A Welfare Analysis of Tax Audits Across the Income Distribution

William C. Boning, U.S. Department of the Treasury
Nathaniel Hendren, MIT and Policy Impacts
Ben Sprung-Keyser, University of Pennsylvania and Policy Impacts
Ellen Stuart, The University of Sydney and Policy Impacts

June 2024

Abstract

We estimate the returns to IRS audits of taxpayers across the income distribution. We find an additional \$1 spent auditing taxpayers above the 90th income percentile yields more than \$12 in revenue, while audits of below-median income taxpayers yield \$5. We construct our estimates by drawing upon comprehensive internal accounting information and audit-level enforcement logs. We begin by estimating the average initial return to all audits of US taxpayers filing in tax years 2010–2014. On average, \$1 in audit spending initially raises \$2.17 in revenue. Audits of high-income taxpayers are more costly, but the additional revenue raised more than offsets the costs. Audits of the 99-99.9th percentile have a 3.2:1 initial return; audits of the top 0.1% return 6.3:1. We then exploit the 40% audit reduction between tax years 2010 and 2014 to examine the returns to marginal audits. We find they exceed the returns to average audits. Revenues remain relatively unchanged but marginal costs fall below average costs due to economies of scale. Next, we use randomly selected audits to examine the impact of an initial audit on future revenue. This individual deterrence effect produces at least three times more revenue than the initial audit. Deterrence effects are relatively consistent across the income distribution. This results in the 12:1 return above the 90th percentile. We conclude by estimating the welfare consequences of audits using the MVPF framework and comparing audits to other revenue raising policies. We find that audits raise revenue at lower welfare cost.

We gratefully acknowledge research assistance from Jamie Emery and Sethu Odayappan. We are also grateful for feedback from conference and seminar participants at the National Tax Association, the Office of Tax Analysis, the 2022 and 2023 ASSA meetings, and the NBER TAPES 2022. Additionally, we thank several people at IRS including Maria Arias-Buchanan, Eric Baltas, Astin Cornwall, Pat Langetieg, Kyle Richison, Ben Schwartz, Alex Turk, and especially John Guyton for their comments, suggestions, and assistance with understanding the audit data and IRS procedures. This research was supported by Policy Impacts, which receives funding from the Spiegel Family Fund. This research was conducted while an author was an employee at the U.S. Department of the Treasury. 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. Any taxpayer data used in this research was kept in a secured IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

IRS estimates suggest that more than $500 billion in tax obligations go unpaid each year (IRS, 2022). Those unpaid tax liabilities are concentrated among taxpayers at the top of the income distribution. For example, the top 10% of earners may owe more than 60% of all unpaid tax liabilities (DeBacker et al., 2020; Guyton et al., 2021; Johns and Slemrod, 2010). Evidence indicates that, on average, tax audits can recoup unpaid taxes and raise revenue in the process. There is, however, very limited evidence on the returns to tax audits across the income distribution (Holtzblatt and McGuire, 2016, 2020).

Do audits of high-income taxpayers generate more revenue per dollar spent on tax enforcement? Does the increased complexity of auditing those high-income taxpayers reduce the "bang for the buck" on enforcement spending? In this paper, we provide a detailed analysis of the returns to in-person tax audits of individuals across the income distribution. Our analysis proceeds in four steps.

We begin by estimating the average costs and average revenue raised from in-person audits.[1] In order to construct the cost and revenue estimates, we rely on two sources of data. First, we use a comprehensive database that tracks enforcement activities by IRS personnel on all audits conducted since 2003. The database contains the revenue raised from each step of each audit. It also contains activity logs recording the time spent on each audit by each IRS employee. When combined with the General Scale (GS) classification for each employee, this allows us to estimate the hourly costs of all direct enforcement activities. Second, we utilize detailed internal IRS business unit accounting information. This provides us with a comprehensive picture of other IRS costs beyond the direct labor cost of hours spent conducting audits. It includes additional labor costs such as wages for non-auditing hours, the cost of management, and fringe benefits of employees. It also includes central overhead costs such as building and technology service costs. This information has, to the best of our knowledge, not been used in academic studies of tax enforcement. We show that this data is, nonetheless, critical to construct accurate measures of the total cost of enforcement activity. The wage costs of hours spent by auditors on direct enforcement activity is only about 20% of the total cost of conducting audits.

[1] We restrict our analysis to focus on in-person audits rather than audits conducted via correspondence. We do so because data limitations prevent a comprehensive analysis of the returns to correspondence audits. In particular, we are unable to measure the marginal costs of correspondence audits or capture the individual deterrence effects of correspondence audits, both of which we measure for in-person audits. We do, however, measure the average returns to correspondence audits and show that average returns to correspondence audits across the income distribution have patterns that are qualitatively similar to our findings for in-person audits.

We combine these measures of audit costs and audit revenue to estimate the average return to audit expenditures. We estimate that each $1 spent on an audit returns an average of $2.17 in initial revenue. We then merge this information with income reported on individual tax returns to study how these costs and revenues vary across the income distribution. We find that audits of higher income taxpayers are more time intensive and more costly than audits of low-income taxpayers. For example, audits of taxpayers with incomes in the bottom 50% of the income distribution cost an average of $5,218 while audits of taxpayers in the top 1% and 0.1% cost an average of $11,382 and $15,170, respectively.[2] That said, these rising costs across the income distribution are more than offset by increasing revenues. We estimate that audits of taxpayers in the bottom 50% of the income distribution produce $0.96 in revenue for each dollar of audit cost. Audits of the top 1% produce $4.25 in revenue and audits of the top 0.1% produce a return of $6.29 for each dollar of audit cost.

In the second step of our analysis, we estimate the costs and revenue associated with a marginal audit expansion. In order to estimate marginal revenues, we exploit the fact that there was a 40% decline in U.S. audit rates for tax returns filed between 2010 and 2014. We examine how revenues changed during these steep cuts in order to assess the returns to reversing them. If audit selection decisions during this time prioritized reducing audits with low rates of return, we would expect that the decline in audits would be associated with an increase in revenues from an average audit. In fact, we see no such pattern. Revenues per hour of auditing remained stable during this period. We show that this is driven by a consistent decline in audit rates across the income distribution and that the audit revenues remained stable within income groups. These patterns align with the IRS's approach of maintaining a balanced portfolio of audits across a range of non-revenue-focused criteria. The patterns suggest that expanding audit rates from 2014 back to their 2010 levels could deliver revenues per audit similar to the average revenue figures we estimate.

We also consider the costs associated with marginal audit expansions. It is natural to expect some economies of scale in the non-direct costs of conducting audit activities. We exploit the detailed nature of our business unit accounting information as well as information from existing IRS budget requests to estimate the magnitude of fixed versus variable costs. The evidence suggests that, in the case of a major audit expansion, approximately 27% of total costs are likely to be fixed.

[2] In particular, audits of taxpayers in the bottom 50% of the income distribution take an average of 28 hours and are conducted by auditors who earn $37 per hour on average. This leads to $1,020 in labor costs for direct audit hours and total costs of $5,218 once non-audit labor and non-labor overhead costs are incorporated. For taxpayers in the top 0.1%, audits take an average of 65 hours and are conducted by auditors who earn an average of $46 per hour. That yields $2,948 in labor costs for direct audit hours and $15,170 in total costs.

Consequently, we expect a marginal audit expansion in the bottom 50% of income would produce \$1.31 in direct revenue for each dollar of audit cost. Audits of the top 1% would return \$5.82 and audits of the top 0.1% would return \$8.62.

In the third part of our analysis, we study how audits may raise revenue in an indirect manner by deterring noncompliance. We focus on one particular form of deterrence known as individual (or specific) deterrence. This refers to a case in which auditing an individual in one year increases taxes paid in subsequent years. We build on earlier work by DeBacker, Heim, Tan, and Yuskavage (2018, henceforth DHTY), who use randomly selected audits as part of the IRS' National Research Program (NRP) to estimate the magnitude of individual deterrence in the US. We extend their analysis by utilizing an additional decade of post-audit data and implementing a stratified matched-control strategy. These extensions allow us to estimate long-run deterrence effects and the heterogeneity in these effects across the income distribution.

We show that audits lead to an increase in future taxes paid that persists over the 14 years we observe in the data. In present discounted value, the payment of these additional taxes raises 3.2 times the revenue raised from the initial audits. We also show that these deterrence effects are similar across the full income distribution (although we lose the power necessary to estimate precise effects in the top 1%).[3] When we combine these deterrence figures with our estimates of the direct returns to marginal audits, we find that auditing individuals in the bottom half of the income distribution produces a return slightly above 5:1. By contrast, audits of individuals in the 90–99th percentiles produce a return of 12:1.[4]

In the fourth and final step of our analysis, we assess the welfare consequences of tax audits across the income distribution. We do so by deriving and estimating the Marginal Value of Public Funds (MVPF) of a change in audit rates. The MVPF of additional tax audits is given by the ratio of equals taxpayers' willingness to pay to avoid the audits divided by the net revenue raised. This measures the welfare cost imposed per dollar of government revenue raised, and can be compared to other methods of raising revenue such as changes in tax rates.

In order to estimate the MVPF of audits, one needs to not only incorporate the revenue raised

[3]This result differs from a conclusion in DHTY that suggests audits do not have positive individual deterrence effects in the top income quintile. As we discuss in Section 5, the absence of year-since-audit fixed effects in DHTY's event study produced spuriously negative deterrence effects at the top of the income distribution that are attributable to mean reversion. Adjusting for that omission results in deterrence effects that are consistent across the income distribution. We are deeply grateful to Alex Yuskavage and the whole DHTY team for sharing their code and providing assistance in conducting these comparisons. We note that including the omitted fixed effects does not substantially alter their conclusions when examining average deterrence effects across the full income distribution.

[4]If we assume that our estimated 3.2X deterrence effect also applies to taxpayers in the top 1%, then the returns to audits at the 99–99.9th percentiles would be 18.2:1 and the return in the top 0.1% would be 36.0:1.

per dollar spent on tax audits, but also include the burden imposed on taxpayers during an audit. We draw upon new survey evidence of recently audited taxpayers that measures the time and money they spent to comply with the audit, and captures how this varies across the income distribution.[5] On average, those in the top 10% (top 1%) of the income distribution spend around 35 (30) hours and $1,500 ($2,000) complying with an audit. While this is a meaningful burden, it is substantially smaller than the revenue collected.[6]

With these inputs, we estimate that additional audits of taxpayers in the bottom half of the income distribution impose a welfare cost of around $1.30 per dollar of net government revenue raised. In contrast, additional audits of taxpayers in the 90–99th percentiles of the income distribution only impose a welfare cost of $1.15 per dollar of net government revenue raised. This pattern lies in stark contrast to MVPF estimates for tax changes across the income distribution, where tax increases of high-income earners typically have much larger MVPFs than tax increases on low-income earners. We also show how the MVPF framework readily incorporates concerns for horizontal equity, which is violated when individuals evade their taxes.[7] The MVPF framework enables one to capture this concern by placing lower social welfare weights on noncompliant taxpayers. We operationalize this idea by exploring the differential burden of audits on compliant versus noncompliant taxpayers. We do so by calculating the fraction of audits that result in a change in assessed tax liability and combining that with new survey data that separately reports the average burden of audits among compliant and noncompliant taxpayers. We find that nearly all of the costs imposed by marginal audits are borne by noncompliant taxpayers. While determinations of policy optimality depend on the welfare weights placed on taxed or audited individuals, this result suggests that increasing audit rates raises revenue at relatively low welfare cost.

Related literature As noted above, a recent body of work calculates the average returns to audits (CBO, 2018, 2020; Holtzblatt and McGuire, 2016, 2020). There is also a robust literature on the size of the tax gap across the income distribution (DeBacker et al., 2020; Guyton et al., 2021; Johns and Slemrod, 2010).[8] There is, however, comparatively little empirical work quantifying the

[5]Our analysis builds upon earlier work by Guyton and Hodge (2014), which reports the average burden across all audited taxpayers and fits a quadratic function in income to model burdens by income.

[6]We also use the MVPF framework to show that a planner who places zero social welfare weight on noncompliant taxpayers would need to assume that audits place at least $300,000 of burden on compliant taxpayers to prefer higher tax rates versus expanded tax audits.

[7]Horizontal equity requires that all taxpayers with the same level of true income pay the same amount in taxes.

[8]Alstadsæter et al. (2019) show that evasion in unreported offshore accounts by tax residents of Norway, Sweden, and Denmark is increasing in income. In particular, they find that the probability of holding offshore accounts increases with income, while, conditional on ownership, the share of income held in such accounts is roughly constant

costs and benefits of tax enforcement across the income distribution.[9] Sarin and Summers (2019) provide a figure for the returns to auditing taxpayers with more than $5 million in income, but their analysis is meant simply as an illustrative calculation rather than a full accounting of costs and revenues across the income distribution.[10]This paper seeks to fill this gap by providing detailed estimates on the returns to audits across the income distribution.

In estimating the total revenue raised by individual audits, we extend upon an existing literature measuring the individual deterrence effects of audits. Our core contribution here is to identify the long-run effects of audits and study how individual deterrence effects vary across the income distribution. We build most closely on DeBacker et al. (2018) by extending our analysis to include additional years of data and by using a matching strategy to allow comparisons within fine-grained income groups.[11] Our work also relates to the large body of evidence on deterrence measured outside the United States (Bjørneby et al., 2021; Hebous et al., 2020; Mazzolini et al., 2022; Kasper and Alm, 2022a,b; Kleven et al., 2011; Advani et al., 2023; Best et al., 2021; Løyland et al., 2023). For example, Kleven et al. (2011) estimate deterrence effects for a single post-audit study year in Denmark and Advani et al. (2023) estimate deterrence effects in the UK up to 8-years post-audit. These papers find per-year deterrence effects between 20% and 35% of initial audit revenue. These results are broadly consistent with our estimates, though we find that the effects persist for at least 14 years.

Our paper also relates to the large literature on the distortionary effects of raising tax revenue. There is an extensive theoretical and empirical body of work quantifying the costs of raising revenue through changes in tax rates. This literature highlights the importance of analyzing heterogeneity in the distortionary cost of taxation across the income distribution (E.g. Saez 2001; Kleven and Kreiner 2006).[12] While that work focuses on optimal tax rates, rather than optimal levels of tax audits, the same basic logic applies. Differences in the returns to audits across the income distribution

across the wealth distribution.

[9]Recent work has also examined the incidence of audit selection algorithms across White and Black taxpayers and also shows audits of Black taxpayers raise less revenue per audit on average (see, e.g., Elzayn et al. (2023)).

[10]In Appendix C, we provide a more detailed comparison between our estimates and those of Sarin and Summers (2019) and with the estimates of average returns provided in Holtzblatt and McGuire (2020). We also note that our estimates relate to recent calculations by CBO on how the return to IRS spending would differ if audits were restricted to taxpayers earning over $400,000 per year. While the report does not publicly provide information on the returns to auditing taxpayers across the income distribution, those figures seem to be an input in the calculation. See `https://www.cbo.gov/system/files/2022-08/58390-IRS.pdf`.

[11]Beer et al. (2020) also examine short-run deterrence effects among self-employed US taxpayers using non-random operational audits, finding results broadly consistent with our estimates.

[12]A progressive social planner would find, at an optimum, that it is costly to raise revenue from high-income taxpayers and less costly to raise revenue from low-income taxpayers (Mirrlees (1976)). This corresponds to higher MVPFs for policy changes that raise revenue from high-income taxpayers.

correspond to differences in the distortionary costs of audits. Our estimates of the returns to tax audits across the income distribution allow for the estimation of those distortionary effects. Our discussion of the MVPF of tax audits also helps to formalize the parallel with tax rates.

Relatedly, our MVPF approach connects to a theoretical literature on optimal tax administration (see, e.g., Kaplow, 1990; Mayshar, 1991; Slemrod and Yitzhaki, 2002). For example, recent work by Keen and Slemrod (2017) develops a general model of optimal tax administration. They show that, at an optimum, the marginal costs and benefits of enforcement should be equated both to each other, and to the marginal costs and benefits of changes in tax rates. Our MVPF approach provides a way of empirically operationalizing this idea by comparing the MVPFs of tax audits to those of modifications to the tax schedule. The MVPF framework generalizes previous work by allowing social preferences to encompass concerns about both vertical and horizontal equity, so that social preferences may differ not only between high and low income taxpayers but also between compliant and noncompliant taxpayers.

