The Role of Tax Preparers in Individual Tax Optimization

Hadar Avivi[*] Katarzyna Bilicka[†] Jakob Brounstein[‡]
Felipe Lobel[§] Alexander Yuskavage[¶]

Preliminary, please do not circulate or cite

Abstract

We examine the role of paid tax preparers in tax optimization of individual taxpayers. First, using the universe of annual personal income tax returns in the US from 2011-2020, we provide novel descriptive evidence characterizing the paid tax preparers and the users of their services. Second, we quantify the effects of paid preparers on tax optimization, using the two-way fixed effects model and leveraging the rotation of clients among preparers. Our findings suggest that tax preparers explain 8% of the post-tax income inequality, highlighting the significant role that paid tax preparation plays in shaping post-tax income disparities.

JEL codes: H20, H22, H23, H24, H26 *Key words:* Tax avoidance, tax preparers, inequality, tax revenue collection

[*]havivi@berkeley.edu, *UC Berkeley*

[†]kat.bilicka@usu.edu, *Utah State University, NBER, CEPR and Oxford University Centre for Business Taxation; Jon Huntsman Business School, Logan, UT, United States,*

[‡]jakob.brounstein@gmail.com, *Institute for Fiscal Studies*

[§]lobel@berkeley.edu, *UC Berkeley*

[¶]Alexander.Yuskavage@treasury.gov, *U.S. Department of the Treasury, Office of Tax Analysis.* This research was conducted while Yuskavage was an employee at the U.S. Department of the Treasury, and at no time was confidential taxpayer data ever outside of the Treasury or IRS computing environment. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury or the Internal Revenue Service. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

Over 55% of American households use a paid tax preparer. Yet, relatively little is known about the impact that access to paid tax preparer technology has on tax optimization of US individual taxpayers.[1] In this paper, we use the universe of US individual income tax returns between 2011 and 2020 to ask questions about the characteristics of individuals who use paid tax preparers and the paid tax professionals who provide these services. Then, we causally estimate the contribution of tax preparers to individual tax optimization across the income distribution and investigate the specific means by which paid tax preparers engage in tax optimization. Finally, we analyze the implications that the services provided by paid tax preparers have on post-income inequality. Broadly, we seek to advance our understanding of the complex landscape of paid tax preparation for individuals in the US.

Tax preparers are specialized professionals that operate typically for the purpose of helping their clientele (individual or business taxpayers) maintain tax compliance and minimize their tax burdens. They combine their knowledge of the tax code (obligations, deductions, credits, legal structures, etc.) with clientele characteristics in order to reduce the amount of tax that their clients have to pay. In countries that feature automatic income tax filing, paid tax preparers tend to assist only taxpayers with more complicated income types; in countries that do not feature automatic income tax filing for most taxpayers, like the United States, paid preparers are widely used across the population.

In order to understand the role of tax preparers in individual tax optimization, we develop a set of novel measures of tax optimization that rely on the comparisons between actual tax paid and the predicted tax paid. We construct the prediction using forward mapping based on broad income and tax-payer characteristics, such as state of residence, filing type, and year.[2] Our preferred tax optimization measure is constructed as 1 minus the ratio of actual tax paid to predicted tax paid, such that we can interpret this measure as a percent of tax saved relative to predicted tax liability based on broad income and individual observable

[1] In future related work we will focus on business and corporate usage of paid tax preparers.

[2] Our framework is analogous to the TAXSIM model of Feenberg and Coutts (1993) and we measure broad income as the reported total income on the tax form and add the other non-taxable net income sources from the additional schedules C,E, and F.

characteristics. Further, to understand how tax preparers affect different margins of tax optimization, we construct analogous measures of deductions and credit optimization.

We start the paper by providing a set of new stylized facts on the use of tax preparers across the US household and how much they contribute to tax optimization. We find that consistently across all the sample years around 65% of US gross fiscal income is handled by paid tax preparers. In 2020, this was around $7.5 trillion (35% of US GDP). While the use of tax preparers is prevalent across the entire income distribution, with around 55% of households using one, on average, richer households are more likely to use tax preparers, with 90% of the top 1 percent of the richest households having one. Households that use a paid tax preparers have lower ETRs conditional on their income and their tax saving increases across the income distribution, such that individuals at the top have 2-3 pp lower ETRs if they use a tax preparer than if they do not. We then find that tax preparers help households in their tax optimization saving them between 5-10pp of their tax liability relative to those households that do not use tax preparers. The effectiveness of tax preparers is most pronounced in deductions optimization for richer households and credits optimization for those in the middle of the income distribution.

In the second part of the paper, we estimate a two-way fixed effect model to quantify the contribution of tax preparers to tax optimization and to isolate the causal effect of those preparers on tax optimization from the household's ability and sorting. We build on a large literature in labor economics that uses matched employer-employee data to separately identify worker and firm fixed-effects (Abowd et al., 1999a; Card et al., 2013a). In our context, to separate the effects of tax preparers from those of individual taxpayers, we rely on taxpayers moving across different tax preparers. Within the largest connected set of taxpayers and tax preparers, under a strict exogeneity assumption, we identify the causal contribution of paid preparers to tax optimization and decompose the variance of the observed tax optimization into the individual, tax preparer and other contributions. Using event studies, we verify that this strict exogeneity assumption holds in our data, by showing that the assignment of taxpayers to preparers is not characterized by payer-preparer match effects. Specifically, we find 1) no pre-event trends in tax optimization around taxpayer switches to different tax

preparers, 2) and that moving to a better (worse) tax preparer increases (decreases) tax optimization, and 3) moving to a worse preparer produces a symmetric reduction in tax optimization as does moving to a better preparer.

We find that tax preparers are most effective at the bottom and top ends of income distribution, with much larger magnitude of deductions optimization, especially for richest taxpayers. Further, we find that tax preparers explain between 14% and 17.5% of variation in tax optimization in the US. We then show that tax preparers also explain 8 and 12% of the variation in post-tax income and final tax payment respectively. This finding highlights that a substantial share of the variation in tax savings is attributable to the variation in quality of tax preparers, emphasizing their crucial role they have in shaping revenue collection and influencing economic inequality.

Relatively little work exists that comprehensively describes tax preparer usage and associated tax savings. Most of the existing work focuses on the effects of audits. Most notably, DeBacker et al. (2019) study the differential impact of random income tax audits in the US based on paid preparer usage, finding slightly lower adjustment rates for paid preparer users, but conditional on adjustment, paid preparer users see a 55% (USD 3,100) higher adjustment. Battaglini et al. (2022) use a control function approach to evaluate tax noncompliance among small business owners in Italy based on paid preparer usage, finding substantial income underreporting among users. Moreover, they find that the tax preparer response to audits diffuses to other peer-clientele of taxpreparers. In this paper, we aim to advance a much broader understanding of paid tax preparation services by focusing on describing the tax preparers and their users and quantitatively characterizing the distribution of tax optimization effects of paid preparers. Our main specification draws separate identification of taxpayer and paid preparer optimization effects from taxpayers that switch paid preparer (see also Zwick (2021)); we justify the validity of our estimates by providing diagnostic tests and additional estimates that make use of random audits.

This paper also contributes to the large literature describing inequality, tax strategy (both evasive and avoiding) and the distributional incidence of the US federal income tax (Auten and Splinter (ming); Guyton et al. (2023); Smith et al. (2022); Piketty et al. (2017); Saez

and Zucman (2016); Piketty and Saez (2003)). These paper have sought to characterize the evolution in income and wealth inequality over time as well as the incidence of tax noncompliance issues across the income distribution (also Brounstein (2023); Londoño-Vélez and Ávila Mahecha (2021); Alstadsæter et al. (2019); Alstadsæter et al. (2018)). Beyond characterizing the value added of tax preparers, our work informs our understanding of income inequality and the US tax code by describing the distribution of tax preparer value added across the income distribution. How does the institution of paid tax preparation service cause the *de facto* progressivity of the US federal income tax code to depart from its statutory progressivity? Our work also seeks to estimate the counterfactual scenarios of income inequality under different allocations of tax preparer value added. Finally, we will identify the exact provisions that facilitate both legal and illicit tax optimization.

2 Data

To understand the role of tax preparers in individual tax optimization, we use the universe of US annual personal income tax returns which allows us to track income of individual taxpayers and their tax preparers over time. In this section we describe the data sources and the main measures of tax optimization we use in this paper. We also provide comprehensive descriptive statistics about tax payers, paid preparers and their interactions.

2.1 Tax Forms

Our dataset is constructed using Forms 1040 filed by US individuals and households annually between 2011 and 2020. This rich dataset includes detailed information on income, deductions, and credits, along with supplementary schedules that break down these amounts into more specific items. A key feature of our data is that taxpayers utilizing professional tax services report their preparers' tax identification number on Form 1040. This enables us to track each individual's history of preparers usage including how often they change tax preparers. Further, because we can identify who the tax preparers are, we are also able to look at their own tax returns and track their clientele over time. To manage computational

demands, we currently utilize a 1% random sample of the entire dataset while drawing upon information from the whole population of paid preparers. This sample includes over 9 million taxpayer-year observations, representing over 2 million unique taxpayers and over 500,000 paid tax preparers.

2.2 Measures of Tax Optimization

The first measure of tax optimization we use is the effective income tax rate (ETR), which is measured as final tax paid divided by broad income. However, the central shortcoming of using ETR as a measure of tax optimization lies in its mechanical relationship with income and consequently tax schedule – higher incomes mechanically see greater ETR. Since we are interested in tax optimization independent of income levels, we develop a novel tax optimization measure based on mappings between actual tax paid and predicted tax paid based on broad income and tax-payer characteristics. Further, to understand how tax preparers affect different margins of tax optimization, we construct deductions and credit optimization measures to disentangle these two margins. Each of our three preferred tax optimization measures quantifies the proportional difference between a stage of the income tax filing process and a measure of taxes paid on that income.

Central to our measures of income tax optimization is our definition of broad un-adjusted income, B, that aims to abstract from all potential tax optimization strategies[3]. This measure of "broad income" takes the reported total income on the tax form and adds the other non-taxable net income sources from the additional schedules.[4] We then design a mapping $T^b = T(B; \Theta)$ from broad income B to federal income tax liability T that takes broad income as wage income, also considering state of residence, filing type, and year, following a simplification of TAXSIM (Feenberg and Coutts, 1993). This mapping features the application of deduction, D, that remove income from tax base, and credits, C, that reduce tax liability

[3]There may be endogenous real activity responses to anticipated tax optimization strategy, as that strategy affects the actual income tax rate that individuals pay. E.g. taxpayers perceiving lower effective income tax rates may generate greater labor earnings. In this version of the paper, we abstract from these endogenous responses, but will analyze them in the future.

