part of the proposition, we get the following lower bound, $f_{10}(y)$, on the utility difference between the cases $a=0$ and $a=1$ (denoted by $\Delta_{10}U$):

$$\Delta_{10}U > f_{10}(y) = -(1-p_1)\Big(c_N(p_1,0,y) - c_N(p_1,\kappa_1,y)\frac{1}{\theta}\big[c_D(p_1,0,y) - c_D(p_1,\kappa_1,y)\big]\Big) \\ + \frac{u'\big(c_N(p_1,0,y)\big)}{u'\big(c_N(p_1,\kappa_1,y)\big)}(p_0-p_1)\tau(1+\theta)g(p_0,0,y)y \tag{C.34}$$

where $\Delta_{10}U = U(p_1,\kappa_1,y) - U(p_0,0,y)$.

Similar to equation C.31, the expression in the big parenthesis boils down to κ_1. Therefore, the lower bound can further be simplified to get the following condition:

$$f(y_{10}) > 0 \iff \kappa_1 < \frac{u'(c_N(p_1,0,y))}{u'(c_N(p_1,\kappa_1,y))}\frac{(p_0-p_1)}{1-p_1}\tau\theta g(p_0,\kappa_0,y)y \tag{C.35}$$

Finally, we can conclude that if the fixed costs for low-and specialized tax preparation, κ_1 and κ_2 respectively, satisfy the following condition, then some agents in the interval $[y', y'']$ will find it optimal to hire a generalized tax preparer:

$$\kappa_1 < min\left\{\frac{u'\big(c_N(p_1,0,y)\big)}{u'\big(c_N(p_1,\kappa_1,y)\big)}\frac{(p_0-p_1)}{1-p_1}\tau\theta g(p_0,\kappa_0,y)y, \kappa_2 - \left(\frac{u'\big(c_D(p_1,\kappa_1,y)\big)}{u'\big(c_D(p_2,\kappa_2,y)\big)}\frac{(p_1-p_2)}{p_1}\tau g(p_2,\kappa_2,y)y\right)\right\} \tag{C.36}$$

This condition corresponds to Assumption 2, which concludes the proof.

□

BIBLIOGRAPHY

BIBLIOGRAPHY

Allingham, M. G. and A. Sandmo (1972a): "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1, 323 – 338.

——— (1972b): "Income tax evasion: A theoretical analysis," *Journal of public economics*, 1, 323–338.

Alstadsæter, A., N. Johannesen, and G. Zucman (2018): "Who owns the wealth in tax havens? Macro evidence and implications for global inequality," *Journal of Public Economics*, 162, 89–100.

——— (2019): "Tax evasion and tax avoidance," *Working Paper*.

Alvaredo, F., A. B. Atkinson, T. Piketty, and E. Saez (2013): "The top 1 percent in international and historical perspective," *Journal of Economic perspectives*, 27, 3–20.

Amromin, G. and P. Smith (2003): "What Explains Early Withdrawals from Retirement Accounts? Evidence from a Panel of Taxpayers," *National Tax Journal*, 56, 595–612.

Angrist, J. D., G. B. Imbens, and A. B. Krueger (1999): "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 14, 57–67.

Angrist, J. D. and A. B. Krueger (1995): "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business & Economic Statistics*, 13, 225–235.

Argento, R., V. L. Bryant, and J. Sabelhaus (2015): "Early Withdrawals from Retirement Accounts During the Great Recession," *Contemporary Economic Policy*, 33, 1–16.

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

Ayres, F. L., B. R. Jackson, and P. S. Hite (1989): "The economic benefits of regulation: Evidence from professional tax preparers," *Accounting Review*, 300–312.

Battaglini, M., L. Guiso, C. Lacava, and E. Patacchini (2019): "Tax professionals: Tax-evasion facilitators or information hubs?" Tech. rep., National Bureau of Economic Research.

Bee, A. and J. Mitchell (2017): "Do Older Americans Have More Income Than We Think?" U.s. census bureau working paper sehsd-wp2017-39.

Bellemare, M. F. and C. J. Wichman (2020): "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 82, 50–61.

Benartzi, S. and R. H. Thaler (1995): "Myopic Loss Aversion and the Equity Premium Puzzle," *The Quarterly Journal of Economics*, 110, 73–92.

Berman, E., J. Bound, and S. Machin (1998): "Implications of skill-biased technological change: international evidence," *The quarterly journal of economics*, 113, 1245–1279.

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2019): "Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches," *The Review of Economic Studies*, 87, 656–690.

Brady, P. J. and S. Bass (2020): "Reconciling Form 1040 and Form 1099-R Data," Soi working paper.