The rest of this paper proceeds as follows. Section 2 provides an overview of the audit process, and then describes our data and sample. Section 3 presents the results for the average costs of audits and average revenue raised, reporting their heterogeneity across the income distribution. Section 4 estimates the returns to marginal audits, rather than average audits. Section 5 studies the individual deterrence effects of audits, measuring the impact on future tax revenue. Section 6 analyzes the welfare consequences of audits using the marginal value of public funds (MVPF) framework. Section 7 concludes.

2 Data and Sample

Our analysis leverages unique data from IRS post-filing enforcement divisions. It contains detailed information on the activities performed by IRS enforcement personnel in the US and the institutional costs of those enforcement efforts. We begin with an overview of how audits work. Next, we discuss the data we have from the audit process. Finally, we discuss how we form our primary analysis sample.

2.1 Audit Overview

In this paper, we focus on in-person audits of individuals.[13] (In the Appendix we discuss results for correspondence audits.[14]) Appendix Figure A1 provides a flow chart that maps out the audit process. The audit process begins when a tax return is reviewed, selected for audit, and designated as a field audit or an office audit. This determination is made based on the expected complexity of the audit. In general, office audits are conducted by Tax Compliance Officers and tend to involve a lower-complexity mix of issues. These audits often involve an interview at an IRS office. Field audits are conducted by Revenue Agents and involve more complex cases. These audits often involve an interview at the taxpayer's home or place of business.[15] In either case, the exam stage begins when an IRS examiner reviews the taxpayer's relevant documents and meets face-to-face with the taxpayer to assess whether they have additional tax liability. This process may involve reviewing documents beyond what was included in the tax return, such as receipts that verify the validity of deductions or bank account records that validate all reported income. In some cases, an examination will expand to include returns filed in additional tax years. For the purposes of our analysis, we define an "audit" to include both the evaluation of an initial tax return and the evaluation of additional returns from other tax years that are triggered by the initial examination.[16]

[13]Formally, this means we focus on audits conducted by the IRS's Small Business/Self-Employment (SB/SE) Division, rather than audits conducted by other divisions such as Large Business and International (LB&I) or Wage and Investment (W&I). Within SB/SE, we focus on audits of individuals rather than small businesses. While business tax returns may be assessed in the process of conducting an individual audit, individual audits begin as a review of taxpayers' individual tax returns. A more detailed discussion of the structure of the IRS can be found in Data Appendix A.2.

[14]We show in Appendix B that the average returns are quantiatively similar for correspondence audits and in-person audits, both on average and across the income distribution. In our primary analysis we focus on in-person audits rather than correspondence audits because our data allows for a fuller evaluation of the returns to in-person audits. In particular, we can measure both the marginal returns and the individual deterrence effects of in-person audits. The marginal costs of correspondence audits are harder to evaluate because these audits rely heavily on algorithms to identify noncompliance. This means that a primary component of the marginal cost of expanding correspondence audits may be the cost of infrastructure development and deployment, which is beyond the scope of our analysis. By contrast, the marginal cost of expanding in-person audits depends primarily on the cost of hiring additional auditors. In addition, we note that there are no random audits conducted solely via correspondence, making estimation of deterrence effects more difficult than for in-person audits. Future work could build on the analyses of Hodge et al. (2015) to measure the returns to correspondence audits including deterrence and overhead costs, as we do here for in-person audits. With those caveats in mind, Appendix Figures A2 and A3 show our results for the average returns to correspondence audits across the income distribution. We find qualitatively similar results to our analysis of in-person audits.

[15]Our analysis groups field and office audits as "in-person" audits. This is because they are jointly administered and there is often no bright line between them.

[16]Here we depart from prior research and IRS statistics that consider each tax year as a separate audit. We do this because it is conceptually useful to consider audits of subsequent tax years conducted at the same time as the initial tax year's audit as part of a single audit process. Our discussion of marginal audits examines the returns to initiating a new audit, and such new audits would likely begin with the evaluation of an initial return. We expect that the costs and revenue raised from examining secondary returns may be systematically different from the cost of examining initial returns.

If the examiner determines that the taxpayer has additional tax liability, the next phase of the audit process depends on whether the taxpayer agrees with the determination. If the taxpayer agrees, then the outstanding amount becomes due and the exam ends. If they disagree, the unagreed upon amount is referred to the IRS's Independent Office of Appeals, where an appeals officer will make a determination. If the taxpayer subsequently disputes the appeals officer's determination, the case then moves to tax court where a final determination about tax liability is made.

Once a taxpayer pays their full tax liability, the audit process is completed. If the taxpayer does not pay the full amount, the case is sent to collections. The collections process starts with the IRS sending notification letters to the taxpayer indicating that they have an unpaid balance. If the taxpayer does not respond to the notifications, the case is handled by the Automated Collection System (ACS) or a local field office. If the case is sent to ACS, ACS personnel will try to contact the taxpayer and work with the taxpayer to find a payment solution.[17] If the case is sent to a field office, a Revenue Officer will work directly with the taxpayer to attempt to resolve the unpaid tax liability.[18]

Most of the analysis in this paper focuses on in-person audits, which are selected for review on the basis of suspected noncompliance. In our analysis of deterrence effects, we also make use of audits conducted as part of the IRS's National Research Program (NRP). The NRP is designed to provide critical information on tax compliance and aid in the IRS's estimate of the tax gap.[19] In order to achieve those goals, the IRS assigns Revenue Agents to examine all areas of a randomly-selected sample of returns. These examinations are more intensive than a typical in-person audit but follow the same basic steps outlined in Appendix Figure A1.

2.2 Data

Our data is primarily drawn from three internal IRS sources. First, we use audit-level data that contains detailed information on IRS enforcement activities. The data includes the time spent and activities performed by IRS enforcement personnel in the US. It also includes information on the

[17] The IRS will not call taxpayers without first attempting to contact them via mail. See `https://www.irs.gov/newsroom/phony-irs-calls-increase-during-filing-season`.

[18] If a case is sent to ACS and ACS is unsuccessful at resolving the unpaid balance, the case may then be sent to a local IRS field office. Most collections cases end in one of four ways: (i) fully paid, (ii) with an installment agreement, (iii) with an offer in compromise, or (iv) deemed "currently not collectible." In the fourth case, the IRS temporarily pauses collection efforts. That may happen, for example, if the IRS has deemed that pursuing collection at that time would cause the taxpayer undue hardship.

[19] Returns are selected from a stratified random sample which over-samples certain populations of particular research interest (e.g., high-income taxpayers or EITC claimants). Information on tax gap estimates can be found at: `https://www.irs.gov/pub/irs-pdf/p1415.pdf`.

revenues raised from each stage of each audit. Second, we use internal accounting data for the audit divisions of the IRS. This provides information on costs other than the direct hourly wage costs for the auditors (e.g., employee benefits, management, rent, IT, etc.). Finally, Section 6 presents results from a random survey of audited taxpayers asking about the time they spent and expenditures they made during the audit process. We defer a detailed discussion of this survey data to the welfare analysis in Section 6.

We begin with a description of the audit-level enforcement data and then discuss the internal IRS accounting data. A more detailed discussion of the data can be found in Appendix A.

Audit-Level Enforcement Data. We observe data on IRS enforcement activities from fiscal years 2003–2021 (which ran from October 1, 2002 to September 30, 2021). The enforcement database includes comprehensive information on the revenue collected via the audit process. For each audit conducted, it reports the tax liability, interest, and penalties assessed at each stage of the audit process, as well as the amounts collected. These data allow us to separately estimate the revenue raised at the exam, appeals, and collection stages. On the cost side, the database includes a detailed log of the hours spent on direct enforcement activity by IRS employees on each audit. It also includes the government pay grade (GS grade) of the associated IRS personnel. We translate this into wage costs using a location-specific GS hourly pay-scale and the zipcode of the taxpayer under audit. Multiplying personnel hours by the hourly wage yields the direct labor costs accrued at each stage of the audit process.[20]

Internal IRS Accounting Data. The cost of conducting audits goes beyond the time cost of direct enforcement activity. In order to identify and incorporate these costs, we utilize detailed internal IRS business unit accounting information from fiscal years 2011–2020.[21] These data have not

[20] We are able to use the administrative enforcement data to directly measure the direct labor costs for both the exam and the appeals stage. The enforcement data also contains hours and GS grade for cases sent directly to collections, but it lacks the same information on the collections process for cases that originate in the exam stage before being sent to collections. In order to estimate the direct labor costs associated with collections for cases that start in the exam stage, we therefore, extrapolate based on the cases that started in collections. We start with the universe of cases that are sent directly to collections. We then identify which of these cases enter field collections, which is the last step of the collections process.

We estimate direct labor costs for the subset of collections cases that entered field collections in the same way as exam and appeals cases. We then create a 10-by-10 index, where one axis plots deciles of total positive income (TPI; the lowest decile restricted to zero TPI) and the other axis plots deciles of the amount assessed. Each cell contains the average cost estimate associated with that combination of TPI-decile and amount assessed-decile. For each audit in our primary dataset that ultimately entered field collections, we determine the relevant decile along both dimensions and apply the corresponding direct labor cost estimate.

[21] We are grateful to the IRS for sharing this internal data for the first time for research purposes to enable a comprehensive analysis of the costs of audits.

previously been used to study the returns to tax audits, but are essential to measure the full costs associated with an audit. These data provide line-item costs for major expenditure types. We combine these line items into three categories: (1) non-direct labor-related costs, (2) organization-wide costs, and (3) central IRS management overhead. Non-direct labor-related costs include time spent by auditors on non-auditing activities (such as training), wage costs for management and support staff, and non-wage costs for employees such as fringe benefits and workers' comp. Organization-wide costs include items such as building costs and information technology costs. Finally, central IRS management overhead includes costs incurred inside the IRS but outside the primary business unit conducting individual audits. It also includes costs for services performed by other government agencies. A detailed accounting of these sub-components can be found in Appendix A.2.

We calculate direct wage costs separately for the exam, appeals, and collection stages of the audit. For the exam stage, we are able to estimate cost multipliers for each major component of overhead: non-direct labor-related costs, organization-wide costs, and general overhead costs. We allocate these costs to each audit in proportion to the audit's direct wage costs, motivated by the idea that non-audit labor costs and fringe benefits should rise with the amount of expenditure on a given audit. Appendix Figure A4 shows the robustness of our results to this alternate methods of cost allocation. When costs are allocated in proportion to labor hours, the rate of return across the income distribution is essentially unchanged. When these costs are allocated on a per-audit basis, fewer costs are assigned to high-income audits (which are more costly on a per-audit basis) and so this flattens the profile of returns across the income distribution. Our baseline approach is intended to be relatively conservative given that our primary results to follow show that the return to audits increases with taxpayer income.

As noted below, these exam costs make up 93% of total audit costs. As for the other audit costs, we apply a per-notice cost for cases that went to collections, a "cost-per-dollar-raised" multiplier for cases spending time in ACS, and an overhead cost per dollar of direct labor cost for audits in field collections. Appendix A provides further details on these costs.

2.3 Sample

We focus our analysis on audits of individual tax returns filed in tax years for which (a) we observe comprehensive measures of costs and (b) sufficient time has passed that nearly all audits have been completed. This leads us to study the universe of audits initiated by reviews of individual tax returns from 2010–2014.[22] We begin with 2010 because tax year 2010 audits take place beginning in 2011, the first year for which we have internal IRS accounting data. We end our sample with the 2014 tax year because this provides us at least 7 years of follow-up to observe the costs and revenues associated with an audit. With these definition, our baseline sample includes approximately 710,000 in-person audits.[23]

Appendix Figure A5 uses data from tax year 2003 to report the revenues that are obtained in each year after the audit. The figure illustrates that almost all revenue from audits is collected within 15 years. Of this, 86% of revenue is obtained within 7 years and 96% of revenue is obtained within 11 years.[24] We use these patterns in our results to impute any remaining revenue we expect to be collected from audits of 2010–2014 tax years.

For each audit, we link the taxpayer to their Form 1040 in the tax year that initiated the audit. We use this information to study the heterogeneity of revenue and costs as a function of the individual's income. Our income measure, total positive income (TPI), is the sum of the various positive income items reported on the return. This is the primary measure used by the IRS to categorize returns by income for audit selection.[25] For all income values, we deflate to constant 2016 dollars using the CPI-U-RS.

3 Average Revenue and Costs per Audit

We begin by estimating the average costs and revenues associated with in-person audits. We report the costs and revenues separately for each stage of the audit process.

Average total costs. Starting with the exam stage, the average in-person audit takes roughly 28.7 hours and is conducted by auditors earning about $38 per hour. This yields $1,097 in direct labor costs for the time auditors spend examining returns.

[22] As noted above, we define an "audit" to include all concurrently audited tax years, even if the review of a 2010–2014 return leads the auditor to examine returns for tax years prior to 2010.

[23] We also study the universe of 4.2 million correspondence audits for tax years 2010 to 2014 and 126,000 NRP audits for tax years 2006 to 2014. As discussed in Section 5, we expand to tax years 2006–2014 when studying the NRP audits to improve precision and increase the length of the post audit period for estimating deterrence effects on future tax revenue.

[24] This repayment pattern is fairly stable across initial tax years.

[25] This measure excludes losses, which avoids concerns about the accuracy of reported losses. Consequently, this approach also treats an individual as high-income if they have both high levels of positive income and large losses. Such an individual is likely to be high-income over the long-run, even if they have a low adjusted gross income (AGI) in a given year. It is important to note that this measure of TPI is determined pre-audit. Below, we discuss how our results could potentially differ if one were to use post-audit income (and we conclude the general patterns of our findings are largely unchanged.)

In order to measure the total cost of these hours spent on the audit, we then need to allocate the overhead cost associated with that auditing work. The first component of overhead is additional labor costs not directly allocated toward an audit. This includes wage costs for non-auditing hours, fringe benefits, training costs, and manager labor costs. We allocate these costs in proportion to wage costs on each audit and find they are approximately $2,115 per audit, roughly double the cost of direct labor hours. Next, we incorporate overhead associated with organization-wide costs. This includes costs like building rent and information technology. These costs add $1,103 per audit in the exam stage. Finally, we incorporate the overhead costs associated with central IRS management. Those result in an additional $1,593 per audit. The $1,097 spent on audit hours comes with an additional $4,811 in overhead costs, for a total of $5,907.

This 4.39:1 overhead ratio is meaningfully larger than the ratio used in recent literature. This divergence occurs because previous work, such as Holtzblatt and McGuire (2020), only incorporates auditor labor costs and associated fringe benefits and thus omits all non-labor overhead costs as well as labor costs for support staff and management. This overhead ratio compares average direct labor costs to the average of all other costs. It is not necessarily a measure of the overhead costs associated with marginal expenditures on audits. In Section 4, we discuss how economies of scale may cause the overhead costs associated with marginal expenditures to fall below this 4.39:1 figure.

After the exam stage, a small subset of taxpayers files appeals or contest their cases in tax court. We estimate that the appeals and tax court stage increases the average cost of an audit by $170. This combines $108 in direct labor hours costs and $62 in additional costs.[26] Additionally, many cases end up in collections, which we estimate leads to an additional $191 in costs.[27] Finally, as noted above, we only observe costs that are accrued in a 7–11 year window after the year a tax return is filed. While nearly all costs accrue within this window, there may be some that are incurred afterward.[28] To account for those potential costs, we plot the trajectory of costs accrued for returns filed in the 2003 tax year. Tax year 2003 lies before our primary sample frame, but provides us 18 years of follow up data.[29] We estimate that in our 7–11 year sample frame we capture 96–99% of

[26]This additional cost figure relies on internal IRS accounting information that explicitly allocates appeals costs from the appeals business unit to the SBSE business unit. We cannot rule out the possibility that some additional overhead costs associated with appeals or tax court stages are allocated to other overhead line items, such as the costs of shared office space. Accounting for these other costs would slightly increase the non-labor costs of appeals, but would not bias our estimate of the overall costs of the audit.

[27]These $191 in costs are due to $22 in direct labor hours costs, $96 in associated overhead costs, $9 in the cost of notices, and $63 in costs associated with the Automated Collection System (ACS).