[4]We use revenues, expenses, and realized capital gains/losses from schedules C,E, and F, as well as negative amounts reported on additional schedules.

directly to mimic the *actual* final tax liability, which can be defined as $T^f = T(B - D) - C$. We summarize the key income and tax concepts we use here in Table 1.

There are several complications that arise with this approach. First, the income tax function is typically non-linear, which makes it difficult to compare the value of credits and deductions at different levels of income. We calculate the tax base equivalent value of credits as the difference between a tax base at a given point and the inverse mapping T^{-1} to the tax base of the difference between implied taxes owed on the fixed base and the credit value. As such, we define the base-equivalent value of refundable credits are: $C_B = T^{-1}(T^f + C) - T^{-1}(T^f)$, where the first term is the base-value of taxes without credits and the second term is the base-value of taxes with credits.[5] Second, because tax base contains several different types of income, such as, for example, labor or capital income, the mapping $T(B; \Theta)$ is not truly bijective. To be able to construct our mapping, we treat all income as wage income.[6] Typically, ordinary wage income sees a higher statutory tax rate than capital income or other types of preferred income, so we argue wage income constitutes a suitable baseline.

Using our mappings, we define our measure of final tax optimization as:

$$Final\ tax\ opt. = 1 - \frac{T^f}{T^b} \tag{1}$$

where T^f is the final tax liability taken directly from income tax returns and T^b is the tax liability based on forward mapping from broad income described above. This measure of final tax optimization is typically bound between 0 and 1 and measures how much tax was saved using deductions and credits. For example, when final tax optimization is 0.4 this means that a given individual, based on their broad income and individual observable characteristics, is paying 40% less in taxes than their broad income would predict.[7]

[5] We limit the maximum base-equivalent value of non-refundable tax credits to the value of the applicable tax base: $C_B^{NR} = \min\left(T^{-1}(T^f), T^{-1}(T^f + C) - T^{-1}(T^f)\right)$. We also map deductions into tax-equivalent in Appendix B.

[6] Of course, tax on wage can be higher or lower than tax on capital across the income distribution. Further, individuals with large share of non-wage income may be inherently different and we explore this heterogeneity empirically in the paper.

[7] Note that this measure is mechanically higher for individuals with de facto zero final tax. We show the

To understand the sources of tax optimization, we split our final tax optimization measure into deductions and credits optimization. The US income tax form 1040 broadly features base/deductions optimization for until defining taxable income[8] Following the calculation of initial tax liability[9], tax optimization broadly focuses on the application of nonrefundable and refundable credits. We can directly infer both of those optimizations from tax returns without any imputations. Since the tax liability is calculated by applying deductions to the tax base first, we start with deductions optimization, which we define as:

$$Deduction\ opt. = 1 - \frac{Taxable\ income}{B} = 1 - \frac{B - D}{B} = \frac{D}{B} \tag{2}$$

Then, after calculating preliminary tax liability based on deductions, tax credits can be used to lower that tax liability further. We define credit optimization as follows:

$$Credit\ opt. = 1 - \frac{T^f}{T(B - D)} = 1 - \frac{T(B - D) - C}{T(B - D)} = \frac{C}{T(B - D)} \tag{3}$$

We summarize our three preferred measures of tax optimization in Table 2. In the Appendix we provide comprehensive descriptive statistics for taxpayers, tax preparers and their interactions. Specifically, Tables A2 and A3 summarize taxpayer and tax preparer characteristics respectively, while Table A5 disaggregates these characteristics by paid preparer type, illustrating substantial differences based on the paid preparer background. Table A1 compares tax optimization measures for taxpayers with and without tax preparers, while Table A4 provides information on the tax preparers clientele. Table A4 reveals that on average, each preparer has 200 clients (corresponding with the mean multiplied by a scalar 1% sample adjustment), although this figure conceals considerable variation. Among the top 95% most popular preparers, the average is approximately 600 clients per year. The typical gross income of these clients is $167,747, which is substantially higher than the $122,717 average for the general taxpayer population, as shown in Table A2. This suggests that individuals who opt for tax preparation services tend to be higher earners.[10]

distribution of this measure in Figure 4.

[8] In the 2021 version of Form 1040, Taxable Income is defined on line 15.

[9] Initial tax liability is calculated in Line 16.

[10] Figure A1 suggest that the size of tax preparer network is small when they serve very high income tax

In principle, all of the measures we develop in this sub-section are independent of income level. While in the paper we prefer measures of tax optimization that correspond with tax that should have been paid on a given income, in the Appendix, we discuss alternative tax optimization measures based on broad income derived from tax liability and show that our findings are robust to those alternatives. We perform the mapping in both directions as, ex-ante, it is unclear whether tax optimization should be measured against a counterfactual disposable income or base-equivalent.

2.3 Stylized Facts

We start our empirical investigation by presenting a set of descriptive facts on (1) the use of tax preparers in the US, (2) tax optimization measures across time and individual income distribution, and (3) tax preparer clientele and networks.

2.3.1 The use of tax preparers

Fact 1: The use of tax preparers is prevalent in the US. In Figure 1 we show that between 2011 and 2020 around 55% of the US households have used a tax preparers. This share has declined in recent years, but only by about 2 percentage points. Further, about 65% of US gross fiscal income is handled by a paid tax preparer. This share has also been stable since 2011. In fact, tax preparers handled about $60 billion of aggregate gross income (AGI) in 2011 and about $75 billion in 2020. The number of paid tax preparers has been relatively stable between 2011 and 2017. It has then declined by around 6.5% since 2017 from 225,000 to 212,00 in 2019. This decline is likely driven by changes in the tax law, which disallowed the deductibility of tax preparer expenses from taxable income.

Fact 2: Richer households are more likely to use a tax preparer. In Figure 2, we document that the use of tax preparers, while prevalent across the income distribution, increases substantially for richer households. Around 40% of households at the 10th percentile

clientele (about 50 clients) and when they serve a very low income tax clientele. Further, the size of the tax preparer network also depends on the qualifications of the tax preparer. Certified Acceptance Agents have the largest clienteles, while Attorneys have one the smallest ones, on average (see Table A5).

of the gross income distribution uses tax preparer services. In turn, 90% of households in the top 1 percent of the gross income distribution use a tax preparer. This share increases slowly until the 85th percentile of the income distribution and sees as exponential increase between 85th and 99th percentile from 60% of households to 90% of households.

2.3.2 Tax optimization measures

Facts 3 and 4: Households that use a paid tax preparer have lower ETRs. This tax saving increases across the income distribution In Figure 3 we plot the effective tax rates and the logarithm of final tax across the income distribution distinguishing between households that use a tax preparer (in red) and those that do not use one (in blue). In Panel A, we show that the ETRs of households that use a tax preparer are consistently lower across the income distribution. In fact, in Panel B, we plot the difference between the ETR paid by households which use a tax preparer and those that do not and find that this difference is stable at around 1 percentage point between 40th and 80th percentiles of income distribution. Then, it increases sharply towards the top end of the income distribution. In Panel C, we demonstrate that this difference is largest for households between 95th and 98th percentiles of income distribution, where it is between 2 and 3 percentage points. This is further underscored by the fact that households who use a tax preparer pay less tax than comparable households with similar gross income who do not (Panel D). Of note is the finding that at the very top of the income distribution (top 1 percent), the households which use a paid preparer have higher ETRs than those that report not to.

Fact 5: Tax preparer users optimize more than those who do not have a tax preparer We present the distribution of our 3 main tax optimization measures in Figure 4. In Panels d-f, we show how these tax optimization measures evolved over time and find that, on average, the amount of overall optimization has declined over time both for those households that use a tax preparer as well as those that do not at about the same pace. In 2011, households that used a tax preparer paid about 54% less tax than their broad income would predict. In 2020, they paid 48% less. Panels e and f show that, on

average, aggregate deductions optimization has remained relatively static between 2011 and 2017 for both taxpayers with and without paid preparer. We observe a sharp increase in deductions starting 2018 following the Tax Cuts and Jobs Act (TCJA), which doubled the standard income tax deduction. We observe the opposite trend for credits optimization, which declined steadily declining between 20111 and 2017, and saw a large step decline in 2018, also following TCJA. Of note, the gap in both the credits and deductions optimization between those who use and do not use a tax preparer has shrunk after the TCJA. These changes likely signify the underlying changes in the tax code for the average household in the income distribution.

In Figure 5, we plot the distribution of our 3 main optimization measures across the entire income distribution and in Figure 6 we zoom in on the top 10 percent of the income distribution. In Panels a-c, we split our optimization measures to show separately the distribution for households which use a tax preparer (in red) and those that do not (in blue). In Panels d-f, we plot the difference in optimization measures between the two groups. Several interesting patterns emerge. First, our measures of optimization decline across the income distribution, suggesting that the difference between what broad income predicts that a household would be paying in taxes and what it is actually paying is lower for higher income households. This is true for both those who use tax preparers and those who do not. Second, tax preparer users optimize more than those who do not have a tax preparer. Since, we are ultimately interested in the difference in these optimization measures between the households that use a tax preparer and those that do not, we focus on these differences.

Fact 6: Richer households that use a tax preparer optimize deductions, while middle to lower income households optimize credits more. The size of the difference in final tax optimization is largest for households at the top end of income distribution, especially those around the 95-96th percentiles of income distribution. In fact, Figure 6 suggests that those households optimize around 10 percentage points more of their total tax liability if they use a tax preparer. Again, similar to the ETR differences, this optimization declines for the top 1 percent of richest households. The magnitude of the difference in final optimization between those households that use a tax preparer and those that do not also

increase around the 40th percentile of income distribution, but this increase is smaller. This two-humped shape of the final optimization plot in Panel d can be explained by differences across the income distribution in the use of credits and deductions optimization. In Panel e, we show that households around the 40th percentile of income distribution who use a tax preparer see a 11 percentage points higher credits optimization than those who do not. While taxpayers who use tax preparers still have higher credit optimization at the top of the income distribution, the magnitude of that optimization is much smaller - around 5 pp. In turn, households around the 95-96th percentiles of income distribution who use a tax preparer see a substantial increase in deductions optimization - around 11pp - relative to those households who do not use a tax preparer. At the same time, the deduction optimization for households at the 40th percentile of the income distribution is 'only' 4pp higher if they use a tax preparer.