Brown, J. R., J. Poterba, and D. P. Richardson (2017): "Do Required Minimum Distribution Rules Matter? The Effect of the 2009 Holiday on Retirement Plan Distributions," *Journal of Public Economics*, 151, 96 – 109.

Butrica, B. A., S. R. Zedlewski, and P. Issa (2010): "Understanding Early Withdrawals from Retirement Accounts," The urban institute retirement policy program discussion paper 10-02.

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

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011): "Adjustment Costs, Firm Response, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*, 126, 749–804.

Choukhmane, T. (2021): "Default Options and Retirement Saving Dynamics," Working paper.

Cranor, T., J. Goldin, T. Homonoff, and L. Moore (2020): "Communicating tax penalties to delinquent taxpayers: Evidence from a field experiment," *National Tax Journal*, 73, 331–285.

Dobbie, W., J. Goldin, and C. S. Yang (2018): "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, 108, 201–40.

Dobbie, W., P. Goldsmith-Pinkham, and C. S. Yang (2017): "Consumer Bankruptcy and Financial Health," *The Review of Economics and Statistics*, 99, 853–869.

Dobbie, W. and J. Song (2015): "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection," *American Economic Review*, 105, 1272–1311.

Doyle, Jr., J. J. (2007): "Child Protection and Child Outcomes: Measuring the Effects of Foster Care," *American Economic Review*, 97, 1583–610.

——— (2008): "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care," *Journal of Political Economy*, 116, 746–770.

Dubin, J. A., M. J. Graetz, M. A. Udell, and L. L. Wilde (1992): "The demand for tax return preparation services," *The Review of Economics and Statistics*, 75–82.

Duflo, E. (2001): "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," *American Economic Review*, 91, 795–813.

Einav, L., A. Finkelstein, and P. Schrimpf (2015): "The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D," *Quarterly Journal of Economics*, 130, 841–899.

Erard, B. (1993): "Taxation with representation: An analysis of the role of tax practitioners in tax compliance," *Journal of Public Economics*, 52, 163–197.

——— (1997): "Self-selection with measurement errors A microeconometric analysis of the decision to seek tax assistance and its implications for tax compliance," *Journal of Econometrics*, 81, 319–356.

Feldstein, M. (1976): "Inflation, Income Taxes, and the Rate of Interest: A Theoretical Analysis," *The American Economic Review*, 66, 809–820.

Feldstein, M., J. Green, and E. Sheshinski (1978): "Inflation and Taxes in a Growing Economy with Debt and Equity Finance," *Journal of Political Economy*, 86, 809–820.

Fleischman, G. M. and T. Stephenson (2012): "Client variables associated with four key determinants of demand for tax preparer services: An exploratory study," *Accounting Horizons*, 26, 417–437.

French, E. (2005): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behavior," *The Review of Economic Studies*, 72, 395–427.

Goda, G. S., D. Jones, and S. Ramnath (2018): "How Do Distributions from Retirement Accounts Respond to Early Withdrawal Penalties? Evidence from Administrative Tax Returns," Working paper.

Goodman, L., K. Mackie, J. Mortenson, and H. Schram (2019): "The Evolution of Leakage and Retirement Asset Flows in the U.S." Ssrn working paper.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89.

Government Accountability Office (2015): "Retirement Security: Most Households Approaching Retirement Have Low Savings," Gao-15-419.

GUVENEN, F. (2006): "Reconciling Conflicting Evidence on the Elasticity of Intertemporal Substitution: A Macroeconomic Perspective," *Journal of Monetary Economics*, 53, 1451 – 1472.

GUYTON, J., P. LANGETIEG, D. RECK, M. RISCH, AND G. ZUCMAN (2021): "Tax Evasion by the Wealthy: Measurement and Implications," *CEPR Working Paper no. DP15851.*

HANSEN, V. J. AND R. A. WHITE (2012): "An investigation of the impact of preparer penalty provisions on tax preparer aggressiveness," *Journal of the American Taxation Association*, 34, 137–165.

HITE, P. A. AND G. A. MCGILL (1992): "An examination of taxpayer preference for aggressive tax advice," *National Tax Journal*, 389–403.

HUMPHRIES, J. E., N. MADER, D. TANNENBAUM, , AND W. VAN DIJK (2019): "Does Eviction Cause Poverty? Quasi-Experimental Evidence from Cook County, IL," NBER Working Paper 26139.

INTERNAL REVENUE SERVICE (2016): "2016 Individual Income Tax Returns Line Item Estimates," Tech. rep.

——— (2017): "2017 Data Book," https://www.irs.gov/pub/irs-soi/17databk.pdf [accessed: June 4, 2018].

INVESTMENT COMPANY INSTITUTE (2019): "Investment Company Factbook," Tech. rep.