[28]The IRS typically has 3–6 years after a return is filed to conduct an audit and make an assessment of tax liability. The specific statute of limitation is determined by the type of noncompliance found on the tax return. Costs that accrue after that 6 year mark are typically part of the collections or appeals process.

[29]While we use estimates from 2003 for our baseline results, we find similar patterns using 2004–2008 returns.

total labor hours spent on audits.[30] The results of this projection exercise can be seen in Appendix Figure A6 panel B. Our forecast suggests that we do not observe $150 in additional future spending on each audit. When we incorporate this estimate, we get a total average cost of $6,418 for each in-person audit.

Average total revenue. How much revenue do audits generate? Figure 1 shows that in-person audits collect an average of $6,194 during the exam stage, $617 during the appeals stage, and an additional $6,259 through collections. These estimates incorporate all revenue collected through 2021. As is the case with our cost estimates, we use a projection exercise to estimate the magnitude of revenue collected outside our observed 7–11 year window. We estimate that 86–96% of all revenue is collected within 7–11 years after filing. This corresponds to additional revenue of $1,212 per audit. The results of the projection can be seen in Appendix Figure A6 panel A.[31]

Putting together results from each stage of the audit, we estimate average total revenue is $14,283 per audit and average costs are $6,418. However, the costs of the audit are, on average, incurred before the revenue is obtained. Appendix Figure A5 demonstrates that average revenues lag average costs by approximately 1 year. Applying a 3% discount rate to revenue collected to align the timing of costs and revenues, the return to an average audit is reduced by $353.[32] This implies total revenue of $13,930, which is 2.17 times higher than the total average cost.

Heterogeneity by Income Next, we analyze how revenues and costs vary by taxpayer income. Figure 2 Panel A presents the average revenues and costs for in-person audits separately by percentiles of taxpayer total positive income (TPI).[33] We split income on the horizontal axis into bins of 5 percentiles. We break out the top 5% into the 95–99th percentiles, 99–99.9th percentiles, and the top 0.1%. The red triangles plot the average costs and the blue dots plot average revenues of audits by income bin.

[30] In particular, we estimate the trajectory of costs accrued in each post-filing year for each decile of TPI. We then aggregate across all years and all observed income bins to produce this average.

[31] Appendix Figure A5 panels A and B presents results separately by income (TPI) bin. This allows for heterogeneity in the repayment patterns by income bin.

[32] Specifically, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued in each year after the audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

[33] Figure 2 presents the heterogeneity in returns to audits as a function of the observed TPI on the originally filed tax return. This is the relevant definition of income from the perspective of the IRS deciding whether to conduct an audit. It is the measure available at the time of the audit. It is also useful, however, to understand the returns to audits as a function of the true income of the taxpayer. Appendix Figure A7 shows that we obtain qualitatively similar, but slightly accentuated, patterns when using imputed measures of post-audit TPI along the x-axis. The returns to audits at the bottom of the income distribution fall slightly while returns at the top of the income distribution increase slightly to more than 8:1.

On average, it costs just over \$5,000 to audit a taxpayer in the bottom half of the income distribution. It costs \$5,221 to audit an individual in the 70–80th percentiles, \$6,863 in the 90–99th percentiles and \$15,170 in the top 0.1%. Appendix Figure A8 shows that this is primarily because audits of higher income taxpayers take auditors longer than audits of lower income taxpayers. For example, auditing taxpayers in the 70–80th percentiles requires an average of 27.8 hours of auditor time while auditing taxpayers in the top 0.1 percent requires 64.6 hours of auditor time.[34]

While audit costs rise with taxpayer income, audit revenues rise even faster. On average, auditing a taxpayer in the bottom half of the income distribution yields \$4,984. Auditing a taxpayer in the 70–80th percentiles yields \$8,270, while auditing a taxpayer in the 90–99th percentiles yields \$14,973 and auditing a taxpayer in the top 0.1% yields \$95,491.

Figure 2 Panel B divides the average revenue by average cost to show the average returns to audits across the income distribution. The 2:1 average return across the full population varies considerably with income. [35] On average, each dollar spent to audit a taxpayer in the bottom half of the income distribution returns \$0.96 in revenue. Each dollar spent to audit a taxpayer in the 70–80th income percentiles returns \$1.58. Auditing a taxpayer in the the 90–99th income percentiles returns \$2.18 for each dollar spent, and auditing a taxpayer in the top 0.1% by income returns \$6.29 for each dollar spent.[36]

4 Marginal Audits

What are the returns to expanding (or contracting) audits? In the previous section we calculated the average cost and average revenue associated with in-person audits. In this section we explore the returns to marginal audits.

The revenue and costs of marginal audits may differ from the average revenues and costs of audits for two key reasons. First, if the audit selection process seeks to maximize revenue per dollar of audit cost, then there may be diminishing marginal revenues associated with additional audits.

[34] Audits of high-income taxpayers are also conducted by more experienced auditors. Those auditors receive higher pay, and so there is some variation in the average hourly costs of audits across the income distribution. For example, as shown in Appendix Figure A8, audits of taxpayers at the 70–80th income percentiles are conducted by auditors earning an average of \$36.86 per hour. Audits of taxpayers at the top 0.1 percentile are conducted by auditors earning an average of \$45.66 per hour.

[35] While our analysis focuses on in-person audits, Appendix Figure A3 shows a similar qualitative pattern for correspondence audits. Average returns per dollar spent on correspondence audits range from below 1 at the bottom of the income distribution to 11.7 in the top 0.1%.

[36] These calculations are produced using expenditure-weighted averages. This is why the average return across the whole population is relatively similar to the return in the 90–99th percentiles. High-income taxpayers are more likely to be audited and audits of high-income taxpayers are more intensive than audits of low-income taxpayers.

Second, economies of scale mean that some costs — particularly costs other than direct labor costs — may not increase linearly with audit hours. This section explores the returns to marginal audit expansions.

(Lack of) Diminishing Marginal Revenue We study the return to marginal audits by exploiting the steep audit rate decline for returns filed between the 2010 and 2014 tax years.[37] The IRS audited 0.92% of all returns from the 2010 tax year, but that rate fell to 0.56% for 2014 tax returns. As shown in Figure 3 Panel A, this represents a 40% decline in the overall audit rate.[38] We use this reduction in audit rates to understand the revenue associated with marginal audits. We conceptualize marginal audits as those that would have been conducted at 2010 levels of audit intensity but not conducted at 2014 levels. This exercise helps us shed light on the potential returns to expanding audit expenditures back to their 2010 levels.[39]

The return to reversing this decline in audit rates depends, in part, on the audit selection process used by the IRS. If audit reductions between 2010 and 2014 prioritized cutting audits with low revenue per unit cost, then audit expansions would have lower returns than average audits. In contrast, if the audit selection process prioritized other criteria — fulfilling statutory requirements, minimizing the fraction of audits that result in no change to tax liability, etc. — then the returns to marginal audits may not differ from the returns to average audits.

Figure 3 Panel A displays the average revenue per audit and average direct wage cost of each audit between 2010 and 2014. Despite the sharp decline in audits in this period, the revenue and cost figures remain essentially unchanged. In 2010, revenues were 11.0 times direct wage costs. In 2014, revenues were 11.2 times direct wage costs.[40] This suggests that the types of audits cut between 2010 and 2014 had similar returns to the audits that were still conducted in 2014.[41] In

[37]We end our analysis with 2014 tax returns because, as discussed in Section 2, audits from subsequent tax years may be ongoing. This means we have a less complete picture of the returns filed in subsequent years. That said, Appendix Figure A6 shows that our conclusions remain similar when studying the continued decline in audit rates for returns filed in tax years 2015 and 2016 and adjusting for expected future revenue using the methods discussed in Section 3.

[38]The audit decline we exploit in our analysis is actually slightly larger than this 40% figure. Our primary analysis focuses on in-person audits, where audit rates declined 47% between 2010 and 2014. In 2010, 0.21% of all returns were selected for in-person audit compared to 0.11% in 2014.

[39]It is important to note that we are only exploring the returns to expanding audits on one specific margin. Other policy changes, such as changes in the audit selection process, could result in different marginal audits with different returns.

[40]We use wage costs here to set aside any discussion about changes to overhead costs over time. We discuss marginal overhead costs in the next subsection.

[41]We note that we find a similar pattern for the revenues per audit for correspondence audits: there was a 38% decline in the audit rate for correspondence audits from 2010 to 2014. The average return to a correspondence audit was $1,194 in 2010 and $1,294 in 2014.

other words, the return to marginal audits was the same as the return to average audits.[42]

That basic pattern is further validated when examining changes in audit patterns across the income distribution. In particular, we find both that the reduction in audit expenditures was relatively consistent across the income distribution and that the returns to audits did not meaningfully change at any income level. Figure 4 Panel A reports the change in the total direct wage costs of audits at each 5-percent TPI bin (and several high-income TPI bins) between 2010 and 2014. Audit expenditures fell between 31% and 42% for the bottom half of the income distribution and fell slightly more in the top decile, with reductions between 46% and 61%.[43] This consistent decline in audits across the income distribution helps explain why the average return to audits did not rise between 2010 and 2014, even though the average return to audits varies across the income distribution. Audit reductions were not concentrated among income groups with low average audit returns. Figure 4 Panel B reports the return to audits in each TPI bin (measured as the revenue per dollar of direct wage costs). It plots the returns in 2014 on the vertical axis against the returns in 2010 on the horizontal axis.[44] We find a slope close to 1 (1.05 with a standard error of 0.03) and intercept close to 0 (0.21 with a standard error of 0.34), showing that, across TPI bins, there is no systematic change in the returns to audits despite significant declines in audit rates.

One potential concern with this conclusion is that there may have been an underlying time trend in the returns to a typical audit between 2010 and 2014. For example, it could be the case that tax evasion fell substantially between 2010 and 2014. That pattern alone would produce a reduction in the returns to audits. This could mean that changes in the audit selection process increased the average return to audits between 2010–2014, but that this increase was masked by a compositional shift in the amount of tax evasion in the first place. We mitigate this concern by drawing upon audit data from the National Research Program (NRP). Each year, the NRP randomly selects a series of returns for review. As noted previously, NRP audits are more intensive than operational audits because they are designed to measure tax compliance. If tax evasion declined over time, then NRP

[42] This is consistent with the findings of Sarin and Summers (2020) who plot aggregate audit rates and revenue collected between fiscal years 2011 and 2018 and show they follow very similar downward trajectories, suggesting stable revenues per audit. It is also broadly consistent with Holtzblatt and McGuire (2020) who argue that the returns to office audits remained flat between 2010 and 2017 while the return to field audits fell slightly. Their approach and sample construction differ from ours in a number of ways. For example, they examine returns filed through 2017 but restrict to audits that are completed within just two years. In particular, for their primary analysis they restrict to 2010 return audits completed before March 31, 2012 and 2017 return audits completed before March 31, 2019.

[43] Here, we report declines in audit expenditures rather than declines in audit rates. We do this to capture any change in audit intensity that could have occurred between 2010 and 2014. In practice, direct labor costs per audit did not meaningfully change across income bins so the decline in audit rates closely resembles the decline in audit expenditures.

[44] Appendix Figure A9 reports the analogous figure for 2010 and 2014 revenue per audit by TPI bin.

audits would detect less evasion in later years. Appendix Figure A10 panel A plots the cost of and revenue from NRP audits over time. There is no systematic decline in the return to NRP audits. Instead, the return to NRP audits rises slightly. Costs remain consistent over time and the revenue raised rises slightly between 2010 and 2014. Appendix Figure A11 repeats this analysis separately by income bin. It shows there is no systematic decline in NRP returns within any income group.

Taken together, this evidence suggests that the marginal audits cut between 2010 and 2014 produced similar revenue (per dollar of auditor wages) to the audits that remained. While estimating the IRS's objective function for audit selection is beyond the scope of this paper, the evidence suggests that audit cuts between 2010 and 2014 were driven by considerations other than the maximization of revenue per dollar of audit cost. This conclusion is also consistent with the message communicated to our research team in internal discussions with IRS officials, who highlighted that the IRS does not have a mandate to solely maximize revenue, but rather is focused on ensuring broad-based compliance with the tax code.

These results suggest that expanding audit rates back to 2010 levels would yield marginal revenue close to average revenue. It is important to note here, however, that these results apply locally to the large (~40%) decline in audits that occured in this period (or the 66% expansion necessary to return to previous audit levels). They do not apply globally. If the IRS were to continually expand audits, it would eventually reach a point of diminishing marginal returns where the marginal revenue collected from new audits diverged from the average revenue collected by previous audits. [45]

Marginal Costs When calculating the total cost of marginal audits, we need to consider the role of non-wage costs and to account for economies of scale. Some costs may not scale proportionately with the number of audits conducted.

For our baseline estimates, we assume that costs are 27% fixed and 73% variable. We arrive at this figure using a combination of business unit accounting information and information from existing IRS budget requests.

As noted previously, the costs of an audit beyond the direct wage costs fall into three broad categories: labor/fringe, organization-wide costs, and central overhead/headquarters costs. La-

[45] We can see this most simply by examining the revenue collected from randomly selected NRP audits. Despite being more intensive than the average audit, NRP audits collect less revenue per audit than operational audits. Appendix Figure A11 Panel A shows that the average revenue per NRP audit for the 99–99.9th percentile is $9,432. This is far below the average of $28,451 collected from in-person operational audits in the same income percentile. This suggests the returns to audits would be much lower if returns were selected for audit at random instead of using the current selection procedures. We also note that the individual deterrence effects we estimate in the next section are another potential source of diminishing returns: if more audits today increase future compliance, they lower the future returns to re-auditing those taxpayers.

bor/Fringe is mostly wages for hours not spent on audits and fringe benefits. Organization-wide costs are primarily the cost of renting and maintaining office space and information technology. Central overhead contains costs shared across IRS business units or incurred by government agencies outside the IRS. For our baseline analysis, we assume that both labor/fringe and organization-wide costs are variable while central overhead costs are fixed. This produces a ratio of total marginal costs to direct labor costs of 3.93:1.[46]

Those results are broadly consistent with, but slightly more conservative than, estimates from existing IRS budget requests. For example, the IRS has published budget estimates from audit expansions as part of the Program Integrity Allocation Adjustment. Total marginal non-labor costs in those requests are approximately 23.6%[47] as large as total labor and fringe costs.[48] By comparison, our baseline estimates assume that total marginal non-labor costs are 25.6% as large as total labor and fringe costs.

While these comparisons help to validate our baseline measures of marginal costs, we also present our results under alternate cost assumptions. We form an upper bound on costs by assuming that marginal costs are equal to current average costs (i.e., all costs are variable and scale proportionally.) We also form a lower bound on costs by assuming that only labor and fringe benefit costs scale proportionally with audit wage costs.

Putting together these results on marginal revenues and marginal costs, Figure 5 reports our estimates for the marginal returns to expanding audits from their 2014 to 2010 levels. We start by combining the average costs and average revenue of audits across the income distribution from Figure 2. Figure 5 then adjusts the average costs downward because the marginal costs fall below the average costs. There is no adjustment made to marginal revenue because we find that restoring audits to 2010 levels would not diminish the marginal returns. The figure reports the returns to marginal audits in each income group. The return to a dollar of spending on marginal audits in the bottom half of the income distribution remains close to $1. The return rises rapidly with income: the return is $2.99 in the 90–99th percentiles, $4.35 in the 99–99.9th percentiles and $8.63 in the

[46] These estimates are a "steady state" measure of the marginal returns to additional audits. In other words, while estimates of overhead include expenditures such as training, they are based on average training costs over time. If a substantial number of revenue agents are hired in a short period of time, training costs may be higher in the short run. This type of adjustment is made explicitly in estimates produced by the CBO. For example, their analysis of recent IRS budget expansion proposals explicitly incorporates rising auditor productivity as new hires are trained in their first three years (CBO, 2018).

[47] See pages 127–129 of https://home.treasury.gov/system/files/266/02.-IRS-FY-2022-CJ.pdf. Total labor costs are $85,074 and total "other direct costs" are $28,452, which consists not only of enforcement costs but also correspondence and document matching. Allocating those costs proportionally by labor costs yields $20,111 of enforcement other direct costs. Taking the ratio of 20,111/85,074 yields 23.6%.