2.3.3 Client-preparer networks

Fact 7: Households switch tax preparers often. Table 3 shows that there is substantial household-level variation in the client-preparer match. Columns (3) and (4) show that a large share of clients use multiple different preparers over the sample period, and 50% of them use at least one different professional paid preparer. Table A6 displays statistics about preparers rotation among clients. On average, clients stay with a preparer for around 50% of their tenure in the data. Conditional on using a professional tax preparer, clients stay with the same preparer for 1.7 years. This variation is important for our empirical approach, as we will leverage the client mobility among different paid preparers to identify the effects that those professionals have on tax optimization.

Figure A1 provides additional evidence on the size of the tax preparer networks. We plot the average number of clients per paid preparer along the average of the tax preparer clientele income distribution. The figure reveals an intriguing inverted-U shape. Tax preparers who serve clients at the median income level have on average 180 clients, whereas those serving clients at the 20th and 80th percentiles average 140 clients each. This pattern is consistent with the specialized nature of services for high-income clients that limits the scope for expanding operations. Conversely, preparers serving lower-income clients often

operate smaller businesses with limited capacity for scaling up their client base.

3 Empirical Approach

While the differences in tax optimization we demonstrate in the previous section indicate that taxpayers who use tax preparers optimize their taxes more, we are interested in identifying the causal contribution of the tax preparers to this tax optimization coming from their effectiveness and not the differences between taxpayers that are due to selection or sorting of households into using more effective tax preparers. To do so, we adapt a two-way fixed effect model, extending the model from Abowd et al. (1999a) to a tax setting. To assess the effectiveness of tax preparers, we rely on within-client variation across tax preparers ("movers") and explore instances where a client uses multiple tax preparers over time.

In our setting, we have a population of N taxpayers, indexed by i, which in every year $t \in \{1, .., T\}$ are characterized by a base income B_{it}, known by the end of the calendar year, and other individual, possibly time-variant, characteristics x_{it}. At the end of each calendar year, taxpayers decide whether to file their taxes alone or use the services of one of $\mathcal{J}$ tax preparers indexed by j, where $j = 0$ indicates using no tax preparer. This last condition allows us to estimate the effectiveness of tax preparers relative to no preparers.[11]

3.1 Econometric Model

Our econometric framework aims to disentangle the different components that drive tax optimization variation. Let the function $J(i,t) \in \{0, 1, ...\mathcal{J}\}$ indicate the identity of the tax preparer of household i at year t. We model the observed tax optimization z_{it} of household i at time t as a linear combination of tax preparer component $\psi_{J(i,t)}$, household component α_i, and an index of observable time-variant characteristics $x'_{it}\beta$:

$$z_{it} = \alpha_i + \psi_{J(i,t)} + x'_{it}\beta + r_{it}, \tag{4}$$

[11] All VITA preparation services are categorized under a single fixed preparer ID.

where r_{it} is the error term. ψ_j represents tax preparer's j effectiveness that is introduced to all its clients, α_i represents a combination of skills, productivity, and motivation that affect both their pay and their resulting tax saving. x_{it} includes the taxpayer's state of residence by year, which accounts for the effects of state-specific tax law and any resulting tax preparation specialization and year-fixed effects. Equation 4 is inspired by a large literature in labor economics that uses matched employer-employee data to separately identify worker and firm fixed-effects (Abowd et al., 1999a; Card et al., 2013a). The model implies that tax optimization can be written as additively separable in tax preparer fixed effect, taxpayer fixed effect, and taxpayer-time-varying characteristics.

We estimate Equation 4 by Ordinary Least Squares (OLS). As noted by Abowd et al. (1999a), Equation 4 is identified only within a connected set of tax preparers that are linked by taxpayers' mobility, i.e. those that move at least once between the tax preparers during the sample period. Table 4 shows descritpive statistics for this subset of taxpayers (movers), contrasting them with those outside the set (non-movers) and the full sample. Notably, 1.3% of taxpayers are excluded from the *largest* connected set, primarily due to filing with extremely low income. We are unable to estimate taxpayer fixed effects for taxpayers who do not switch preparers, nor can we determine the effectiveness of tax preparers who serve a strictly non-changing clientele base over the observation period.

3.2 Identification Assumption

To identify the underlying taxpayers and tax preparer fixed effects using the model outlined in Equation 4, we requires the strict exogeneity assumption, which posits that the assignment of taxpayers to preparers is uncorrelated with payer-preparer matching. We state this assumption as follows:

$$E[r_{it}|X, \psi, \alpha] = 0,$$

where $X = (x_{11}, ..., x_{NT})$, $\psi = (\psi_0, .., \psi_{\mathcal{J}})$, and $\alpha = (\alpha_1, .., \alpha_N)$. This assumption allows for unrestricted dependence of preparer mobility on tax preparers and taxpayers' effects. For

example, highly skilled taxpayers may be more likely to use the services of highly effective tax preparers than low-skilled taxpayers. However, assuming mean independence between the error term and the entire path of future preparers rules out mobility with respect to these time-varying and/or match effect shocks.

If the variation across tax preparers is mainly attributable to sorting on unobserved characteristics, then taxpayers who switch between tax preparers will not necessarily experience a proportional change in their tax optimization measures. If, on the other hand, tax preparers are characterized by a stable latent effectiveness applied to all of their clients, then we might find that taxpayers who switch to a tax preparer whose other taxpayers experience higher tax optimization gains, will experience a gain as well, regardless of their original preparer.

Card et al. (2013a) propose a simple empirical test to evaluate the plausibility of this assumption, which we follow here. In Figure 7, we present a series of simple event studies that analyze taxpayers' transitions to different tax preparers. We assign tax preparers to their quartile of the distribution of clientele leave-out mean optimization,[12] where 1 indicates bottom 25th percentile in the distribution of the tax optimization (least effective) and 4 indicates top 25th percentile (most effective).

We then track clients' tax optimization behaviors over time, both before and after they switch preparers. We find that clients transitioning from the least to most effective preparers experience significant increases in their tax optimization. Conversely, moving from most to least effective tax preparer results in a symmetric reduction in tax optimization. The figures also suggest that different types of tax preparer movers experience different levels of tax optimization benefits and losses. Furthermore, prior to switching preparers, clients demonstrate a stable tax optimization behavior, suggesting no-pre trends. These findings validate the exogeneity assumption.

[12] I.e. for a taxpayer k with preparer j and a set of $K-1$ other taxpayers $-k$ that also employ tax preparer j in year $\bar{t}$, the leave-out mean optimization of taxpayer i's preparer j is defined as $\frac{1}{K-1}\sum_{i\in -k} x_{i\bar{t}}$.

3.3 Variance Decomposition

Using Equation 4, we can decompose the variance of the observed measures of tax optimization as follows:

$$\begin{aligned}\text{var}(z_{it}) = \text{var}(\alpha_i) + \text{var}(\psi_{J(i,t)}) + \text{var}(x'_{it}\beta) + \text{var}(r_{it}) \\ + 2\text{cov}(\alpha_i, \psi_{J(i,t)}) + 2\text{cov}(\alpha_i, x'_{it}\beta) + 2\text{cov}(\psi_{J(i,t)}, x'_{it}\beta), \quad (5)\end{aligned}$$

Under our identification assumption, each estimated component, $\hat{\psi}_j$, is an unbiased, albeit noisy estimate of the true preparer effect. As such, replacing each term in this equation with its empirical analog produces a biased estimate for the variance components. As noted by Kline et al. (2020a) and Bonhomme et al. (2023), such bias could be non-negligible. To correct for this bias, in future drafts, we will apply Kline et al. (2020a) (hereafter, KSS), leave-out biased correction estimate.[13]. Computing the KSS estimator requires focusing on a leave-one-out connected set sample, which means that the sample remains connected even after each (i, t) has been taken out. Kline et al. (2020a) provides the conditions under which the leave one out variance component is a consistent estimate for the true variance component.

In Table 5 we present the results from our preliminary variance decomposition using the 3 main optimization measures in Panel A and logarithm of tax payments and logarithm of post-tax income in Panel B. Depending on the specific tax optimization measure, the contribution of paid preparers ranges from 14% to 17.5%. Paid tax preparers contribute much more to explaining the variation in credits optimization than in deductions optimization. At the same time, variation across states and years plays a relatively larger role in explaining variation in credits optimization than in deductions optimization.

Findings in Panel B indicate that tax preparers explain 8 and 12% of the variation in post-tax income and final tax payment respectively. This finding highlights that a substantial share of the variation in tax savings is attributable to the variation in quality of

[13]We will use the Python package PYTWOWAY from Bonhomme et al. (2023)

tax preparers, emphasizing their crucial role they have in shaping revenue collection and influencing economic inequality.

Interestingly, the negative correlation between taxpayer and preparer fixed effects, as shown in Table 5, suggests that taxpayers with a higher capacity for tax optimization are less likely to seek out effective tax preparers. This aligns with the intuition that highly skilled clients may not require as much assistance from professional preparers, thus alleviating inequality in tax savings generated by tax preparers.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999a). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999b). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Alstadsæter, A., Johannesen, N., and Zucman, G. (2018). Who owns the wealth in tax havens? macro evidence and implications for global inequality. *Journal of Public Economics*, 162:89–100. In Honor of Sir Tony Atkinson (1944-2017).

Alstadsæter, A., Johannesen, N., and Zucman, G. (2019). Tax evasion and inequality. *American Economic Review*, 109(6):2073–2103.

Auten, G. and Splinter, D. (Forthcoming). Income inequality in the united states: Using tax data to measure long-term trends. *Journal of Political Economy*.

Battaglini, M., Guiso, L., Lacava, C., and Patacchini, E. (2022). Tax professionals and tax evasion. *Working Paper*.

Bonhomme, S., Holzheu, K., Lamadon, T., Manresa, E., Mogstad, M., and Setzler, B. (2023). How much should we trust estimates of firm effects and worker sorting? *Journal of Labor Economics*, 41(2):291–322.

Boning, W. C., Hendren, N., Sprung-Keyser, B., and Stuart, E. (2023). A welfare analysis of tax audits across the income distribution. Working Paper 31376, National Bureau of Economic Research.

Brounstein, J. (2023). Can countries unilaterally mitigate tax haven usage? evidence from an ecuadorian transaction tax. *Working paper*.

Card, D., Cardoso, A. R., Heining, J., and Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics*, 36(S1):S13–S70.