JAKOBSEN, K., K. JAKOBSEN, H. KLEVEN, AND G. ZUCMAN (2019): "Wealth Taxation and Wealth Accumulation: Theory and Evidence From Denmark," *The Quarterly Journal of Economics*, 135, 329–388.

KEANE, M. P. (2011): "Labor supply and taxes: A survey," *Journal of Economic Literature*, 49, 961–1075.

KLEPPER, S. AND D. NAGIN (1989): "The role of tax preparers in tax compliance," *Policy Sciences*, 22, 167–194.

KLEVEN, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464.

KLEVEN, H. J., M. B. KNUDSEN, C. T. KREINER, S. PEDERSEN, AND E. SAEZ (2011): "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark," *Econometrica*, 79, 651–692.

KLEVEN, H. J. AND M. WASEEM (2013): "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics*, 128, 669–723.

KLING, J. R. (2006): "Incarceration Length, Employment, and Earnings," *American Economic Review*, 96, 863–876.

Laibson, D., P. Maxted, A. Repetto, and J. Tobacman (2018): "Estimating Discount Functions with Consumption Choices with the Lifecycle," Working paper.

Laibson, D. I., A. Repetto, J. Tobacman, R. E. Hall, W. G. Gale, and G. A. Akerlof (1998): "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity*, 1998, 91–196.

Larrimore, J., J. Mortenson, and D. Splinter (2019): "Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions," *Journal of Human Resources*.

Loeffler, C. (2013): "Does imprisonment alter the life course? Evidence on crime and employment from a natural experiment," *Criminology*, 51, 137–166.

Long, J. E. and S. B. Caudill (1987): "The usage and benefits of paid tax return preparation," *National Tax Journal*, 35–46.

Maestas, N., K. J. Mullen, and A. Strand (2013): "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, 103, 1797–1829.

Masken, K., M. Mazur, J. Meikle, and R. Nord (2008): "Do Products Offering Expedited Refunds Increase Income Tax Noncompliance?" *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 101, 106–117.

Mehra, R. and E. C. Prescott (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

Miller, E., S. Orlett, and A. Turk (2014): "Uncollectible vs. Unproductive: Compliance Impact of Working Collection Cases That are Ultimately Not Fully Collectible," in *IRS-TPC Research Conference: Advancing Tax Administration.*

Miller, J. J., S. Nikaj, et al. (2016): "The response of delinquent taxpayers to more aggressive collection," *National Tax Journal*, 69, 77–102.

Mortenson, J. A., H. R. Schramm, and A. Whitten (2019): "The Effects of Required Minimum Distribution Rules on Withdrawals from Traditional Individual Retirement Accounts," *National Tax Journal*, 72, 507–542.

Mueller-Smith, M. (2015): "The Criminal and Labor Market Impacts of Incarceration," Working paper.

OASDI (2020): "The 2020 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds," Available at https://www.ssa.gov/oact/tr/2020/.

Perez-Truglia, R. and U. Troiano (2018): "Shaming Tax Delinquents," *Journal of Public Economics*, 167, 120–137.

Phillips, M. D. (2014): "Deterrence vs. gamesmanship: Taxpayer response to targeted audits and endogenous penalties," *Journal of Economic Behavior & Organization*, 100, 81–98.

Rosen, S. (1981): "The economics of superstars," *The American economic review*, 71, 845–858.

Rouwenhorst, G. K. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Structural Models of Wage and Employment Dynamics*, ed. by T. Cooley, Princeton: Princeton University Press.

Sabelhaus, J. (2000): "Modeling IRA Accumulation and Withdrawals," *National Tax Journal*, 53, 865–876.

Scheuer, F. and J. Slemrod (2020): "Taxation and the Superrich," *Annual Review of Economics*, 12, 189–211.

Slemrod, J. (1989): "The return to tax simplification: An econometric analysis," *Public Finance Quarterly*, 17, 3–27.

Slemrod, J., B. Collins, J. L. Hoopes, D. Reck, and M. Sebastiani (2017): "Does credit-card information reporting improve small-business tax compliance?" *Journal of Public Economics*, 149, 1–19.

Slemrod, J. and N. Sorum (1984): "The compliance cost of the U.S. individual income tax system," *National Tax Journal*, 461–474.

TIGTA (2015-10-042): "Improvements Can Be Made to Educate and Notify Taxpayers of Required Minimum Distributiona Requirements From Individual Retirement Arrangements," Tech. rep.

Yuskavage, A., J. DeBacker, B. Heim, and A. Tran (2019): "Seeking professional advice: The effects of tax preparers on compliance," in *112th Annual Conference on Taxation*, NTA.