[48] This figure is also broadly consistent with internal IRS estimates of large scale audit expansions.

top 0.1%.

5 Individual Deterrence Effects of Audits

Thus far, our analysis focuses on the direct revenue obtained from assessments made during an audit. Audits may also raise revenue in an indirect manner by deterring future noncompliance. In this section, we examine one particular form of deterrence known as individual or specific deterrence. This form of deterrence, distinct from the concept of general deterrence, refers to a situation where auditing an individual in one year encourages greater tax compliance in future years. This can be an important mechanism through which audits affect revenue.[49]

In order to estimate individual deterrence effects, we draw upon random audits conducted by the National Research Program (NRP) and build on the approach developed by DeBacker et al. (2018) (DHTY). As in DHTY, we construct a treatment group of individuals audited by the NRP and compare their taxes paid to a group of control individuals not selected for random audit. In our analysis, we construct our control group using stratified matching. The NRP divides taxpayers into strata on the basis of return characteristics such as total positive income (TPI), EITC receipt, and the presence of self-employment income. For each audited observation, we select a matched control observation from the same stratum and same tax year using a nearest-neighbor approach. In particular, we match each selected return to the non-selected return within the same NRP stratum that has the closest level of TPI. Stratified matching produces a control group with similar pre-treatment income to the treatment group, especially for the highest-income taxpayers. We then track income reported and taxes paid in each year after the initial audit.

Our approach builds on DHTY in two additional key ways. First, we use additional data to substantially increase the duration of the observed post-audit period. We use random audits from 2006–2014 and estimate the impact on behavior up to 14 years post-audit. In comparison, DHTY draw upon random audits from 2006–2009 and estimate the impact on behavior up to 6 years post-audit. Second, our event study framework incorporates year-since-audit fixed effects. As we detail below, the omission of these controls in DHTY, in combination with the use of a top income quintile threshold, results in a mismatched comparison between treated and control taxpayers. That

[49] In general, the literature on tax audits distinguishes between two potential sources of deterrence – individual (or specific) deterrence and general deterrence. General deterrence refers to a situation whereby people increase compliance in response to an increase in the likelihood of being audited (e.g. as in the classic model of Allingham et al. (1972)). This means that auditing one individual encourages other individuals to change their behavior. We do not quantify the role of general deterrence in this paper, but return to a discussion of general versus individual deterrence in our discussion of the welfare impacts of audits.

comparison produces a spurious negative deterrence effect at the top of the income distribution.[50]

Figure 6 Panel A plots results from our primary event study for the full population of audited individuals. We find clear and persistent deterrence effects. The figure displays the difference in mean taxes paid between taxpayers selected for an NRP audit and those in the matched control group. We weight each observation by the inverse of the NRP sampling probability to ensure that our results measure the average deterrence effect across the full population of taxpayers (as opposed to the distribution of audited taxpayers). As one would expect, the difference in taxes paid between treatment and control is statistically indistinguishable in the years prior to the NRP audit. In the years following the audit, a clear gap emerges. Starting in Year 2, we find a statistically significant deterrence effect.[51] The yearly impact on taxes paid is around $300 per audited taxpayer, which is approximately 30% of the $1,026 in revenue collected via the initial audit.[52] The results also demonstrate that the individual deterrence effect is highly persistent. Taxes paid remain elevated up to 14 years after the initially audited return, although estimates become less precise in the final years of the event-study window.[53]There is no clear change in the magnitude of the effect over the full 14-year window.[54] When we sum the deterrence effect over 14 years and apply a 3% discount rate, we estimate that an average NRP audit produces $3,258 in additional taxes paid via the individual deterrence channel.[55] This effect is roughly 3.2 times the revenue collected from the initial NRP audit.[56]

[50]We again thank Alex Yuskavage and the whole DHTY team for their openness, sharing their code, and providing assistance in conducting these comparisons. We note that including the year-since-audit fixed effects does not substantially alter their conclusions when examining average deterrence effects across the full income distribution.

[51]The presence of a smaller treatment effect in Year 1 is consistent with the fact that the audit process itself is not always conducted within a year of the initial tax filing. As a result, taxpayers may file their Year 1 taxes before the NRP process is complete.

[52]The magnitude of this effect is broadly consistent with the previous literature, which finds annual deterrence effects ranging from 20% to 35%. We note that Advani et al. (2023) find a slight decay in point estimates and increase in standard errors over time such that 8 years post-audit they are unable to reject a null of no effect. The slight decay over time is consistent with our findings for audits of taxpayers with business income (Appendix Figure 12). In contrast to Advani et al. (2023), however, we continue to find statistically significant effects after 8 years.

[53]The sample size shrinks because the requisite time has not elapsed since later waves of the NRP.

[54]Appendix Figure A13 shows that these results are also consistent across NRP audit waves. Panel A shows the deterrence event study separately for NRP waves selected from tax years 2006–2009 and 2010–2014. Panel B then reports the same results for NRP waves for tax years 2006–07, 2008–09, 2010–11, and 2012–14. The magnitude and trajectory of the treatment effects are similar across waves.

[55]It is worth noting that the sign of this deterrence effect is not obvious ex ante. Slemrod et al. (2001) argue that some taxpayers learn during the audit process that it is optimal to report lower levels of earnings because they may be able to avoid paying more taxes. Moreover, if an individual believes that the probability of audit has risen, they may perceive a higher effective tax rate and respond by working less.

[56]Appendix Figure A12 explores how deterrence effects differ for taxpayers with and without business income (as measured by income on Schedule C, E and F). For taxpayers without business income the future revenue collected is roughly 3.81 [1.30, 6.39] times initial NRP revenue. For taxpayers with business income, the ratio is 2.81 [1.47, 4.15]. The event study graphs show that, consistent with the patterns observed in DHTY, the deterrence effects for business owners are less persistent than the deterrence effects for other types of filers.

Appendix Figure A14 explores deterrence effects among a sample of individuals chosen for typical in-person audits, Next, we explore how these deterrence effects vary across the income distribution. Table 1 reports the results of the treatment-control difference-in-difference specification separately by bins of TPI.[57] We display both raw estimates and, to limit the influence of outliers, estimates where collected tax revenue is winsorized at the 99th percentile of the population distribution.[58] Broadly, we find significant deterrence effects across the income distribution. Figure 6 Panel B summarizes these results by plotting the ratio of the estimated deterrence effects to upfront revenue collected from the NRP audit.[59] The effect is statistically significant in nearly all income bins and the magnitudes are consistent with the overall deterrence ratio of 3.2 across the income distribution.[60] While we lack the power to estimate precise effects in the top 1%, the point estimates remain similar at the 99–99.9th percentiles and in the top 0.1%.[61]

It is worth noting that our finding of positive deterrence at the top of the income distribution lies in contrast with previous estimates. Notably, DTHY find a negative deterrence effect in the top quintile of income. We find that their negative effect is driven by the omission of year-relative-to-audit fixed effects when restricting their sample to audits of high-income returns. The omission of these fixed effects means that treated audits in one year are compared against a set of control audits selected across multiple different years. The concern here is primarily one of mean reversion: filers classified as high-income in 2006 may have regressed farther to the mean during 2010–2014 than filers classified as high-income in 2009. Without including year-since-audit fixed effects, the time path of treatment interacted with year since audit partially captures this mean reversion. This produces a spurious negative effect for filers classified as high-income. The inclusion of years-relative-to-audit fixed effects in their initial specification recovers estimates broadly consistent with the findings in this paper.

Having assessed deterrence effects overall and across the income distribution, Figure 7 multiplies

rather than those chosen for NRP audit. These audits are non-random, but we nonetheless match audited individuals to non-audited individuals on the basis of income and return characteristics. The difference in differences estimates suggest a multiplier of approximately 2.5, consistent with the results in our NRP sample. These alleviates any concerns that the results from the NRP sample do not generalize to typical in-person audits.

[57] Here we adopt a difference-in-difference approach rather than a simple difference to improve precision, but the results remain similar across those two specifications.

[58] Winsorization at the 99th percentile is consistent with the approach taken in DHTY.

[59] We express both of these in present discounted value.

[60] In our baseline specification we winsorize collected tax revenue at the 1st and 99th percentiles of the distribution of taxes collected. Table 1 shows that these estimates remain consistent across the income distribution without winsorization. Appendix Figure A15 repeats the exercise found in Figure 6B to present a side-by-side comparison of the winsorized and unwinsorized deterrence ratios. While the unwinsorized estimates are noisier at the top of the income distribution, the broad trajectory of our results remains the same.

[61] For those in the top 1%, we cannot reject the hypothesis that audits generate 5x additional revenue from deterrence, nor can we reject that they have large negative effects on deterrence.

these ratios of revenue from deterrence to revenue from the initial NRP audit by the revenue raised from operational audits. Panel A takes the baseline R/C figures reported in Figure 5 and applies the overall deterrence multiplier of 3.2.[62] This produces a new estimate of the total revenue raised per marginal dollar spent on audits.[63] As before, the estimated returns to audits increase in income for individuals in the top half of the income distribution. For example, we estimate that each dollar spent on auditing a taxpayer in the 70–80th income percentiles produces a return of $9.06. Each dollar spent auditing a taxpayer in the 90–99th income percentiles produces a return of $12.48. We use these figures below to conduct a welfare analysis of marginal audits.

Panel B shows how these estimates vary when using deterrence multipliers calculated within TPI bins. The confidence intervals increase as the deterrence effects are estimated with more uncertainty, but the trajectory of the point estimates remain the same. For example, when using the TPI-bin individual deterrence multiplier of 6.29x for taxpayers at the 90–99th percentiles, we get a total return of 21.8:1.[64] When following the same approach we get a point estimate of 24.4:1 for taxpayers at the 99–99.9th percentiles and a point estimate of 44.0:1 in the top 0.1%.[65]

6 Welfare Analysis of Marginal Audits

We find that audits, particularly audits of high-income taxpayers, yield revenue that far exceeds their costs. What does that imply about the welfare consequences of tax audits? How should we think about the tradeoff between greater tax enforcement and alternative policies that raise revenue, such as higher tax rates?

[62]We omit the top 1% of taxpayers from this figure since the aggregate confidence interval associated with the overall multiplier of 3.2 does not reflect the statistical imprecision of deterrence effects seen for this group in Figure 6B.

[63]The logic of this approach is that the randomly selected NRP audits can be used to calculate a ratio of future revenue to initial revenue, and that ratio can then be applied to estimate the long-run return to in-person audits. The key assumption is that future behavioral changes across NRP and in-person audits are a constant multiple of the size of the audit adjustment. This is consistent with our results discussed above in Appendix Figure A14 that finds similar deterrence effects for in-person audits using a matching strategy. It is also consistent with a model where taxpayers learn from an audit that certain reporting behaviors are not permitted. Our approach assumes that the extent to which NRP audit adjustments induce learning is the same as the extent to which in-person audit adjustments induce learning. In practice, NRP and in-person audits are similar. They are conducted by the same staff, and while NRP audits are more intensive than traditional in-person audits, that is primarily because they examine more line-items than in-person audits. That said, the use of NRP audits to measure deterrence effects may be a slightly conservative assumption. Individuals subject to NRP audits are told they were randomly selected, while those chosen for traditional in-person audits have to reason to believe that their audit was triggered by content of their previous tax filings. Those subject to in-personal audits may have larger behavioral responses because they infer that similar filing practices will result in additional audits in the future.

[64]NRP audits of individuals in the 90–99th income percentiles produce an average of $1,758 in initial revenue. We estimate they generate $11,055 in additional revenue via deterrence effects in the subsequent 14 years. This figure is 6.29 times the initial revenue.

[65]As noted above, we omit the top 1% from our primary results because we do not have the necessary power to precisely estimate the deterrence multiplier.

In this section, we consider the welfare consequences of tax audits. In particular, we derive and estimate the marginal value of public funds (MVPF) for a change in the audit rate at each point of the income distribution. Along the way, we provide new evidence from a survey of audited taxpayers on the burden of audits across the income distribution. We use our MVPF estimates to examine desirability of expanded audits relative to other potential methods of raising revenue such as changes in the tax rate.

A key advantage of the MVPF framework is that it provides a transparent way to incorporate welfare weights that vary across the population.[66] In our setting, this not only enables different welfare weights on high- versus low-income taxpayers (i.e. vertical equity) but it also allows for different welfare weights on compliant versus noncompliant taxpayers (i.e. horizontal equity). We show that in the case of marginal audit expansions, nearly all the costs are borne by noncompliant taxpayers. So, if a social planner were to place a low welfare weight on costs imposed on noncompliant individuals, increases in audit enforcement would raise revenue at a particularly low social welfare cost.

The MVPF of tax audits captures the welfare cost imposed by an audit per each \$1 government raised. It is defined as the ratio of the willingness to pay to avoid the audit divided by the net revenue to the government that is raised by the audit.

$$MVPF^{audit} = \frac{\text{WTP to Avoid Audit}}{\text{Net Govt Revenue Raised}}$$

For any revenue-raising policy, a lower MVPF means the policy raises funds more efficiently. As a point of reference, a simple non-distortionary tax would have an MVPF of 1. The revenue raised by the government would be exactly equal to the individual beneficiary's willingness to pay to avoid the tax. If the MVPF of a tax audit were 1.1, that would imply that the audit imposes \$1.10 in private welfare costs for each \$1 in government revenue it raises. Assessing the social welfare impact of the expansion then requires applying social welfare weights to judge the value of the \$1.10 in the hands of individuals impacted by a marginal audit. We discuss these considerations in further detail below when comparing audits to other methods of raising revenue.

Appendix D provides a formal model derivation of the MVPF of individual tax audits in a wide class of dynamic models with audits and evasion. We focus the main text on the intuition of the MVPF components that we derive from this model.

Let us first consider the the numerator of the MVPF, the willingness to pay to avoid the audit.

[66] See Hendren and Sprung-Keyser (2020) who extend the analysis of Mayshar (1990) to allow for arbitrary welfare weights on individuals in the economy.

There are two key terms in the numerator. First, audited individuals experience a welfare loss equal to the amount of additional money they are required to pay the tax authority. We let R denote this revenue raised by the audit. In Appendix D, we show that this term includes not only the initial revenue raised by the audit but also additional revenue raised in the future as the result of the individual deterrence effect of the audit.[67,68]

Second, the individual cost of being audit includes not only the revenue paid to the government but also other costs to the taxpayer of complying with the audit. We let B denote the monetized value of this taxpayer burden. The taxpayer burden captures an individual's willingness to pay to avoid the cost of representation by specialists, the time costs of an audit, and other hassles associated with an audit. Combining the revenue paid with the taxpayer burden yields the willingness to pay to avoid an additional audit, $R + B$.

Next, we turn to the denominator of the MVPF, which captures the net cost to the government of an additional audit. This is equal to the marginal revenue raised by the audit, R, minus the marginal cost the government pays to conduct the audit, which we denote by C.

When we combine the willingness to pay, $R + B$, and net cost to the government, $R - C$, this yields the formula for the MVPF of:

$$MVPF^{audit} = \frac{R + B}{R - C}.$$

[67]Individual deterrence revenue enters into the numerator of the MVPF, despite the fact that individuals are changing their earnings behavior in response to an initial audit. In other words, the application of the envelope theorem does not mean individuals are indifferent to these deterrence effects. This is because individual deterrence include a "mechanical" response to the changes in one's constraints that arise after being audited. In particular, the penalties for repeat offenses are quite large and the burden of those changes in penalties are part of the mechanical cost of the initial audit. In Appendix D we show that under quasilinearity utility, the present discounted value of net tax revenue paid, R, perfectly captures the individuals' willingness to pay to avoid the audit. We also discuss how this conclusion differs if the mechanism driving individual deterrence is an increase in future audit probabilities instead of higher future penalties. In particular, if future audit rates do increase, then the MVPF should include both the additional government costs and additional burden of those audits (the willingness to pay component remains the same present discounted value of net revenue from the audit).