Card, D., Cardoso, A. R., and Kline, P. (2015). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, 131(2):633–686.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly journal of economics*, 131(2):633–686.

Card, D., Heining, J., and Kline, P. (2013a). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015.

Card, D., Heining, J., and Kline, P. (2013b). Workplace Heterogeneity and the Rise of West German Wage Inequality*. *The Quarterly Journal of Economics*, 128(3):967–1015.

DeBacker, J., Heim, B. T., Tran, A., and Yuskavage, A. (2019). Seeking professional advice: The effects of tax preparers on compliance. *Working Paper*.

Feenberg, D. R. and Coutts, E. (1993). An introduction to the taxsim model. *Journal of Policy Analysis and Management*, 12(1):189–194.

Guyton, J., Langetieg, P., Reck, D., Risch, M., and Zucman, G. (2023). Tax evasion at the top of the income distribution: Theory and evidence. Working Paper 28542, National Bureau of Economic Research.

Kaçamak, Y. (2022). The role of information aggregators in tax compliance. Technical report.

Kline, P., Saggio, R., and Sølvsten, M. (2020a). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Kline, P., Saggio, R., and Sølvsten, M. (2020b). Leave-out estimation of variance components. *Econometrica*, 88(5):1859–1898.

Londoño-Vélez, J. and Ávila Mahecha, J. (2021). Enforcing wealth taxes in the developing world: Quasi-experimental evidence from colombia. *American Economic Review: Insights*, 3(2):131–48.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, 66(4):281–302.

Piketty, T. and Saez, E. (2003). Income inequality in the united states, 1913–1998. *The Quarterly Journal of Economics*, 118(1):1–41.

Piketty, T., Saez, E., and Zucman, G. (2017). Distributional National Accounts: Methods and Estimates for the United States*. *The Quarterly Journal of Economics*, 133(2):553–609.

Saez, E. and Zucman, G. (2016). Wealth inequality in the united states since 1913: Evidence from capitalized income tax data. *The Quarterly Journal of Economics*, 131(2):519–578.

Smith, M., Zidar, O., and Zwick, E. (2022). Top wealth in america: New estimates under heterogeneous returns. *The Quarterly Journal of Economics*, 138(1):515–573.

Zwick, E. (2021). The costs of corporate tax complexity. *American Economic Journal: Economic Policy*, 13(2):467–500.

Table 1: Key income and income tax concepts

Concept	Interpretation	Definition
Broad income	Sum of all labor income and net capital income	B
Tax on broad income	Income tax imputed from broad income	$T^b = T(B; \Theta)$
Gross income	Total income from line 9, Form 1040	
Deductions	Sum of all deductions that apply to taxable income (box x,x, in Form 1040)	D
Credits	refundable and non-refundable credits	C
Taxable income	Income after application of main deductions,Form 1040 line 15	$B - D$
Final tax	Total tax less refundable credits and other tax payments, Form 1040 line 24 – 32	T^f
Implied final tax base	Tax base imputed from final tax	T^{-1}(Final tax)

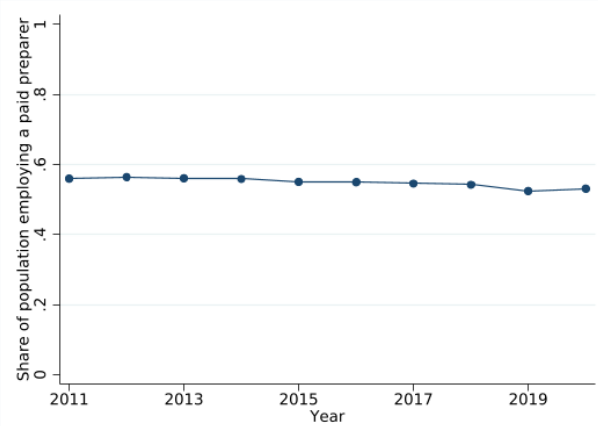
Note: This table lists the key income and income tax benchmarks that we take and define. The specific line items and names correspond with references to the 2022 US Federal Income Tax Form 1040. The "Definition" column presumes the existence of a mapping from tax base to income tax $T(\cdot)$ as well as an inverse mapping from income tax to $T^{-1}(\cdot)$ that also considers filing year and filing status.

Table 2: Key tax and base optimization dependent variable concepts

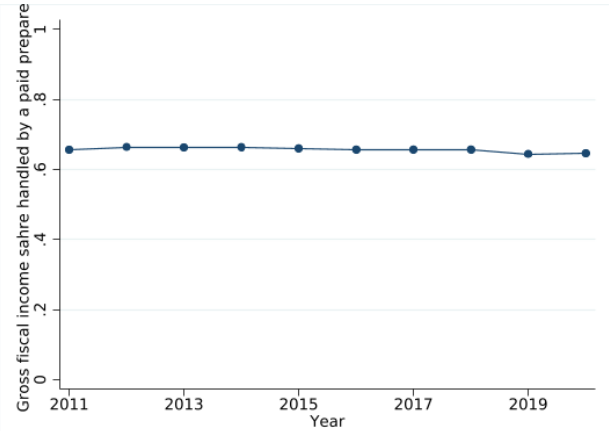
Concept	Interpretation	Definition
Final tax opt.	implied tax optimization between final tax and imputed tax from broad measure	$1 - \frac{T^f}{T^b}$
Deduction optimization	base optimization attributable to deductions	$1 - \frac{B-D}{B}$
Credit optimization	tax optimization attributable to credits	$\frac{C}{T(B-D)}$
Effective tax rate (ETR)	Average tax rate	$\frac{T^f}{B}$
ETR diff	Baseline minus effective tax rate	$\frac{T^b}{B} - \frac{T^f}{B}$

Note: This table lists they key tax optimization concepts and definitions from the paper.

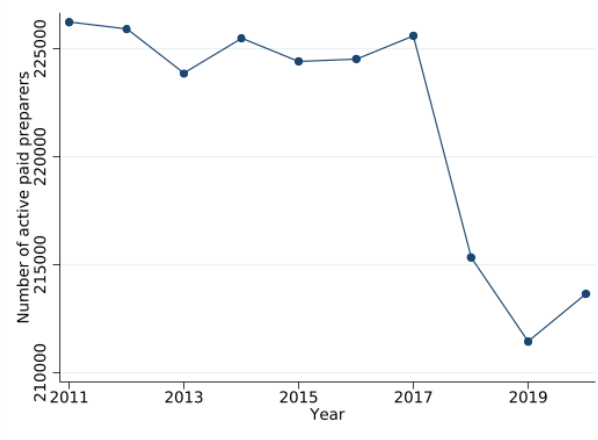
Figure 1: Number and share of paid preparer users over time



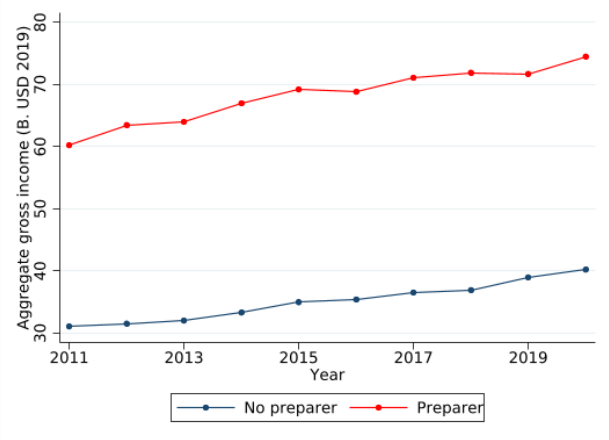
a Share of population using preparers



b Share of income handled by paid preparers



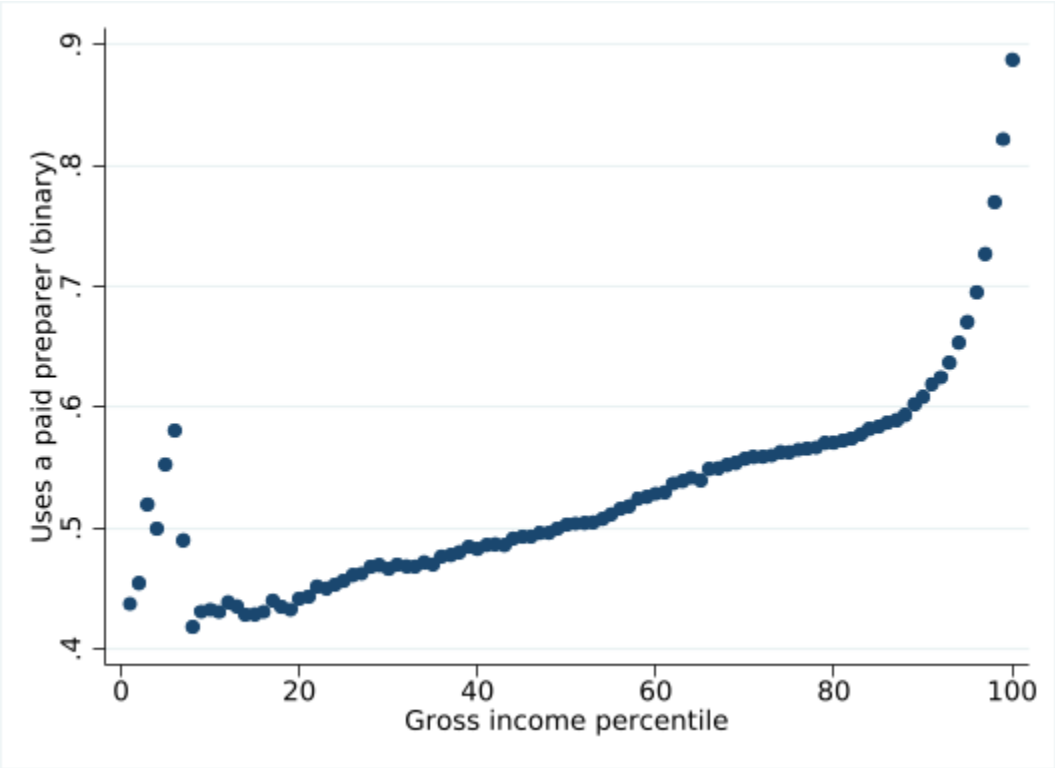
c Number of paid preparers



d Gross income handled by tax preparers

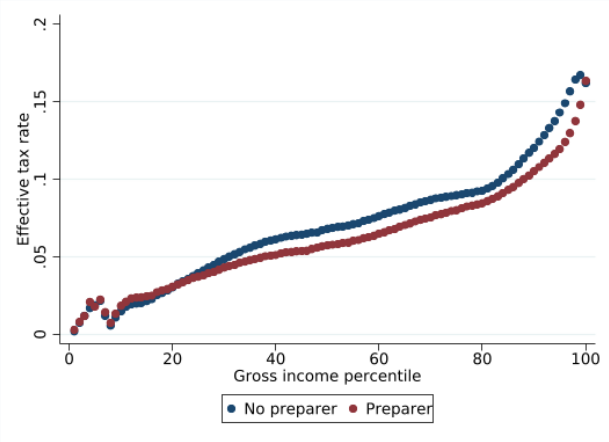
Note: This figure plots the number and shares of taxpayers using tax preparers and income handled by tax preparers over time. These plots are generated from a 1% sample of taxpayers, so aggregate measures should be multiplied by 100.

Figure 2: Use of paid preparers across income distribution

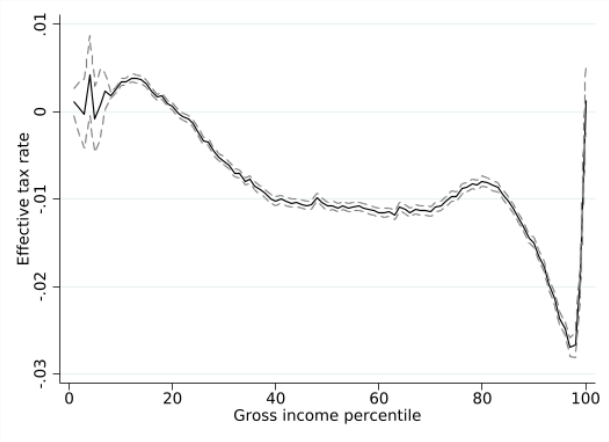


Note: This figure plots the share of population that uses a paid preparer across the income distribution. We bin the population in 100 equal sized bins and show the share of individuals who use a paid preparer in each of those bins.

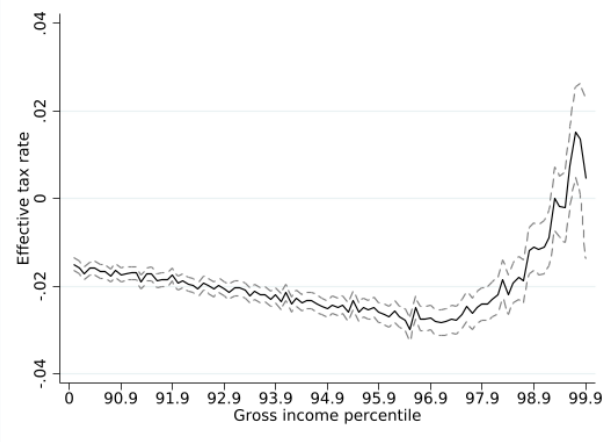
Figure 3: ETRs and taxes paid across income distribution



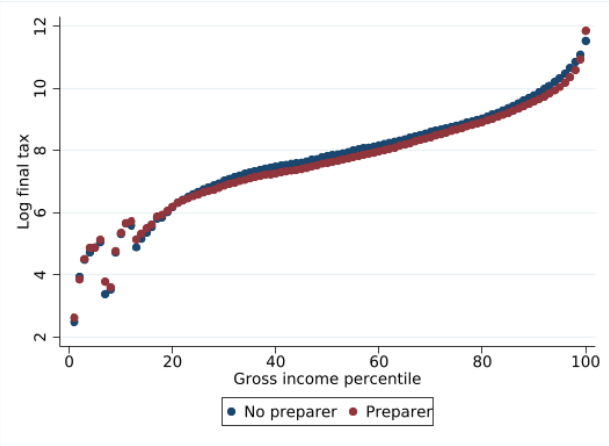
a ETR.



b The difference between ETRs.



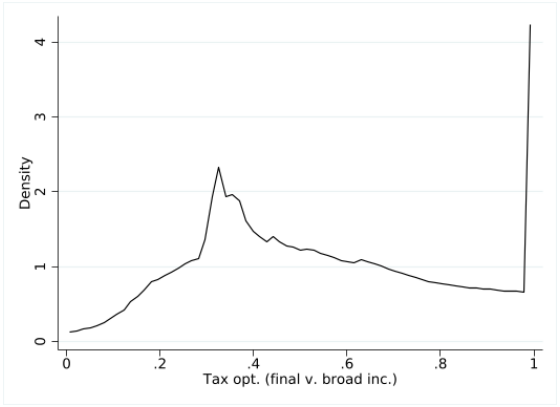
c The difference between ETRs, top 10% income distribution



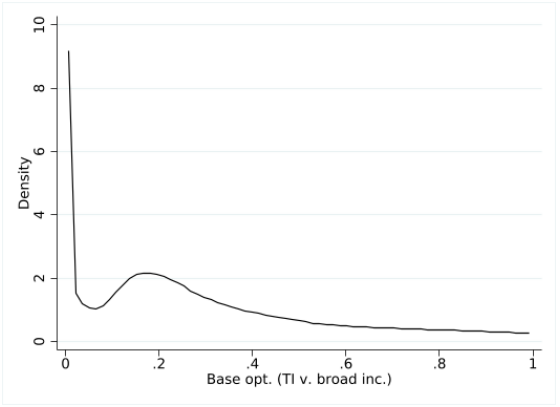
d Log tax.

Note: This figure plots the distribution of ETR and logarithm of final tax paid across the income distribution. In panel A, we plot effective tax rate (ETR) calculated as the ratio of final tax paid relative to the broad income. In red, we plot the averages for each income percentile for taxpayers who use tax preparers, in blue we plot the averages for each income percentile for taxpayers who do not use tax preparers. In panel b, we plot the difference between these two time series across income distribution and in panel c, we zoom in on the top 10 percent of the richest taxpayers. In Panel d, we plot logarithm of final tax for taxpayers who use tax preparers in red and those who do not in blue.

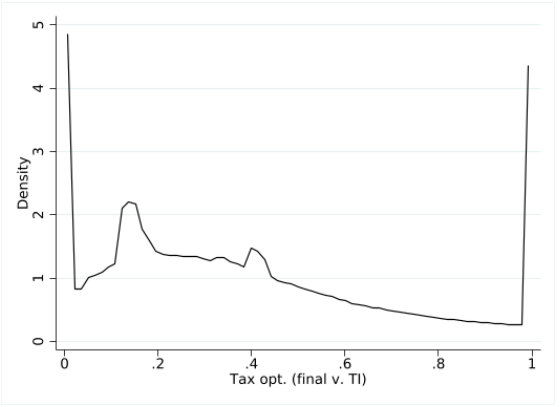
Figure 4: The distribution of main optimization measures and their evolution across years



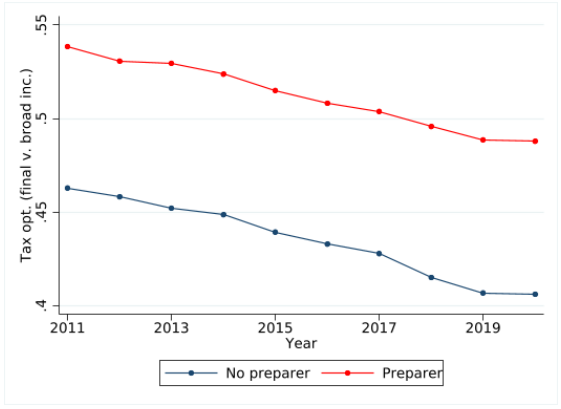
a Final optimization.



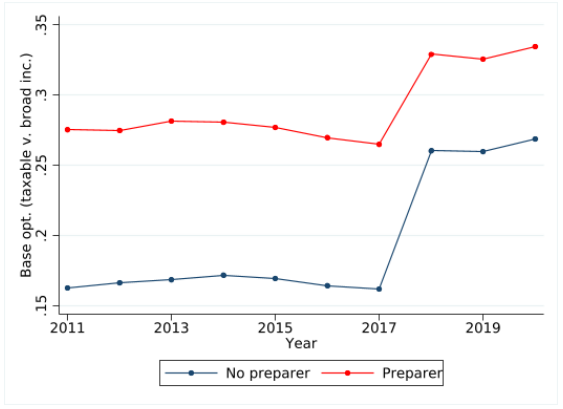
b Deductions optimization.



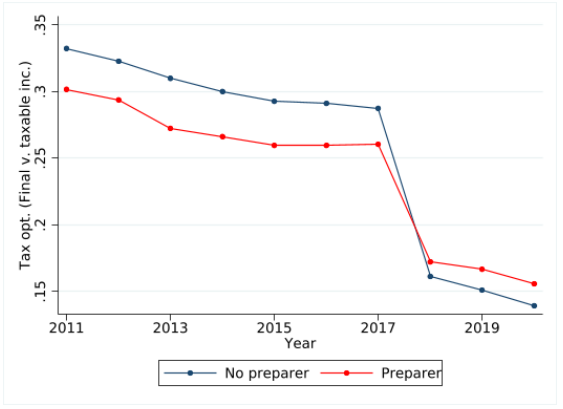
c Credits optimization.



d Final optimization.



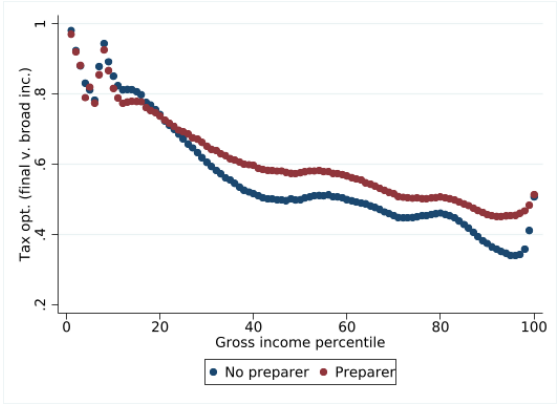
e Deductions optimization.



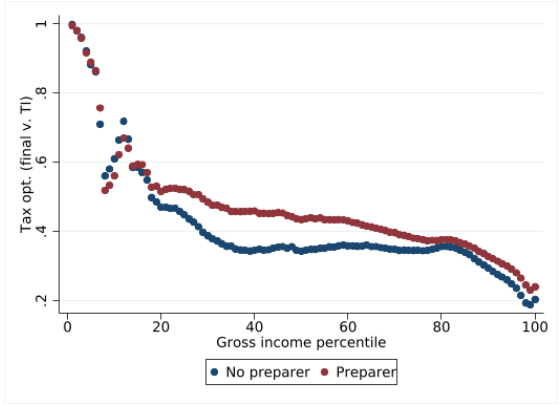
f Credits optimization.