[68]While we do not measure general deterrence effects in this paper, we show in Appendix D.4.2 that these effects can enter the denominator of the MVPF but not the numerator. In a benchmark model where all penalties are financial and all taxpayers are risk-neutral, general deterrence effects are zero. This is because taxpayers minimize tax liability, and so the envelope theorem means their behavioral responses do not affect tax revenue. In a more general case where individuals are risk averse or have non-financial costs of being audited, the effect of general deterrence on government revenue is non-negative. In that more general case, the general deterrence effects are omitted from the numerator due to the envelope theorem. They are the result of a behavioral response rather than a mechanical effect of policy. This means that if such general deterrence effects were present, our benchmark estimates of the MVPF would be likely upper bounds relative to the MVPF with general deterrence effect included. One point of caution with this conclusion is that there is some empirical work suggesting general deterrenence can produce opposite-signed revenue effects due to negotiation concerns or higher implicit tax rates. Most notably, (Slemrod et al., 2001) consider the effect of notices to high-income taxpayers warning them of likely scrutiny on their future returns. The authors find this leads to reductions in tax liability reported. We leave a detailed empirical analysis of general deterrence as an important direction for future work.

Dividing the numerator and denominator by the cost of the audit, we obtain:

$$MVPF^{audit} = \frac{\frac{R}{C} + \frac{B}{C}}{\frac{R}{C} - 1}. \tag{1}$$

Here, R/C is the revenue raised per dollar of marginal spending and B/C is the taxpayer burden per dollar of government cost. Intuitively, a higher value of R/C pushes the MVPF of audits closer to 1. If the cost to conduct the audit is small compared to the revenue collected, the policy is closer to a non-distortionary tax. By contrast, if R/C is low, then there are large distortionary costs associated with the audit, and so the MVPF is higher.[69]

In order to estimate the MVPF of tax audits across the income distribution, we use the empirical estimates of revenues and costs constructed in this paper. The key remaining input into our MVPF calculation is the the taxpayer burden of audits, B.

Burden of Audits

We provide new insights into the burden of audits using information from a 2023 survey conducted by the IRS.[70] The survey asked recently audited taxpayers about the hours that they spent responding to the audit and any financial expenses they incurred when complying with the audit (e.g. payments to lawyers or accountants). The IRS then constructed a measure of the total monetized burden of these audits using a taxpayer-specific wage imputation to translate time burden into financial burden.[71]

We obtain the average value of these survey-derived burden estimates separately by bins of taxpayer income (TPI).[72] Within each TPI bin we also obtain separate burden estimates for compliant and noncompliant taxpayers. This latter split is useful in the analysis to follow as it allows us to consider the possibility that welfare weights differ between those two groups.

Panel A of Appendix Figure A16 splits the population of taxpayers by income bin and reports the hours taxpayers spent complying with audits. On average, individuals spend around 30 hours complying with their audits. This varies from around 20 hours at the bottom of the income distribution to slightly below 40 hours in the 90th–99th percentiles of TPI. In the top 1%, this falls to just below 30 hours. Panel B reports the total monetary spending (e.g. on accountants and lawyers) as a result of the audit across the income distribution. This is around \$500 per audit across the

[69] Analogously, the MVPF of the income tax exceeds 1 to the extent to which higher taxes cause reductions in taxable income.

[70] The survey analyzed a random sample of taxpayers whose audits closed in 2022, covering tax returns filed in 2019, 2020, and 2021.

[71] This procedure is roughly equivalent to dividing total year earnings by 2000 (the expected number of hours worked in a year.) This approach relates to work by Guyton and Hodge (2014) who use an earlier version of this survey to measure taxpayer burden.

[72] Due to sample size considerations, we are not able to separate the 99.9+ and 99–99.9 TPI percentile bins.

bottom nine deciles of the income distribution, but rises to around \$1,500–\$2,000 above the 90th income percentile. Panel C combines these two measures to report the total monetized burden of the audit using a taxpayer-specific imputed wage.[73] This yields a total burden of an audit that rises from around \$650 in the bottom half of the distribution to \$30,000 in the top 1% of the income distribution.[74]

Marginal Value of Public Funds

With these inputs, we can calculate the MVPF of audits across the income distribution. Figure 8 Panel A reports the construction of the MVPF of expanding audits among those with TPI in the 20–30th percentiles. Recall, we estimate that at the 20–30th percentiles every \$1 in audit spending generates \$1.14 in upfront revenue and \$3.63 in future revenue from deterrence effects. This generates total revenue of \$4.78 per dollar of marginal spending. The willingness to pay to avoid the audit is the sum of this \$4.78 in revenue and \$0.14 in taxpayer burden, for a total of \$4.92. Each dollar of government expenditure raises \$3.78 in net revenue (\$4.78 in upfront and deterrence revenue minus \$1 of initial spending). We take the ratio of these two terms to get an MVPF of 1.30. Figure 8 Panel B repeats this calculation for expanding audits among those in the 90–99th percentiles of the TPI distribution. Here, we find an MVPF of 1.15. This MVPF is closer to 1 because the policy has a higher return on expenditure, R/C.

Would it be welfare enhancing to audit more high-income individuals and fewer low-income individuals? This depends on the social welfare weights placed on the audited high- and low-income individuals. For any two policy changes, A and B, with MVPFs given by $MVPF_A$ and $MVPF_B$, raising revenue through policy A to finance reduced revenue (or increased spending) on policy B increases social welfare if and only if

$$\eta_A MVPF_A < \eta_B MVPF_B \tag{2}$$

where η_A and η_B are the social marginal utilities of income of the policy A and B beneficiaries.[75] The left-hand side of Equation (2) measures the social welfare cost from raising revenue through

[73] Roughly 40% of in-person audits result in no change in assessed tax liability. Appendix Figure A17 shows how this "no change" rate varies across the income distribution. In particular, the no change rate is around 40% in the bottom half of the income distribution and rises to nearly 60% in the top 1% of the distribution. Below, we consider the possibility that one might place different welfare weights placed on compliant and noncompliant taxpayers. Such an analysis requires burden information separately by compliance status. Appendix Figure A18 shows the estimated burden imposed on individuals with no change in their tax liability. As compared to those with a change in their tax liability, these compliant individuals spend slightly more time on their audits and spend a roughly similar amount of money.

[74] We note these burden numbers are somewhat higher than those constructed in Guyton and Hodge (2014). The average burden of in-person audits in this sample is \$5,710 for audits concluding in calendar year 2022, whereas Guyton and Hodge (2014) find an average burden of \$3,200 for audits concluding in calendar year 2011.

[75] See Saez and Stantcheva (2016) for a detailed discussion of the social marginal utility of income. Formally, let η_i

policy A and the right-hand side measures the social welfare gain from expending resources through policy B. Given our MVPF estimates, this means that if a social planner places equal welfare weights on the willingness to pay of low- and high-income audited individuals, it would be welfare enhancing to increase audits on the 90–99th percentiles rather than increasing audits on the 20–30th percentiles. In fact, this calculation suggests that such a policy is welfare enhancing as long as the average welfare weight on audited high-income individuals is less than 1.23 times the welfare weight placed on audited low-income individuals.

Comparison to tax and transfer policies A key benefit of the MVPF approach is that it can be used to conduct welfare comparisons of policies across broad policy domains. For example, we consider how the MVPF of tax audits differs from the MVPF of other tax and transfer policies. Hendren and Sprung-Keyser (2020) show that for a change in tax rates, the MVPF is given by

$$MVPF^{tax} = \frac{1}{1 - FE}$$

where FE is the effect of the behavioral response to the policy on the government budget. The behavioral response to higher tax rates usually creates a negative fiscal externality, so tax rate increases usually have MVPFs greater than one.[76]

Figure 9 compares the MVPF of tax audits to a wide range of tax and transfer policies within the United States. The horizontal axis reports the quantiles of the income distribution, displaying how these MVPFs vary with income. The black squares report MVPF estimates of tax and transfer policies analyzed in Hendren and Sprung-Keyser (2020). They include policies such as the 1993 EITC expansion and the 1986 top tax rate reduction and draw upon existing causal estimates to compute the MVPF. Next, the gray triangles report estimates from Hendren (2020), which measures the MVPF of a small change in the tax schedule at each point of the income distribution.[77] Finally, the purple circles show the MVPF of expanding tax audits as estimated above.

be the social marginal utility of income of individual i in society, as defined in Saez and Stantcheva (2016): giving \$1 to individual i increases the policymaker's social welfare function by η_i. The social welfare impact of raising \$1 in government revenue from expanded audits is then given by $\eta^{audit} MVPF^{audit}$, where η^{audit} is the incidence weighted average social marginal utility of income of those being audited, $\eta^{audit} = \sum_i \eta_i \left(R_i + B_i\right) / \sum_i \left(R_i + B_i\right)$. Here, R_i and B_i is the financial and burden cost of the audit of individual i. We note this is an average of weights on both compliant ($R_i = 0$) and noncompliant ($R_i > 0$) taxpayers, with the mix determined in part by the efficacy of the targeting of audits. We discuss heterogeneous welfare weights further in Appendix D.6.

[76] Intuitively, the deadweight loss associated with tax audits is driven by the compliance burden and administrative cost of conducting the audits; the deadweight loss from changes to the tax schedule is driven by the impact of the behavioral response to the tax change on tax revenue.

[77] Hendren (2020) arrives at these estimates by combining estimates of labor supply elasticities, tax rates, and tax data on the shape of the income distribution.

Broadly speaking, the MVPFs of tax and transfer policies increase with income. The MVPFs of taxes and transfers targeting low-income individuals are near one, while the MVPFs of taxes and transfers at the top of the income distribution are around 1.5, 2, or higher. This increasing pattern is consistent with the canonical model of optimal taxation with a progressive social planner. If the social planner wishes to redistribute from high-income individuals to low-income individuals, then it should be willing to impose a greater welfare cost on the high-income individuals when raising revenue. At an optimum, the ratio between the MVPFs for tax policies targeting high-income and low-income people should be equal to the ratio of the social marginal utilities of income placed on those two groups. Put another way, if, at an optimum, the MVPF of a high-income tax increase is 2 and the MVPF of a low income tax increase is 1, that implies the social planner values $1 in the hands of a low-income person as much as it values $2 in the hands of a high-income person.

While the MVPFs of tax and transfer policies increase with income, the MVPFs of marginal audit expansions display the opposite pattern. The MVPF of tax audits of high-income individuals is meaningfully lower than the MVPF of audits on lower income individuals.[78] Such a pattern could only constitute an optimum if the social planner placed a lower welfare weight on money in the hands of a low-income audited individual than in the hands of a high-income audited individual. This suggests that current audit rates are inconsistent with canonical social preferences.

We can see this pattern in another way by directly comparing the MVPF of tax audits to the MVPF of tax increases on high-income taxpayers. The MVPF of tax audits falls well below the MVPF of tax increases: audits impose a lower private welfare cost per dollar of government revenue raised. If current policy were at an optimum, this would suggest that the social planner values money in the hands of high-income audited taxpayers more than in the hands of high-income non-audited taxpayers.

Thus far, our discussion has only considered a single average welfare weight placed on audited individuals at a given income level. This welfare weight can be interpreted as an incidence-weighted average of the welfare weights across all those audited individuals. In practice, one might want to

[78] These calculations do not incorporate the general deterrence effects of audits. If audits of taxpayers at all income levels had similar general deterrence effects, the MVPFs of audits would fall across the board. If, by contrast, the threat of audits disproportionately drove low-income taxpayers not to claim the benefits for which they are eligible, that could change the relative ordering of the MVPFs for audits of high versus low-income taxpayers. In order to explore this, Appendix Figure A19 examines audits of individuals who claimed the EITC. It examines the effect of the initial audit on future EITC claiming behavior. It explores the results separately for those individuals whose initial EITC claim was corrected downward (noncompliant) versus those whose initial EITC claim was not corrected downward (compliant). Taxpayers whose initial EITC claim was corrected downward also claim less EITC in future years. This is an individual deterrence effect. By contrast, individuals whose initial EITC claim was not corrected downward do not change their EITC claiming behavior in future years. This suggests that experiencing the audit did not cause a chilling effect on the taxpayers that prevented them from making future claims.

might want differentiate between between particular subsets of audited individuals. Here, we use the MVPF framework to consider the role of differential welfare weights on compliant and noncompliant taxpayers. Appendix Figure 17 shows that roughly 40 percent of audited individuals experience no change in their tax liability. Appendix Figure 18 reports the burden of audits separately among taxpayers with and without additional assessed tax liability. Combining these burden estimates with the estimates of the audit "no change" rate, we can decompose the total MVPF into the incidence on those with and without additional assessed tax liability. Consider, for example, the MVPF of expanding audits for taxpayers in the 90–99th income percentiles. The MVPF of 1.15 indicates that raising one dollar of revenue incurs total private welfare costs of \$1.15. Decomposing the welfare cost, raising \$1 from audits of taxpayers in the 90th–99th income percentiles imposes private welfare costs of \$1.08 on high-income taxpayers who are not compliant and private welfare costs of \$0.07 on high-income taxpayers who are compliant. If the social planner places a low welfare weight on noncompliant taxpayers, this suggests that audits of high-income taxpayers are a particularly efficient means of raising revenue.[79]

This basic conclusion holds even if the IRS survey data has substantially understated the true burden of audits on high-income taxpayers. Suppose, for example, that the social planner puts no weight on the private welfare costs that audits impose on noncompliant taxpayers. How large would the burden of audits on compliant taxpayers need to be in order to overturn the conclusion that audit increases raise revenue at lower welfare cost than tax rate increases on top earners? As we show in Appendix D.6, in order to overturn that conclusion, an audit would need to impose average burden on a compliant top earner around \$300,000. That \$300,000 figure is more than 10 times the audit burden estimate found in the IRS survey data.

7 Conclusion

In this paper, we conduct a detailed analysis of the returns to tax audits across the income distribution. We estimate that the average IRS audit of a tax return filed between 2010 and 2014 produced \$2.17 in revenue for each dollar spent on the audit. We find that audit returns varied substantially across the income distribution. Audits of higher income taxpayers are more time intensive and more

[79]Here, we can again compare increases in the tax rate with changes in audit rates. Supposing that the social planner placed no welfare weight on noncompliant taxpayers, and that the social planner placed the same weight on compliant taxpayers subject to audit and on the average . The planner would prefer raising revenue from individuals in the top 1% of the income distribution by increasing income tax rates for those individuals rather than by increasing audits of those individuals only if the MVPF of those tax rate increases were less than or equal to 0.08. This is far below the MVPF of 1.5 for top rate increases found in the literature.

costly than audits of lower income taxpayers but obtain much more revenue per dollar spent. We find that the average return to IRS audits rises from \$0.96 in the bottom half of the distribution to \$2.18 in the 90–99th percentiles and \$6.29 in the top 0.1%.

Next, we examine the return to marginal audits, exploiting the sharp decline in audit rates for returns filed between 2010 and 2014. We find that revenue generated from marginal audits is indistinguishable from the revenue generated by average audits, but marginal costs fall below average costs due to economies of scale.

We then use random audits conducted by the IRS National Research Program to estimate individual deterrence effects of audits. We find that auditing individuals once results in a persistent increase in tax revenue collected from those individuals. Measured over 14 years, the future revenue collected is 3.2 times the return to the initial audit. We find that these deterrence effects are present across the full income distribution and can be measured with precision for all but the top 1% of taxpayers. Combining our results, we estimate that a marginal audit of a taxpayer in the 90–99th percentiles produces a return of \$12.5:1.

Lastly, we use the MVPF framework to consider the welfare implications of these results. We estimate that the MVPF of auditing taxpayers in the 90–99th income percentiles is 1.15. By contrast, the MVPF of auditing taxpayers in the 20–30th income percentiles is 1.30. These results suggest that current audit levels across the income distribution are inconsistent with a wide range of canonical social preferences.

There is ample room for future work to build upon our findings. For example, our analysis measures the individual deterrence effect of IRS audits, but it does not measure general deterrence. If auditing an individual has spillover effects on the decisions of others that further increase government revenue, that could substantially increase the returns to audits. Moreover, our measures of average and marginal costs capture the steady state that existed for filings between tax years 2010 to 2014. Hiring and training new auditors requires a substantial upfront cost before those auditors begin to yield revenue. Future work should explore how the time path of hiring costs and the accumulation of auditor expertise shape the returns to audits. Finally, the analysis here is restricted to IRS in-person audits of individuals. While our work captures audits of individual income generated by businesses, future work should further examine the returns to auditing businesses themselves.