Note: This figure plots the distribution (panels a-c) and changes over time (panels d-f) for our 3 main measures of tax optimization; final optimization, deductions optimization and credits optimization. Each of those measures is defined in Table 2.

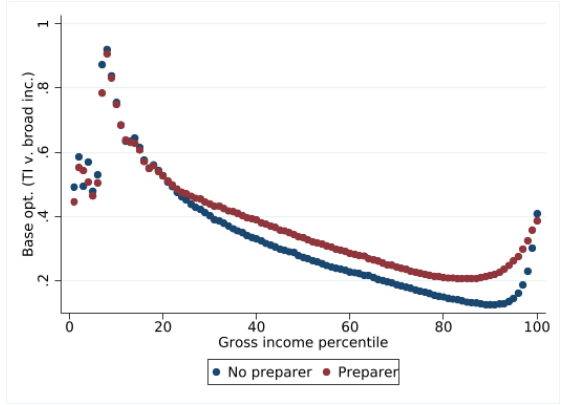
Figure 5: Main optimization measures across income distribution



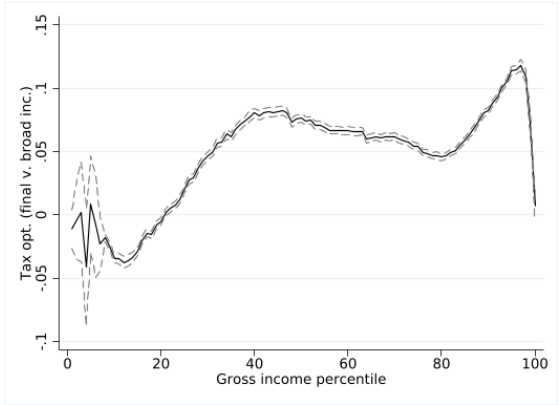
a Final optimization.



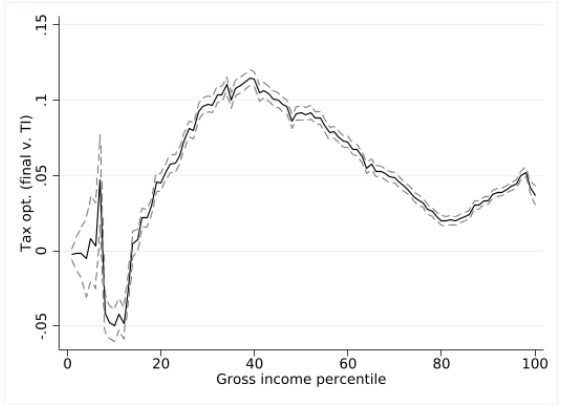
b Credits optimization.



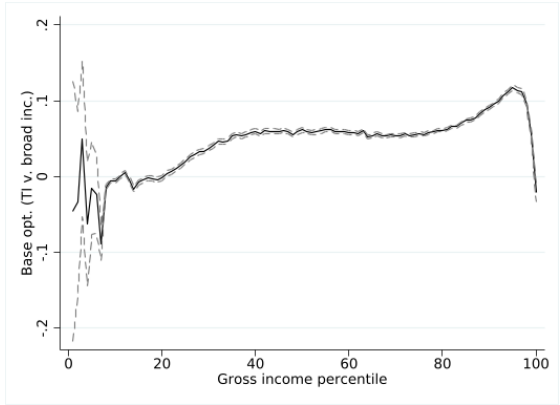
c Deductions optimization.



d Final optimization.



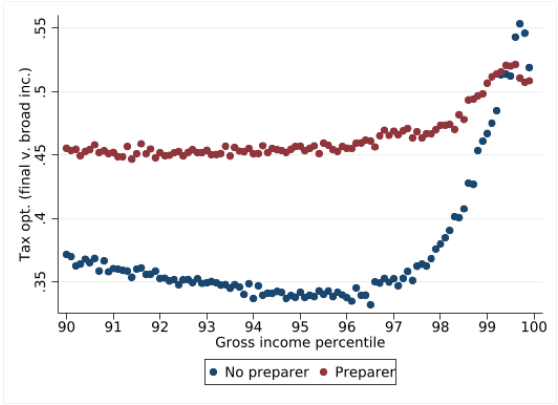
e Credits optimization.



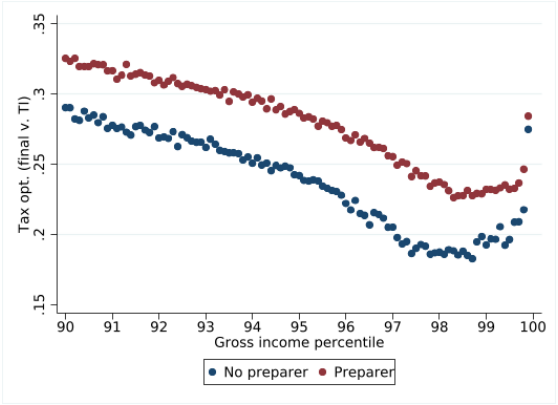
f Deductions optimization.

Note: This figure plots the distribution of our 3 main measures of tax optimization across individual income distribution. Each of those measures is defined in Table 2 and is labelled in each figure subtitle. In Panels a-c, we show the distribution of these measures for taxpayers who use a tax preparer in red and the distribution for taxpayers who do not use a tax preparer in blue. In Panels d-f, we plot the difference between tax optimization between taxpayers who use and do not use a tax preparer.

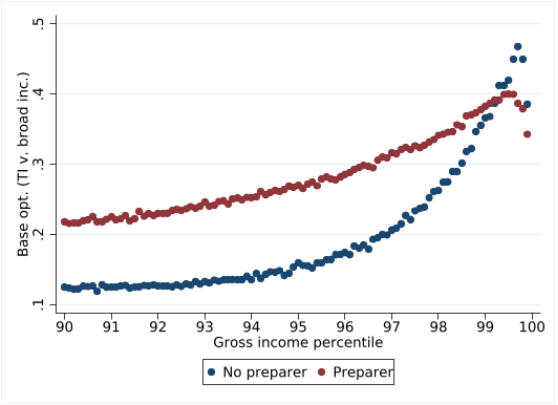
Figure 6: Main optimization measures across the top 10% of income distribution



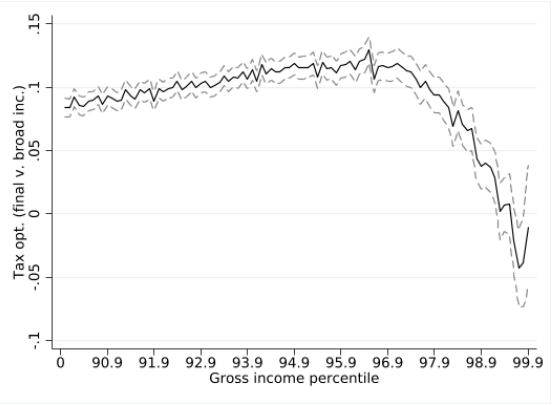
a Final optimization.



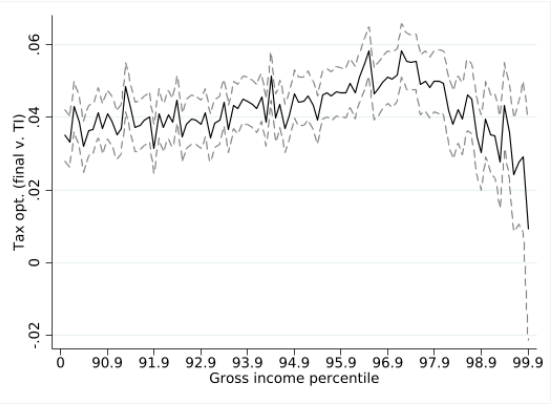
b Credits optimization.



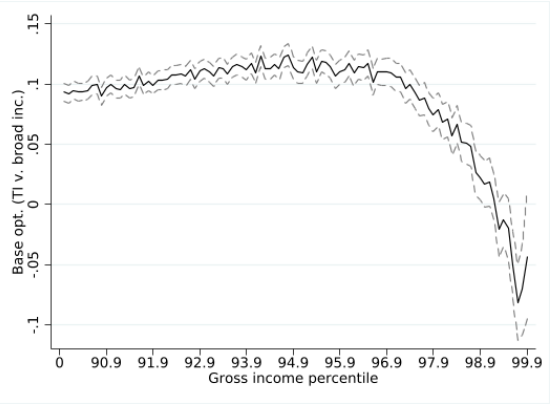
c Deductions optimization.



d Final optimization.



e Credits optimization.



f Deductions optimization.

Note: This figure plots the distribution of our 3 main measures of tax optimization across individual income distribution among the top 10 percent richest taxpayers. Each of those measures is defined in Table 2 and is labelled in each figure subtitle. In Panels a-c, we show the distribution of these measures for taxpayers who use a tax preparer in red and the distribution for taxpayers who do not use a tax preparer in blue. In Panels d-f, we plot the difference between tax optimization between taxpayers who use and do not use a tax preparer.

Table 3: Time with paid preparer

	(1)	(2)		(3)	(4)		(5)	(6)
Years any preparer used			Different preparers used			Years w/ a given preparer		
	Number	share		Number	share		Number	share
1	27,224,467	18%	1	74,433,507	49%	1	159,950,000	52%
2	17,804,359	12%	2	37,625,222	25%	2	48,098,828	16%
3	14,162,471	9%	3	19,001,652	13%	3	26,097,618	9%
4	12,020,545	8%	4	9,798,187	6%	4	17,334,822	6%
5	10,354,087	7%	5	5,162,913	3%	5	12,522,076	4%
6	9,081,268	6%	6	2,744,145	2%	6	9,554,249	3%
7	8,297,658	5%	7	1,429,969	1%	7	7,569,505	2%
8	7,875,522	5%	8	691,448	0%	8	6,308,714	2%
9	8,825,501	6%	9	277,726	0%	9	5,836,362	2%
10	35,585,926	24%	10	71,254	0%	10	13,637,805	4%

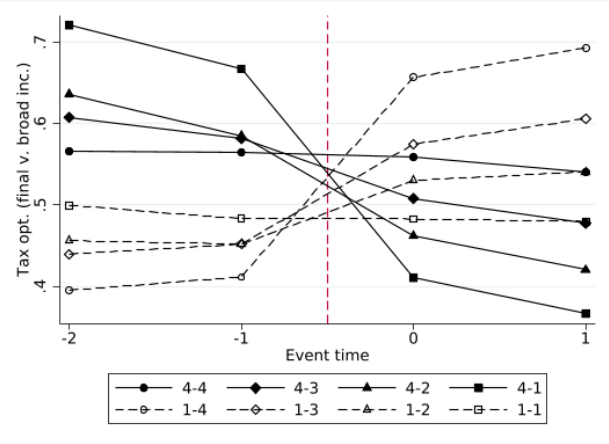
Note: This table summarizes the number and shares of years any paid preparer was used by an individual taxpayer (columns 1 and 2), number of different tax preparers used by an individual (columns 3 and 4), and the number of years with the same tax preparer (columns 5 and 6). We look across all years 2011 - 2020 here.

Table 4: Characteristics of taxpayers and paid preparers.

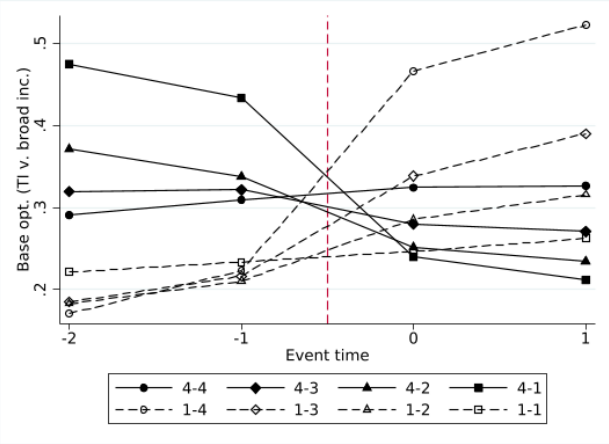
	Panel A: Tax preparers			
	Full sample	Non-movers	Movers	Difference
N	9.589e+06	238822	9.350e+06	
Unique tax preparers	505376	43155	462221	
% CPA	0.241	0.372	0.229	
% missing	0.380	0.240	0.393	
Unique clientele per year	2.558	1.159	2.689	-1.530
	(565.583)	(.517)	(591.397)	(.87)
Gross income	135251	167365	132252	35113
	(5489364)	(958947.9)	(5732410)	(9977.237)
ETR	0.0700	0.0630	0.0710	-0.00700
	(.054)	(.06)	(.054)	(0)
Baseline minus effective tax rate	0.0880	0.0990	0.0870	0.0110
	(.045)	(.05)	(.045)	(0)
Final tax optimization (tax)	0.580	0.639	0.574	0.0640
	(.259)	(.273)	(.257)	(.001)
	Panel B: Tax payers			
N	9.589e+06	238822	9.350e+06	
Unique tax payers	1.712e+06	52376	1.660e+06	
% has paid preparer	46.93	56.32	46.70	9.625
	(17.565)	(18.27)	(17.481)	(.093)
Gross income	122717	259646	119219	140427
	(1684178)	(1780223)	(1681507)	(7688.087)
Taxable income	90601	184234	88209	96026
	(785706)	(1419786)	(762485.1)	(6064.15)
Taxes paid	14702	37588	14118	23471
	(191115.1)	(320359.6)	(186608.5)	(1438.745)
ETR	0.0780	0.0840	0.0770	0.00700
	(.054)	(.067)	(.053)	(0)
Baseline minus effective tax rate	0.0810	0.0970	0.0810	0.0160
	(.043)	(.056)	(.043)	(0)
Final tax optimization	0.534	0.560	0.533	0.0270
	(.254)	(.277)	(.253)	(.001)

Note: This table reports the summary statistics of tax preparers (Panel A), and of tax payers (Panel B). Column (1) reports the statistics over the full sample of tax preparers and tax payers respectively, column (2) over the sub-sample of individuals who are excluded from the fixed effects estimations and column (3) over the sub-sample of movers - the largest connected set of taxpayers and tax preparers. Movers are individuals who switch tax preparer or start using one during the sample period. Column (4) reports the difference between columns (1) and (3).

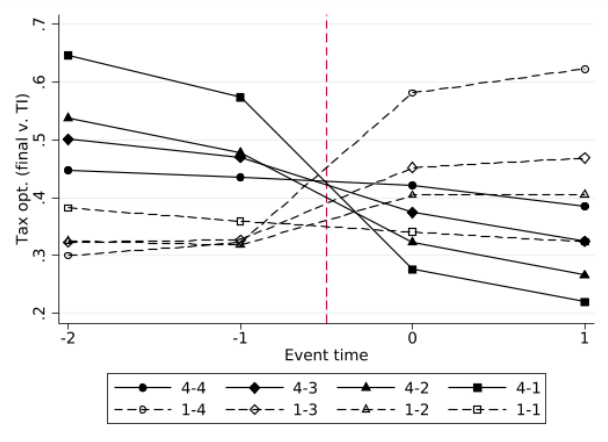
Figure 7: Mean tax optimization for taxpayers that switch tax preparer by quartiles of Changes in tax preparer fixed Effects.



a Final optimization



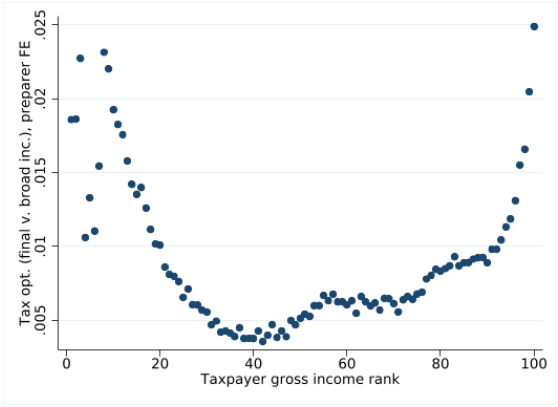
b Deductions optimization



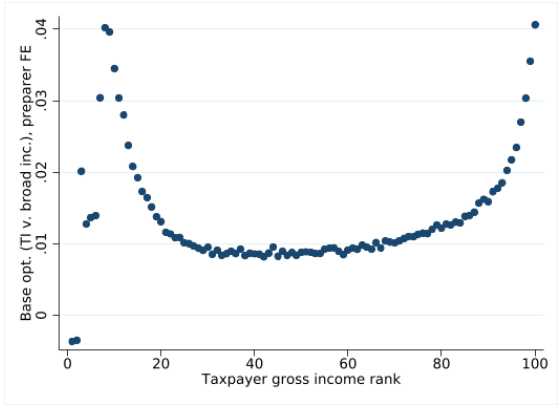
c Credits optimization

Note: This figure plots event studies for individuals who switch a tax preparer across our 3 main measures of tax optimization. Each of those measures is defined in Table 2 and is labelled in each figure subtitle. We show switches between tax preparers with different levels of effectiveness, where 4 is the top quartile of the effectiveness distribution, while 1 is the bottom quartile. 1-4 means a switch from the least effective to the most effective tax preparer.

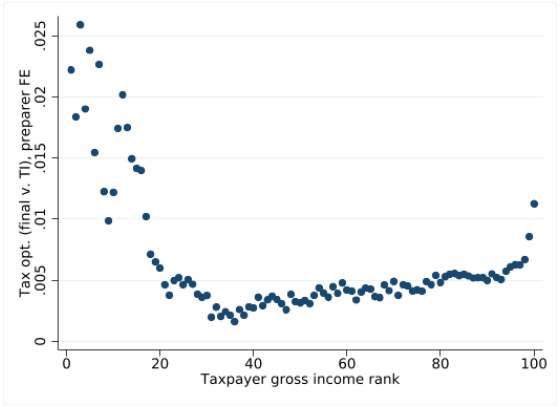
Figure 8: AKM results taxpayer and tax preparer FEs by income rank.



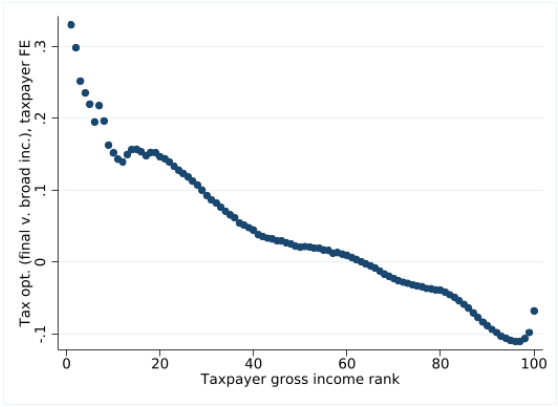
a Final optimization



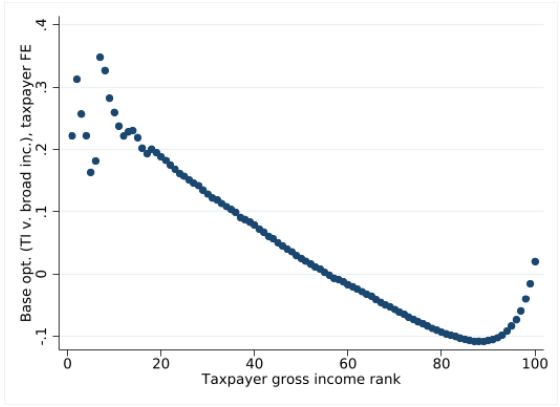
b Deductions optimization



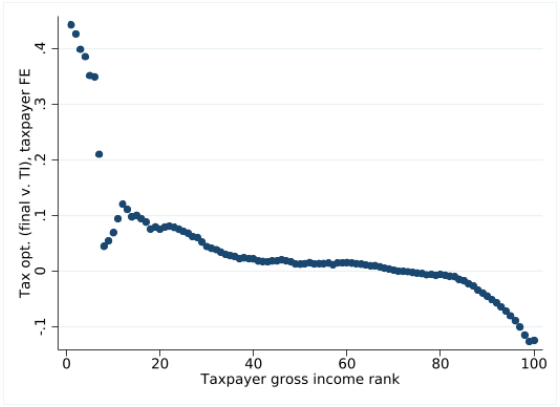
c Credits optimization



d Final optimization



e Deductions optimization



f Credits optimization

Note: This figure plots the distribution of fixed effects from our main estimation across the income distribution of taxpayers for each tax optimization measure. Each of those optimization measures is defined in Table 2 and is labelled in each figure subtitle. In Panels a-c, we plot tax preparer fixed effects and in panels d-f, we plot taxpayer fixed effects.