References

Advani, A., W. Elming, and J. Shaw (2023, May). The dynamic effects of tax audits. *The Review of Economics and Statistics 105*(3), 545–561.

Allingham, M. G., A. Sandmo, et al. (1972). Income tax evasion: A theoretical analysis. *Taxation: critical perspectives on the world economy 3*(1), 323–338.

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

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of political economy 76*(2), 169–217.

Beer, S., M. Kasper, E. Kirchler, and B. Erard (2020). Do audits deter or provoke future tax noncompliance? evidence on self-employed taxpayers. *CESifo Economic Studies 66*(3), 248–264.

Best, M., J. Shah, and M. Waseem (2021). Detection without deterrence: Long-run effects of tax audit on firm behavior. Technical report, Tech. rep., Mimeo, University of Manchester.

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

Border, K. C. and J. Sobel (1987). Samurai accountant: A theory of auditing and plunder. *The Review of economic studies 54*(4), 525–540.

CBO (2018). Options for reducing the deficit: 2019 to 2028. Technical report, Washington, DC: Congressional Budget Office.

CBO (2020). Trends in the internal revenue service's funding and enforcement. Technical report, Washington, DC: Congressional Budget Office.

DeBacker, J., B. Heim, A. Tran, and A. Yuskavage (2020). Tax noncompliance and measures of income inequality. *Tax Notes, February 17.*

DeBacker, J., B. T. Heim, A. Tran, and A. Yuskavage (2018). Once bitten, twice shy? the lasting impact of enforcement on tax compliance. *The Journal of Law and Economics 61*(1), 1–35.

Elzayn, H., E. Smith, T. Hertz, A. Ramesh, R. Fisher, D. E. Ho, and J. Goldin (2023). Measuring and mitigating racial disparities in tax audits.

Guyton, J. and R. Hodge (2014). The compliance costs of irs post-filing processes. In *An IRS-TPC Research Conference: Advancing Tax Administration*.

Guyton, J., P. Langetieg, D. Reck, M. Risch, and G. Zucman (2021). Tax evasion at the top of the income distribution: theory and evidence. Technical report, National Bureau of Economic Research.

Hebous, S., Z. Jia, K. Løyland, T. O. Thoresen, and A. Øvrum (2020). Do audits improve future tax compliance in the absence of penalties? evidence from random audits in norway.

Hendren, N. (2020). Measuring economic efficiency using inverse-optimum weights. *Journal of public Economics 187*.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

Hodge, R. H., A. H. Plumley, K. Richison, G. Yismaw, N. Misek, M. Olson, and H. S. Wijesinghe (2015). Estimating marginal revenue/cost curves for correspondence audits. *IRS Research Bulletin*.

Holtzblatt, J. and J. McGuire (2016). *Factors Affecting Revenue Estimates of Tax Compliance Proposals*. Congressional Budget Office, Joint Committee on Taxation.

Holtzblatt, J. and J. McGuire (2020). Effects of recent reductions in the internal revenue service's appropriations on returns on investment. In *IRS Research Bulletin: Proceedings of the 2020 IRS/TPC Research Conference*.

IRS (2022). Federal tax compliance research: Tax gap estimates for tax years 2014–2016. Technical report, Research, Applied Analytics & Statistics, Internal Revenue Service.

Johns, A. and J. Slemrod (2010). The distribution of income tax noncompliance. *National Tax Journal 63*(3), 397–418.

Kaplow, L. (1990). Optimal taxation with costly enforcement and evasion. *Journal of Public Economics 43*(2), 221–236.

Kasper, M. and J. Alm (2022a). Audits, audit effectiveness, and post-audit tax compliance. *Journal of Economic Behavior & Organization 195*, 87–102.

Kasper, M. and J. Alm (2022b). Does the bomb crater effect really exist? evidence from the laboratory. *FinanzArchiv 78*(1-2), 87–111.

Keen, M. and J. Slemrod (2017). Optimal tax administration. *Journal of Public Economics 152*, 133–142.

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*(3), 651–692.

Kleven, H. J. and C. T. Kreiner (2006). The marginal cost of public funds: Hours of work versus labor force participation. *Journal of Public Economics 90*(10-11), 1955–1973.

Løyland, K., O. Raaum, G. Torsvik, and A. Øvrum (2023). Evaluating compliance gains of expanding tax enforcement. *OFS Working Paper*.

Mayshar, J. (1990). On measures of excess burden and their application. *Journal of Public Economics 43*(3), 263–289.

Mayshar, J. (1991). Taxation with costly administration. *The Scandinavian Journal of Economics*, 75–88.

Mazzolini, G., L. Pagani, and A. Santoro (2022). The deterrence effect of real-world operational tax audits on self-employed taxpayers: evidence from italy. *International Tax and Public Finance 29*(4), 1014–1046.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of public Economics 6*(4), 327–358.

Saez, E. (2001). Using elasticities to derive optimal income tax rates. *The review of economic studies 68*(1), 205–229.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45.

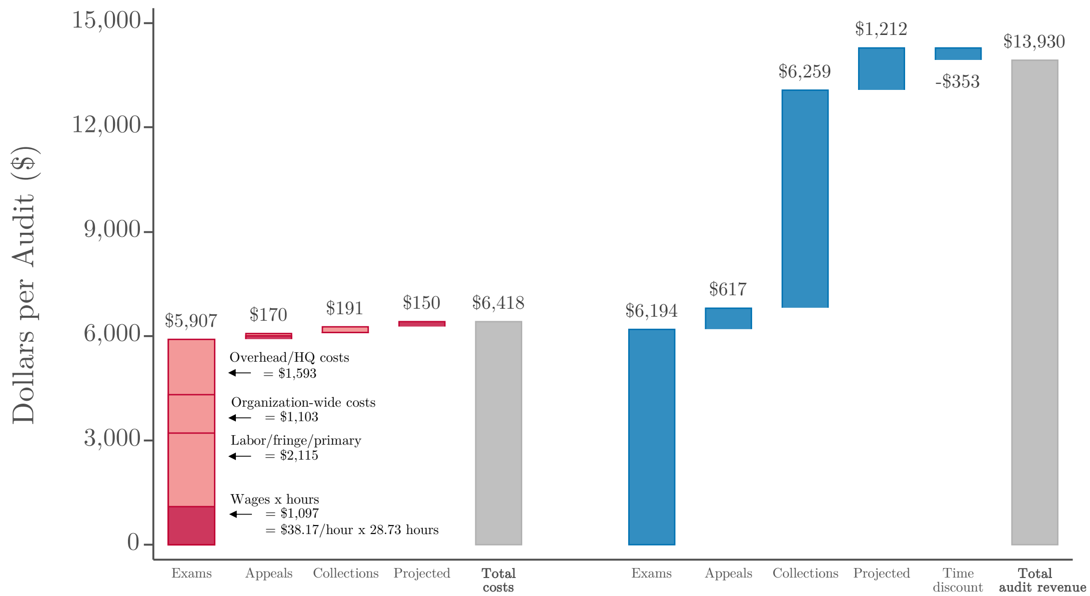
Sarin, N. and L. H. Summers (2019). Shrinking the tax gap: approaches and revenue potential. Technical report, National Bureau of Economic Research.

Sarin, N. and L. H. Summers (2020). Understanding the revenue potential of tax compliance investment. Technical report, National Bureau of Economic Research.

Slemrod, J., M. Blumenthal, and C. Christian (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of public economics 79*(3), 455–483.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. In *Handbook of public economics*, Volume 3, pp. 1423–1470. Elsevier.

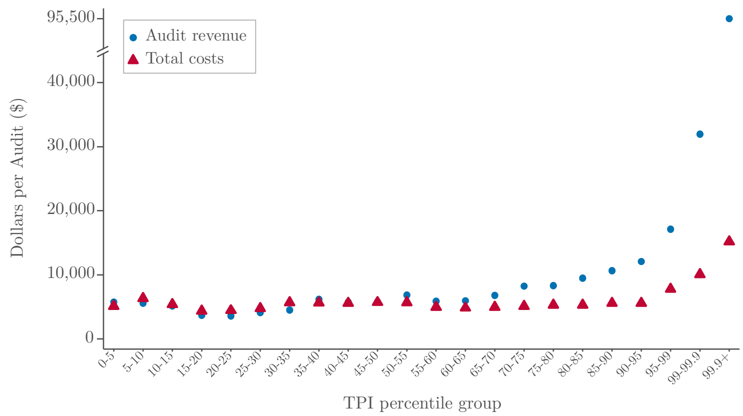
FIGURE 1: Average Costs and Revenue Raised per In-Person Audit



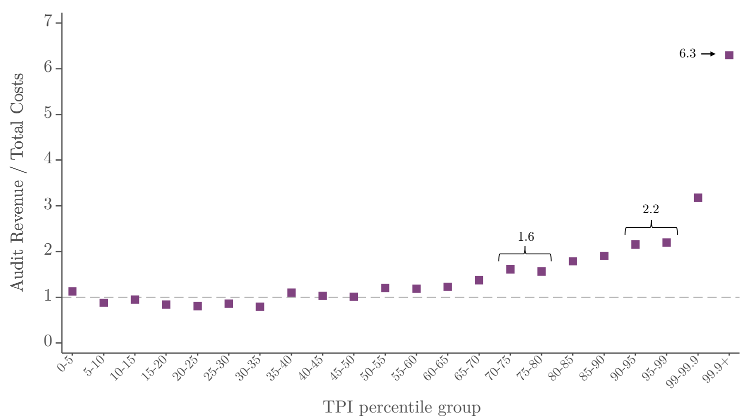
Notes: This figure presents the average total costs and revenue raised per in-person audit of a tax return filed for tax years between 2010 and 2014. Total wage costs (auditors' wages times hours spent on the exam) are shown in dark red and additional costs are shown stacked on top in lighter red. Additional costs include labor/fringe/primary, organization-wide, and overhead/HQ costs. Together, these additional costs are 4.39, 0.57, and 4.15 times total wage costs at the exam, appeals, and collections stage respectively using average multiplier values from 2011–2015. Revenue raised at each stage of the audit process is shown in blue and includes revenue raised from additional tax liability, penalties and interest. For details on each stage of the audit process, see Appendix Figure A1. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

FIGURE 2: Average Costs, Revenue and Revenue over Costs per In-Person Audit, by Income Group

A. Average Total Costs and Audit Revenue



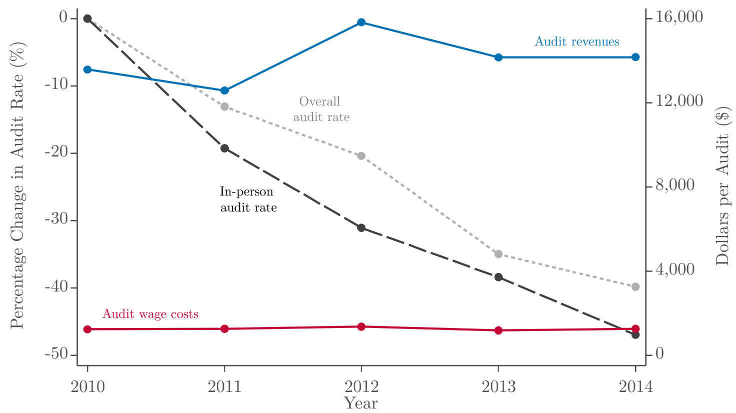
B. Average Audit Revenue over Total Costs



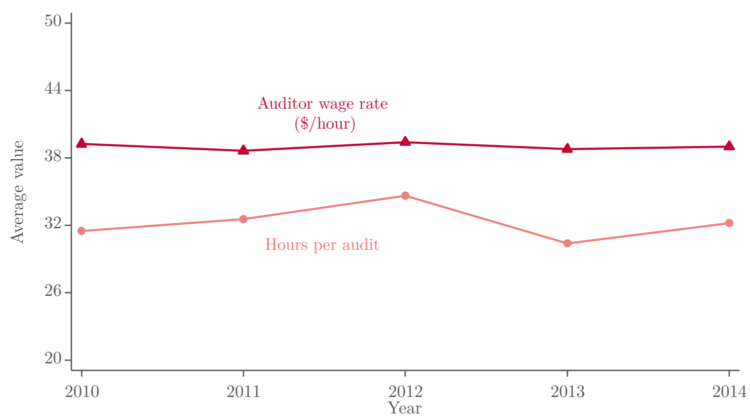
Notes: Panel A presents the average total costs and revenue raised per in-person audit of a tax return filed for tax years 2010–2014 by the taxpayer's total positive income (TPI). Panel B shows the ratio of the average revenue and costs per audit by TPI. The x-axis groups TPI into bins of five percentiles and splits out the top bin into the 95–99th and 99–99.9th percentiles and the top 0.1%. Total costs are the sum of labor costs (auditors' wages x hours spent on exam) and additional costs (labor/fringe/primary, organization-wide, and overhead/HQ costs), which are allocated in proportion to direct labor costs. Total revenue is the sum of revenue raised from additional tax liability, penalties and interest. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

FIGURE 3: Audit Probability, Revenue Collected and Wage Costs per In-Person Audit, by Year

A. Audit Probability, Audit Revenue Collected and Wage Costs



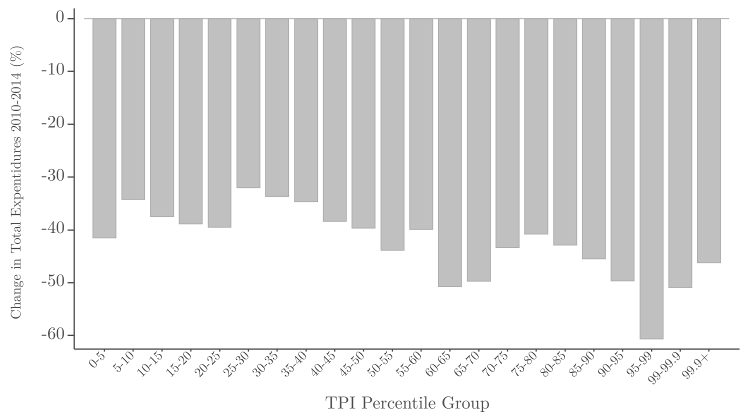
B. Components of Wage Costs



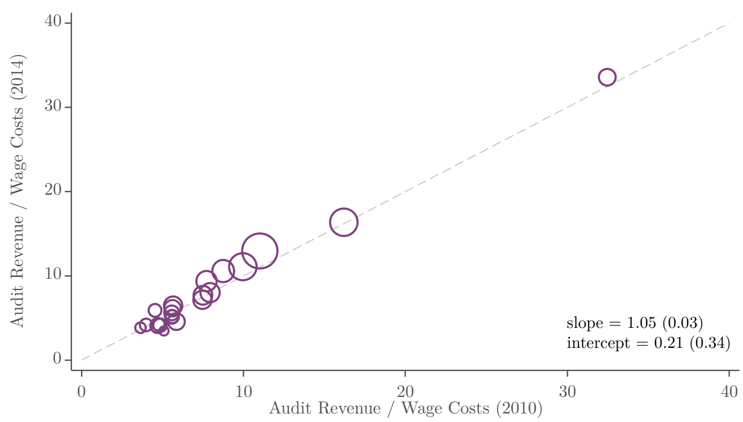
Notes: Panel A presents the percentage change in overall and in-person audit rates, total revenues raised, and direct labor costs (auditors' wages x hours spent on exam) per in-person audit for each tax year from 2010 to 2014. Panel B shows each component of labor costs (auditors' wages and hours worked per audit) by year. Total revenue is the sum of revenue raised from additional tax liability, penalties and interest. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

FIGURE 4: Change in Audit Expenditures and Average Total Revenue over Labor Costs for In-Person Audits

A. Change in Audit Expenditures Between 2010-2014

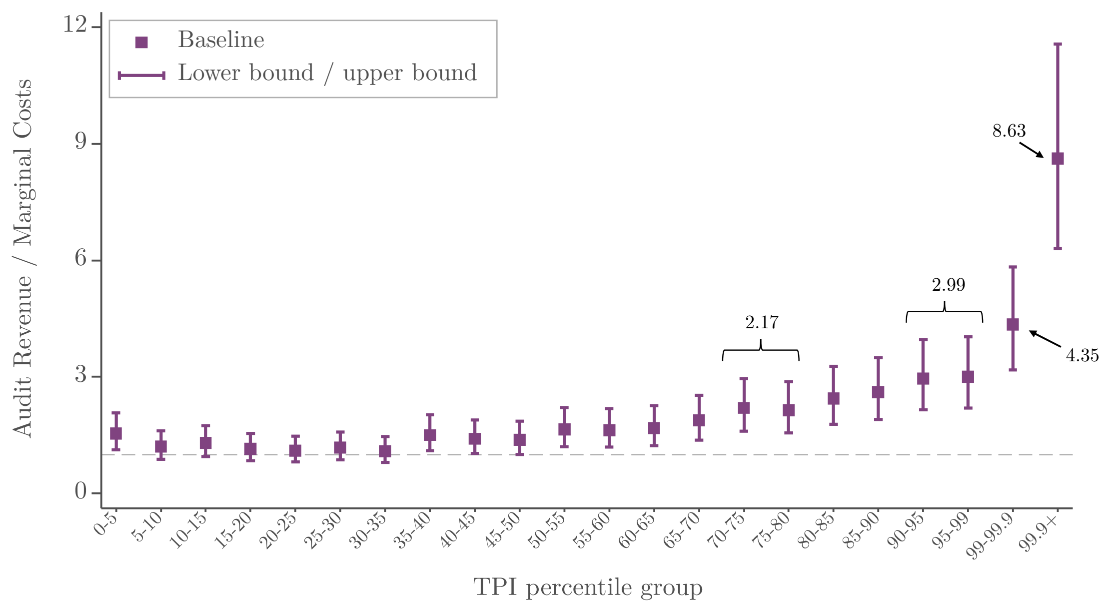


B. Change in Audit Revenue Over Wage Costs between 2010-2014



Notes: Panel A shows the percentage change in total direct wage costs of audits between the first tax year (2010) and the last tax year (2014) in our sample frame, by taxpayers' total positive income (TPI). Total direct wage costs are the product of labor costs per audit (auditors' wages x hours spent on exam) and the number of audits in each TPI bin. TPI is grouped into bins of five percentiles and the top bin is split into the 95–99th and 99–99.9th percentiles and the top 0.1%. Panel B shows the return (revenue per dollar of direct wage cost) to audits in each TPI bin for tax year 2014 against the values for tax year 2010. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

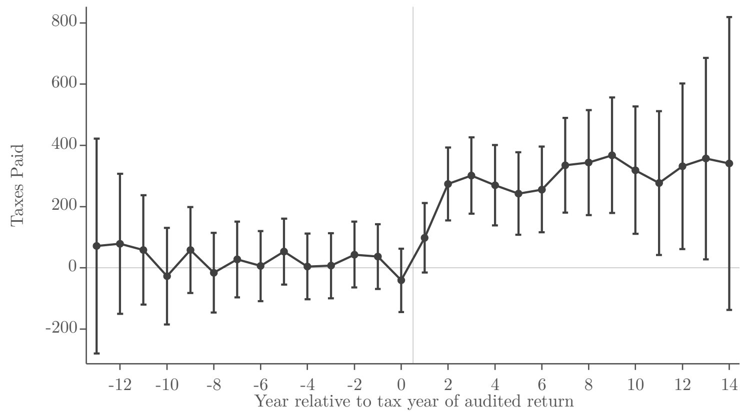
FIGURE 5: Average Audit Revenue over Marginal Costs



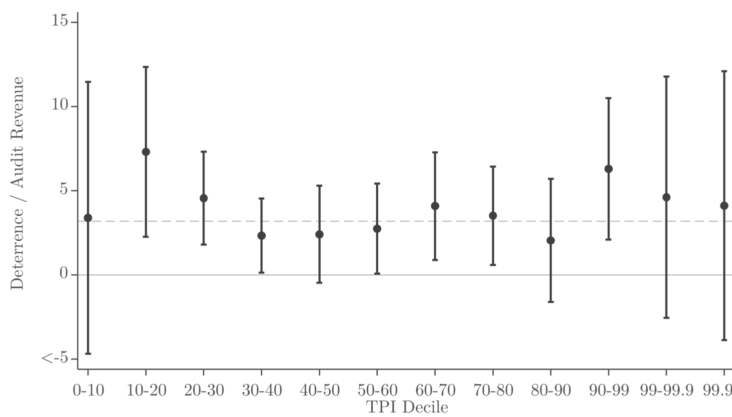
Notes: This figure shows the ratio of average revenue raised to total marginal costs (calculated using marginal overhead estimates) per in-person audit of a tax return filed for tax years 2010 to 2014 by the taxpayers' total positive income (TPI). Our baseline estimates use a marginal cost multiplier of $2.93 per dollar of audit wage costs. This estimate subtracts internal BU allocations and imputed costs (i.e. headquarters costs and costs from other IRS divisions and parts of the government other than IRS) from the total costs used to construct the average overhead multiplier. The lower bound case uses a marginal cost estimate of $1.93 per dollar of audit wage costs and includes only "primary" non-direct labor costs as variable costs (i.e. labor, benefits, training, travel). The upper bound case assumes all non-direct labor costs are variable and therefore uses a marginal cost estimate equal to the average overhead multiplier of $4.39 per dollar of audit wage costs. The x-axis groups TPI into bins of five percentiles and splits out the top bin into the 95–99th and 99–99.9th percentiles and the top 0.1%. Total costs are the sum of labor costs (auditors' wages x hours spent on exam) and marginal additional costs, which are allocated in proportion to direct labor costs. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

FIGURE 6: Estimated Deterrence Effects

A. Within-taxpayer impact of audits on future tax payments



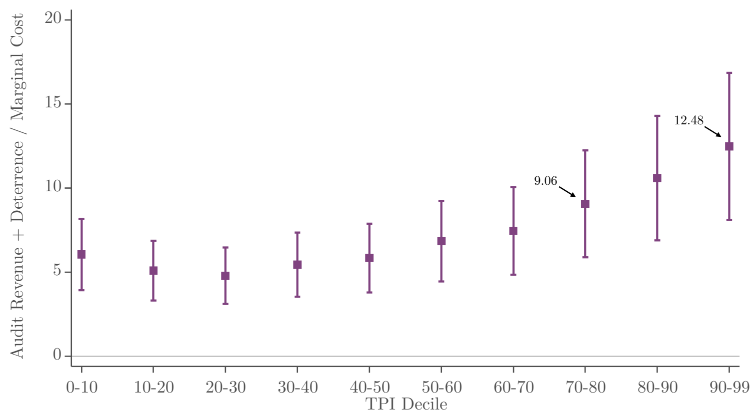
B. Deterrence effect over initial audit revenue, by income



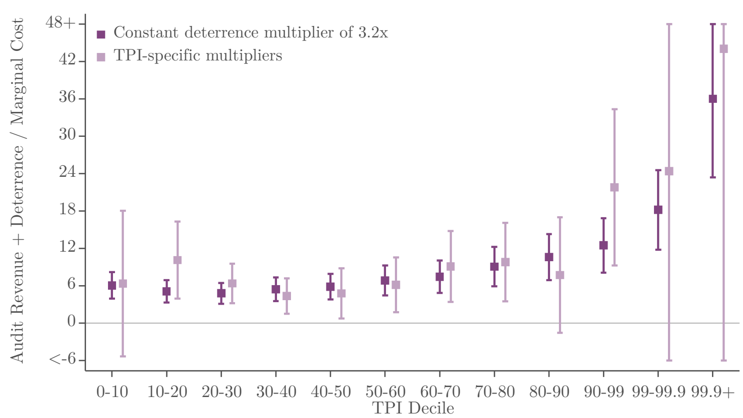
Notes: Panel A presents estimates of the change in taxes paid each year post-audit for the full population of individuals selected for random audit by the National Research Program (NRP) for tax years between 2006 and 2014. The control group is a matched sample of individuals not selected for random audit. Collected tax revenue is winsorized at the 99th percentile of the population distribution to limit the influence of outliers. The plotted estimates show the difference in taxes paid between control and treated individuals in each year in a single difference specification. Panel B shows the results of the treatment-control difference-in-difference specification scaled by revenue collected directly by the audit separately by bins of TPI. The reported deterrence effects are calculated as the ratio of the net present value (NPV) of total additional taxes paid post-audit to the NPV of upfront revenue raised by NRP audits. The dashed gray line shows the average multiplier across TPI bins associated with the single difference estimates from Panel A.

FIGURE 7: Deterrence Effect plus Initial Audit Revenue over Marginal Costs, by Income

A. Incorporating Overall Deterrence Effect

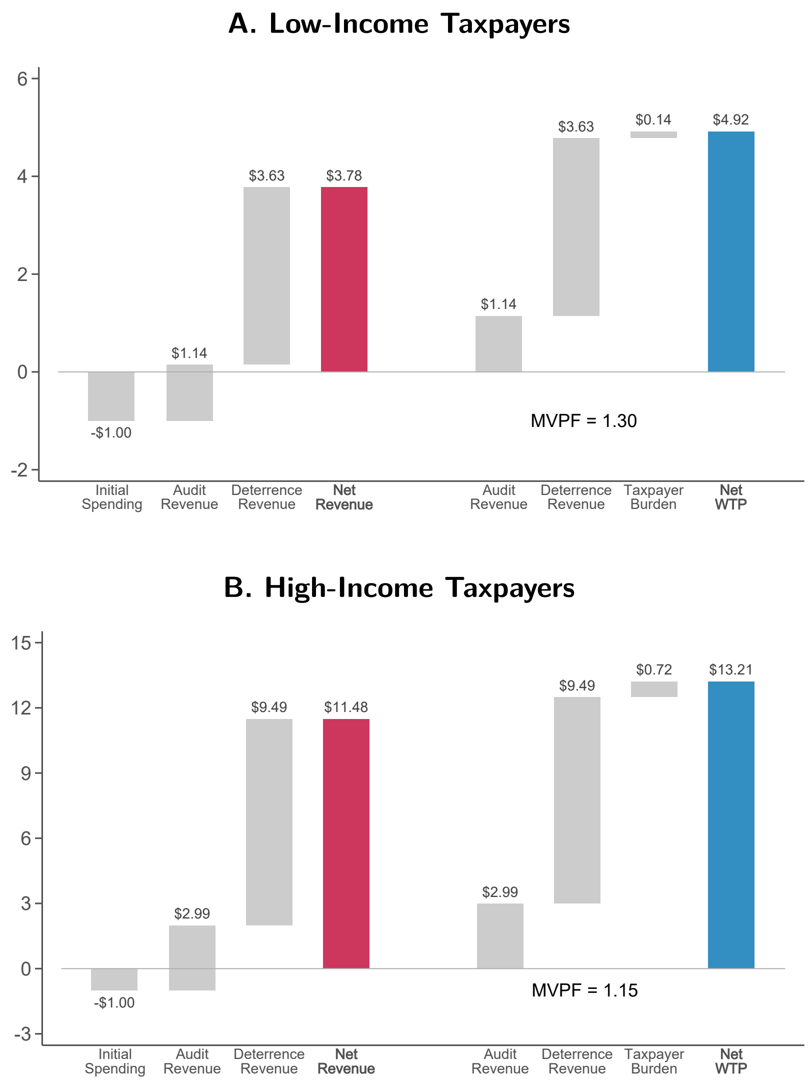


B. Incorporating TPI-Bin Individual Deterrence Effects



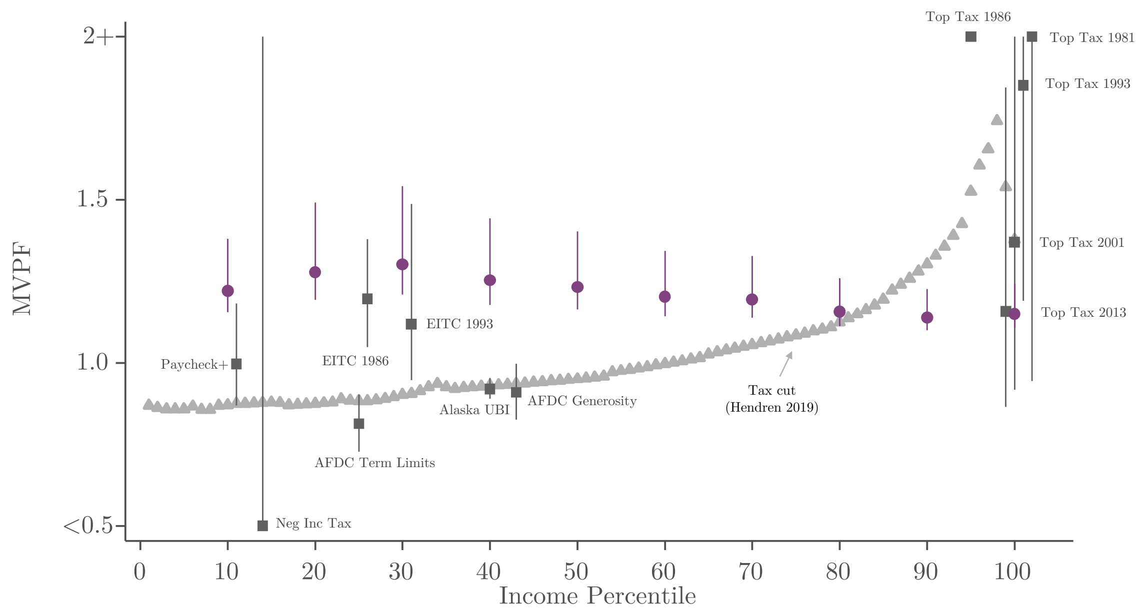
Notes: This figure shows the robustness of total revenue raised (deterrence and upfront audit revenue) per marginal dollar spent on audits across the income distribution to using different estimates of deterrence effects. These estimates are the product of one plus the deterrence multiplier and the baseline estimates of audit revenue divided by marginal costs (reported in Figure 6). The series plotted in darker purple uses the overall average deterrence multiplier from the single difference specification shown in Figure 6A. The series plotted in lighter purple uses the TPI bin-specific multipliers from a difference-in-differences specification shown in Figure 6B. In both cases, deterrence multipliers are the net present value (NPV) of total additional taxes paid post-random audit divided by the NPV of upfront revenue raised by the NRP audit. Upfront audit revenue per in-person audit of a return filed for tax years between 2010 and 2014 includes projected revenue collected and costs accrued after the observed 7–11 year post-audit sample window. These revenues are discounted using a 3% discount rate to align with the timing of costs. Baseline estimates of marginal costs incorporated a non-direct labor costs multiplier of $2.93 per dollar of audit wage costs. This estimate subtracts internal BU allocations and imputed costs (i.e. headquarters costs and costs from other IRS divisions and parts of the government other than IRS) from the total costs used to construct the average overhead multiplier.

FIGURE 8: The Marginal Value of Public Funds of Marginal Tax Audits



Notes: This figure presents the components of the MVPF of expanding audits among those with in the 20th–30th percentiles of the TPI distribution (Panel A) and the 90–99th percentiles of the TPI distribution (Panel B). The MVPF is the taxpayer's willingness to pay to avoid audit (shown in blue) divided by the net government revenue raised by the audit (shown in red). Net government revenue is upfront and deterrence revenue minus the cost of conducting the audit. Upfront revenue is the revenue raised per dollar of marginal spending (reported in figure 7). The deterrence effect is calculated by multiplying direct revenue by the relevant deterrence multiplier shown in Figure 7 Panel B. Net willingness to pay to avoid an audit includes the upfront taxes paid as a result of the audit and the downstream additional taxes paid due to deterrence effects, as well as the financial and time costs incurred in responding to the audit, which we estimate using an IRS survey. Time costs are monetized using taxpayer-specific average wages.