Table 5: Variance decomposition

Panel A: Tax optimization measures						
	Final Opt.		Deductions Opt.		Credits Opt.	
	Component	Share	Component	Share	Component	Share
Variance of taxpayer effect	0.0456	71.90	0.0461	74.10	0.0558	69.31
Variance of preparer effect	0.00920	14.48	0.00880	14.14	0.0141	17.51
Variance of state x year effect	0.00100	1.600	0.00220	3.565	0.00540	6.710
Variance of residual	0.0211	33.24	0.0172	27.65	0.0263	32.63
2cov(taxpayer, preparer)	-0.0130	-20.46	-0.0120	-19.27	-0.0205	-25.44
2cov(state x year effect, taxpayer + preparer)	-0.000500	-0.752	-0.000100	-0.181	-0.000600	-0.723
Panel B: Post-tax income and tax liability						
	log tax payments		log post-tax income			
	Component	Share	Component	Share		
Variance of taxpayer effect	1.969	82.93	0.703	85.77		
Variance of preparer effect	0.286	12.05	0.0649	7.927		
Variance of state x year effect	0.0409	1.724	0.0182	2.216		
Variance of residual	0.503	21.18	0.105	12.88		
2cov(taxpayer, preparer)	-0.382	-16.07	-0.0565	-6.897		
2cov(state x year effect, taxpayer + preparer)	-0.0430	-1.810	-0.0155	-1.894		

Note: This table outlines each component of variance decomposition and the share that each component contributes. In Panel A, we perform variance decomposition for each of the 3 main tax optimization measures. Each of those optimization measures is defined in Table 2 and is labelled in each column title. In Panel B, we use as outcome variable logarithm of tax payments and logarithm of post-tax income and perform the same variance decomposition as in Panel A.

Appendices

A Additional Results

Table A1: Comparison of individual tax optimization measures for individuals with and without tax preparers.

	No preparer	Preparer	Dif.	Dif. (controls)
Gross income	84336	147564	63227	-13508
	(6805045)	(2083121)	(3841.655)	(11073.08)
AGI	73990	118637	44648	-5688
	(661543.7)	(888153.1)	(1269.972)	(2405.617)
Taxable income	65338	110548	45210	-5321
	(661677.7)	(888518.1)	(1270.492)	(2405.678)
ETR	0.0750	0.0770	0.00200	-0.00900
	(.049)	(.056)	(0)	(0)
Final tax optimization	0.460	0.485	0.0240	0.0630
	(.264)	(.27)	(0)	(0)
No. individuals	1.026e+06	1.087e+06		
Taxpayer-years	4.447e+06	5.001e+06		

Note: This table presents summary statistics on US individual taxpayers from 2011 to 2019. For variables definition, refer to Table 1.

Table A2: Summary Statistics on Taxpayers

	Mean	SD	N	Missing
Age	46.93	17.57	9.555e+06	33299
Gross income	122717	1.684e+06	9.589e+06	0
AGI	98946	785420	9.589e+06	0
Taxable income	90601	785706	9.589e+06	0
Taxes paid (pre-credit)	14762	191112	9.589e+06	0
Taxes paid	14702	191115	9.589e+06	0
Final base opt.	0.474	0.264	9.589e+06	0
Final tax opt.	0.534	0.254	9.589e+06	0
Tax savings	13969	545303	9.589e+06	0
Log tax savings	8.447	1.139	9.589e+06	10
Baseline ETR	0.159	0.0460	9.589e+06	0
Final ETR	0.0780	0.0540	9.589e+06	0
ETR optimization	0.0810	0.0430	9.589e+06	0
Base opt. (AGI)	-0.0330	39.59	9.589e+06	0
Base opt. (TI)	0.297	0.253	9.124e+06	464950
Tax opt. (TI)	0.382	0.287	8.342e+06	1.247e+06
Tax savings (TI)	6052	145255	8.342e+06	1.246e+06
Log tax savings (TI)	7.213	2.278	8.326e+06	1.263e+06
ETR (TI)	0.105	0.0680	9.571e+06	17288
ETR opt. (TI)	0.0540	0.0380	8.343e+06	1.246e+06
Log credits	1.781	3.151	6.802e+06	2.787e+06
Log credits (alt.)	1.906	3.276	6.802e+06	2.787e+06

Note: This table presents summary statistics on US individual taxpayers from 2011 to 2019. In addition to the mean (reported in the first column), the table reports distributional information across several percentiles. For variables definition, refer to Table 1.

Table A3: Summary Statistics on Professional Preparers

_	Mean	SD	p5	p25	p50	p75	p95	N
% Female	0.296							
% College	0.139							
Year of birth	1963	15.70						477120
AGI	107019	221060						481400
Fraction of panel active	0.439	0.334	0.100	0.100	0.300	0.700	1	505375
Distinct clientele (average over active years)	1.026	0.833	0.222	0.500	1	1	2	505375
Client tenure probability	0.619	0.298	0.167	0.357	0.563	1	1	282267
Probability repeat client from last year	0.662	0.328	0	0.500	0.750	0.933	1	355338
Probability repeat client next year	0.667	0.327	0	0.500	0.769	0.933	1	355338
Distinct preparers	505375							

Note: This table presents summary statistics on professional tax preparers from 2011 to 2019. In addition to the mean (reported in the first column), the table reports distributional information across several percentiles.

Table A4: Summary Statistics on Preparers' Clientele

_	Mean	SD	N
Number of annual clientele	2.373	2.238	2.216e+06
Clientele gross income	167747	2.846e+06	2.216e+06
Clientele AGI	127744	933752	2.216e+06
Clientele tax obligation	22084	247628	2.216e+06
Clientele baseline effective tax rate	0.166	0.0460	2.216e+06
Clientele final effective tax rate	0.0780	0.0490	2.216e+06
ETR optimization	0.0870	0.0400	2.216e+06
Clientele tax savings	21971	993427	2.216e+06
Final tax optimization	0.549	0.216	2.216e+06
Final base optimization	0.488	0.223	2.216e+06
Log credits	2.292	3.050	1.881e+06
Log credits (alt.)	2.439	3.153	1.881e+06
Unique preparers	505375		
Preparer-years	2.216e+06		

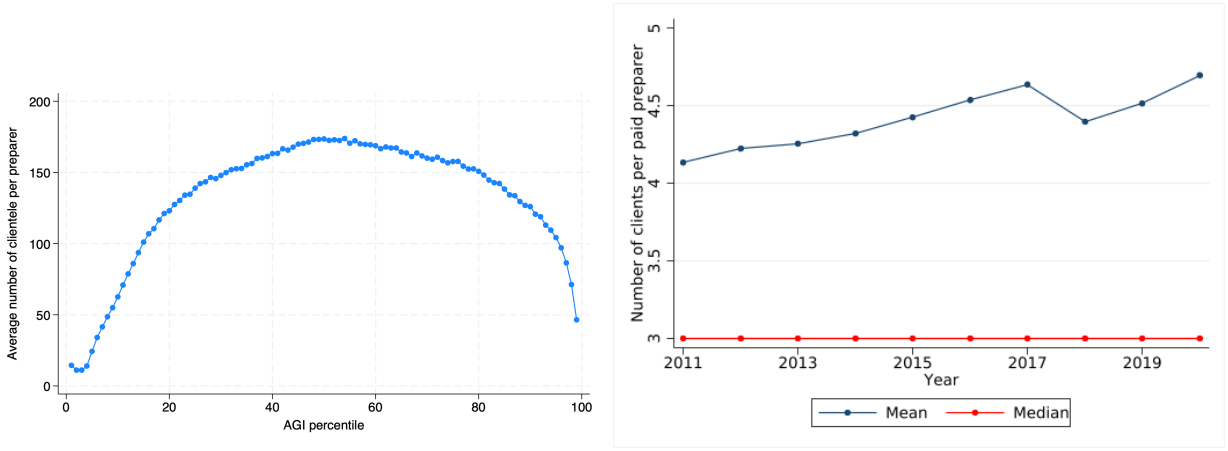
Note: This table presents descriptive information on preparers' clientele from 2011 to 2019. In addition to the mean (reported in the first column), the table reports distributional information across several percentiles. For variables definition, refer to Table 1.

Table A5: Summary Statistics on Preparers' Clientele by Professional Category

	Attorney	CAA	CPA	EA	SRTP
Number of annual clientele	2.194	3.855	2.858	3.569	3.819
	(2.299)	(4.081)	(3.117)	(3.537)	(4.315)
AGI	104241	41822	119271	75310	55548
	(108435.1)	(38091.86)	(106934.6)	(66638.44)	(48718.19)
Tax obligation	16013	2413	19050	9091	4837
	(25056.36)	(7226.771)	(25195.08)	(14491.44)	(9929.234)
Average tax rate	0.086	0.048	0.092	0.075	0.059
	(.072)	(.041)	(.065)	(.053)	(.048)
Tax opt. (total)	0.499	0.624	0.492	0.514	0.584
	(.437)	(.306)	(.371)	(.345)	(.328)
Base opt. (total)	0.471	0.594	0.456	0.479	0.549
	(.387)	(.298)	(.331)	(.313)	(.309)
Ded. + cred. inc. share	0.399	0.697	0.396	0.397	0.487
	(.953)	(1.137)	(.875)	(.748)	(.844)
Distinct preparers	7525	3310	131387	47642	23390
Preparer-years	37524	20100	824691	302201	134648

Note: This table disaggregates paid preparer clientele characteristics by paid preparer type. CAA refers to "Certified Acceptance Agent". CPA refers to "Certified Public Accountant"; EA refers to "enrolled agents". SRTP refers to "Supervised Registered Tax Preparer". Standard errors are clustered on the taxpayer level.

Figure A1: Tax preparer networks



a Average Tax Preparer Network by Client AGI b Number of clients per tax preparer

Note: This figure plots the mean number of clients of paid preparers along AGI rank of paid preparer users. For example among paid preparer users of 80th percentile AGI, paid preparers have on average 150 clients per year.

Table A6: Taxpayers x Preparer Usage

_	Mean	SD	p5	p25	p50	p75	p95	N
Fraction of years with preparer	0.516	0.449	0	0	0.500	1	1	1.712e+06
Number unique preparers	1.740	1.150	1	1	1	2	4	1.712e+06
Raw years per preparer	2.163	1.421	0.500	1	2	3	5	1.712e+06
Years per preparer (conditional on usage)	1.774	2.494	0	0	1	2.250	9	1.712e+06

Note: This table presents descriptive information on preparers usage by taxpayers from 2011 to 2019. In addition to the mean (reported in the first column), the table reports distributional information across several percentiles. For variables definition, refer to Table 1.