FIGURE 9: The MVPFs of Revenue Raising and Transfer Policies, by Income



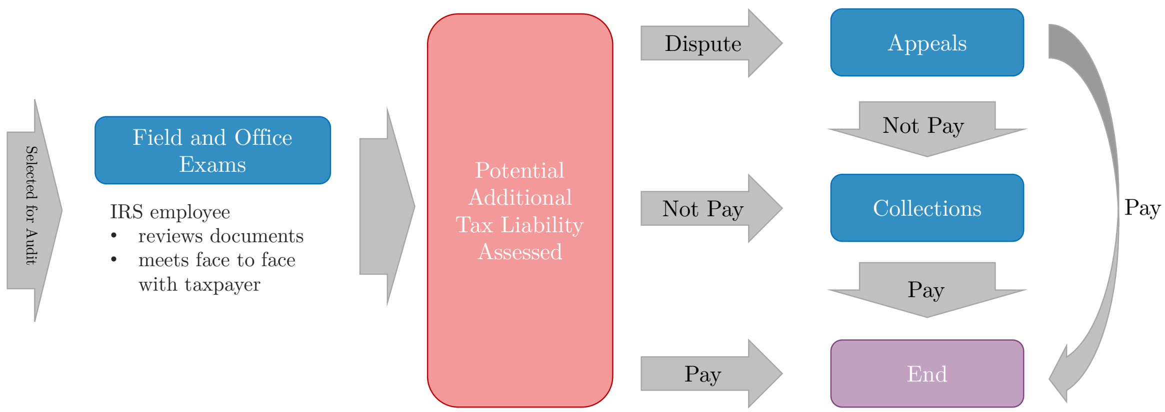
Notes: This figure compares the MVPF of tax audits to the MVPFs of a wide range of tax and transfer policies within the United States. The horizontal axis reports the quantiles of the income distribution, displaying how these MVPFs vary with income. The gray triangles report MVPF estimates for tax and transfer policies analyzed in Hendren and Sprung-Keyser (2020) which draw upon existing causal estimates to compute the MVPF for these policies. The black squares report estimates from Hendren (2020), which measures the MVPF of a small change in the tax schedule at each point of the income distribution. The purple circles show the MVPF of expanding tax audits by TPI decile as constructed in Figure 7.

Table 1: Deterrence effects versus mechanical audit revenue, by income

	Winsorized			Unwinsorized		
TPI Decile	Mechanical audit revenue	PDV total tax incl. self-emp minus EITC [95% CI]	Deterrence Multiplier	Mechanical audit revenue	PDV total tax incl. self-emp minus EITC [95% CI]	Deterrence Multiplier
0-10	400.89	1354.76 [-1883.91, 4593.42]	3.38 [-4.70, 11.46]	426.20	1510.64 [-2244.12, 5265.40]	3.54 [-5.27, 12.35]
10-20	444.89	3248.50 [999.72, 5497.27]	7.30 [2.25, 12.36]	462.36	7161.34 [-1748.29, 16070.97]	15.49 [-3.78, 34.76]
20-30	640.78	2920.62 [1153.00, 4688.24]	4.56 [1.80, 7.32]	663.15	2861.65 [847.11, 4876.20]	4.32 [1.28, 7.35]
30-40	764.16	1782.24 [102.27, 3462.20]	2.33 [0.13, 4.53]	785.02	2145.27 [134.08, 4156.45]	2.73 [0.17, 5.29]
40-50	869.07	2095.34 [-404.53, 4595.20]	2.41 [-0.47, 5.29]	901.44	1225.10 [-2600.50, 5050.69]	1.36 [-2.88, 5.60]
50-60	1000.79	2745.07 [62.29, 5427.85]	2.74 [0.06, 5.42]	1061.92	3036.64 [-192.24, 6265.51]	2.86 [-0.18, 5.90]
60-70	1105.77	4517.27 [987.27, 8047.26]	4.09 [0.89, 7.28]	1183.69	6421.21 [1251.31, 11591.12]	5.42 [1.06, 9.79]
70-80	1174.29	4119.87 [698.47, 7541.26]	3.51 [0.59, 6.42]	1250.96	4168.65 [416.83, 7920.47]	3.33 [0.33, 6.33]
80-90	1302.59	2661.52 [-2108.13, 7431.17]	2.04 [-1.62, 5.70]	1435.94	1412.17 [-7132.02, 9956.36]	0.98 [-4.97, 6.93]
90-99	1757.80	11054.78 [3664.31, 18445.26]	6.29 [2.08, 10.49]	2199.03	16031.50 [-4391.24, 36454.23]	7.29 [-2.00, 16.58]
99-99.9	3027.57	13945.91 [-7774.43, 35666.24]	4.61 [-2.57, 11.78]	6663.11	-53104.94 [-2.13e+05, 106903.60]	-7.97 [-31.98, 16.04]
99.9+	3567.96	14653.04 [-13843.91, 43149.98]	4.11 [-3.88, 12.09]	13576.52	-1597725.00 [-3.92e+06, 728385.40]	-117.68 [-289.02, 53.65]

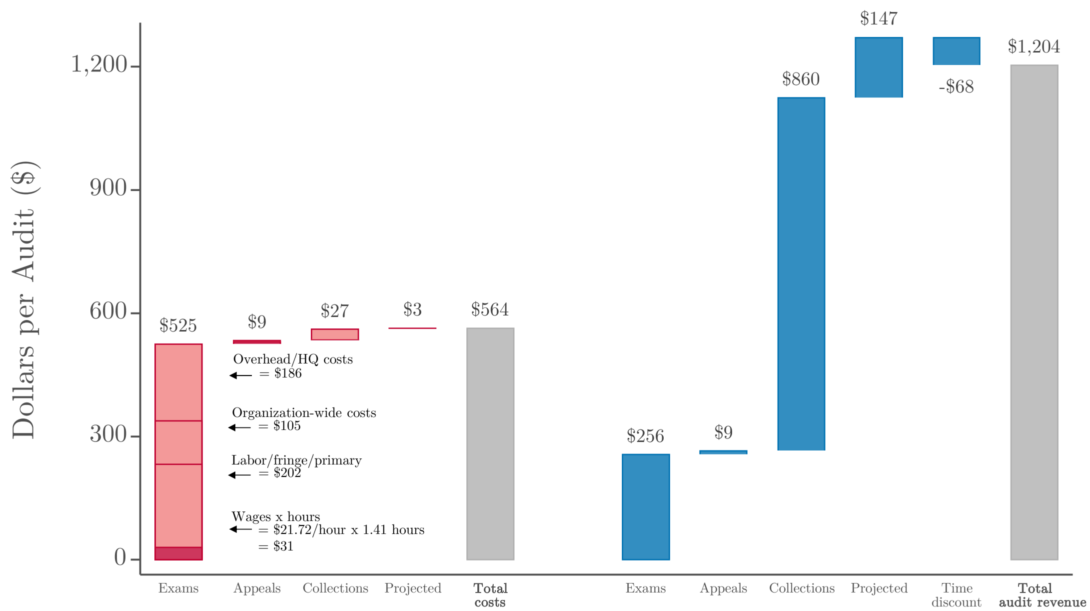
Notes: This table presents individual deterrence effects versus mechanical audit revenue by deciles of taxpayer's total positive income (TPI). Deterrence effects are estimated from a difference in differences regression that compares total taxes paid up to 13 years post-audit for individuals selected for random audit versus a matched control group who was not selected for random audit. Column (3) presents deterrence estimates winsorized at the 99.9th percentile and column (5) presents the unwinsorized results. Mechanical and deterrence revenues are discounted back to the audit tax year using a 3% discount rate. The deterrence multiplier is then calculated as the ratio of deterrence and mechanical revenue. 95% confidence intervals are reported in square parentheses.

APPENDIX FIGURE A1: Overview of the Audit Process



Notes: This flow chart provides an overview of the audit process beginning when a tax return is selected for a field or office exam. The exam begins when an IRS employee reviews the taxpayer's relevant documents and meets face-to-face with the taxpayer to determine any adjustments to tax liability. If there is no audit adjustment then the enforcement process concludes. If the taxpayer agrees with the audit adjustment then the change in tax liability is assessed, while if the taxpayer disagrees the case is heard by the IRS' Independent Office of Appeals and may be further appealed to tax court for a final determination and assessment. If the taxpayer does not pay any additional assessed liability, the case is sent to collections. In practice, few exams are appealed or sent to collections. Our estimates include the taxes, penalties, and interest collected and costs accrued at the exam, appeals, and collections stages of the audit process.

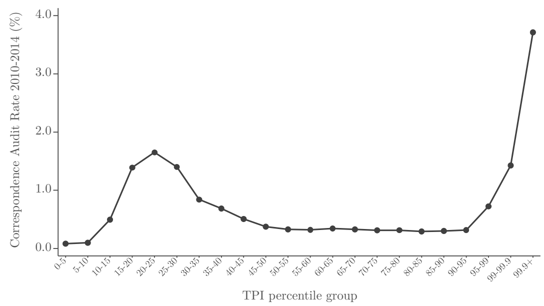
APPENDIX FIGURE A2: Average Costs and Revenue Raised per Correspondence Audit



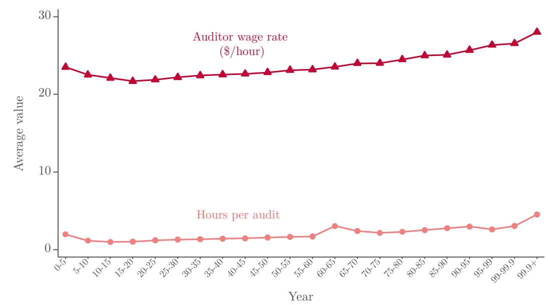
Notes: This figure presents the average total costs and revenue raised per correspondence audit of a tax return filed for tax years 2010–2014. Total wage costs (auditors' wages times hours spent on exam) are shown in dark red and additional costs are shown stacked on top in lighter red. Additional costs include labor/fringe/primary, organization-wide, and overhead/HQ costs. Together, these additional costs are 13.94, 0.57, and 4.15 times total wage costs at the exam, appeals, and collections stages respectively using average multiplier values from tax years 2011–2015. Revenue raised at each stage of the audit process is shown in blue and includes revenue raised from additional tax liability, penalties and interest. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

APPENDIX FIGURE A3: Correspondence Audits Across the Income Distribution

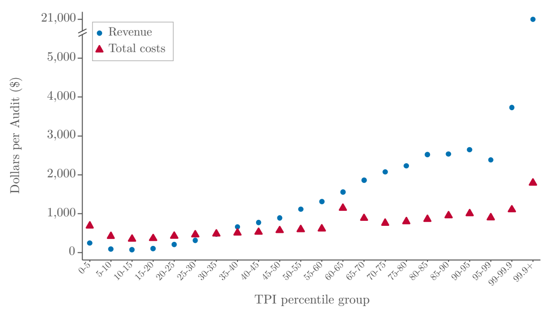
A. Audit Rates



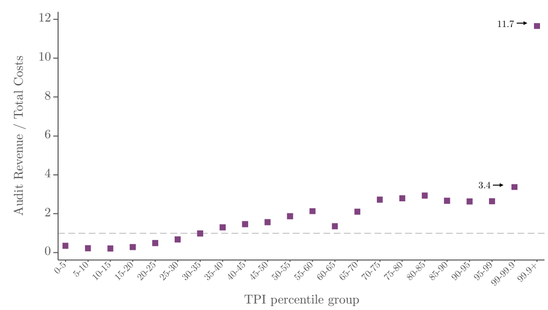
B. Wage Costs



C. Audit Revenue and Total costs



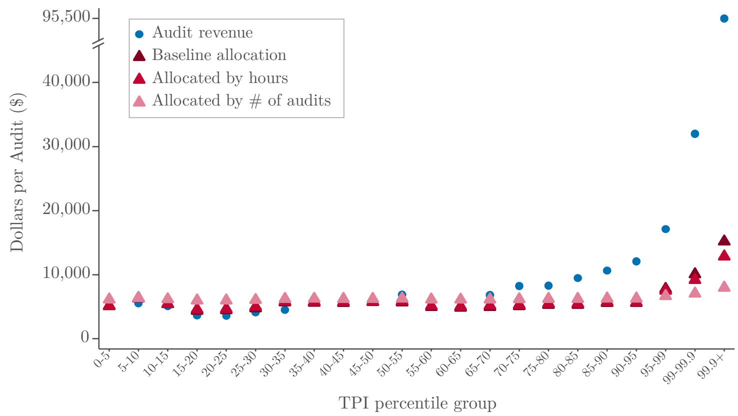
D. Audit revenue over Total Costs



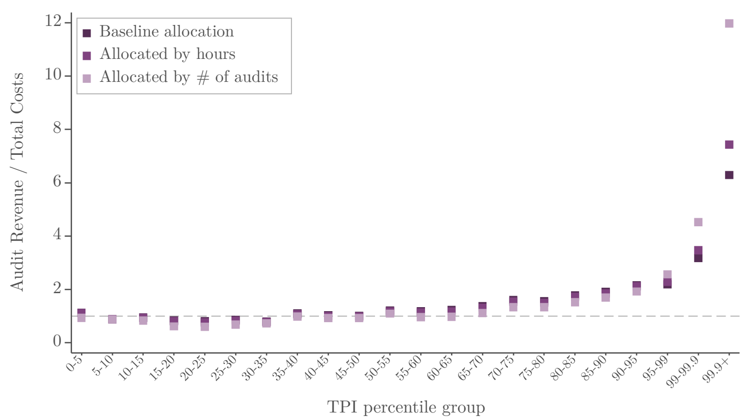
Notes: This figure presents the average audit rate, costs and revenues for correspondence audits of tax returns filed for tax years 2010–2014 by the taxpayer's total positive income (TPI). Panel A presents the average audit rate for correspondence audits. Panel B shows each component of labor costs (auditors' wages and hours worked per audit) per correspondence audit. Panel C presents the average total costs and revenue raised per audit, and Panel D shows the ratio of the average revenue and costs per audit by TPI. The x-axis groups TPI into bins of five percentiles and splits out the top bin into the 95–99th and 99–99.9th percentiles and the top 0.1%. Total costs are the sum of labor costs (auditors' wages times hours spent on exam) and additional costs (labor/fringe/primary, organization-wide, and overhead/HQ costs) which are allocated in proportion to direct labor costs. Total revenue is the sum of additional tax liability, penalties and interest collected. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

APPENDIX FIGURE A4: Average Costs, Revenue and Revenue over Costs per In-Person Audit with Alternative Non-Direct Labor Cost Allocations, by Income Group

A. Average Costs and Revenue



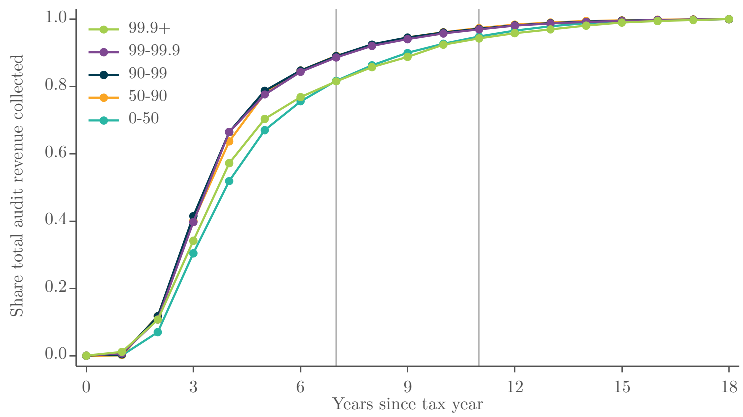
B. Average Revenue over Costs



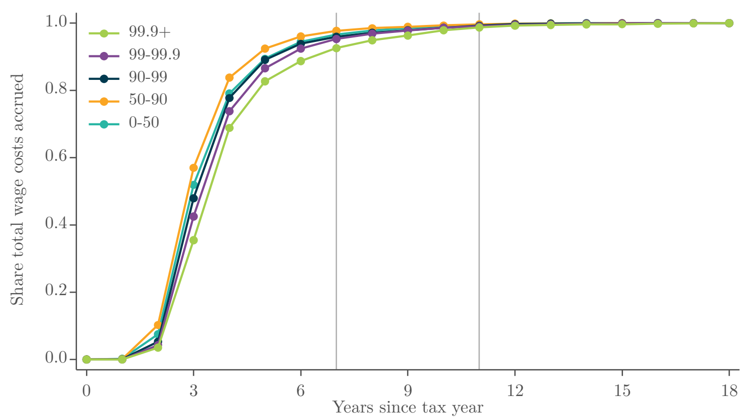
Notes: This figure presents the average total costs accrued and revenue raised per in-person audit (Panel A) and the ratio of average revenue and costs (Panel B) by the taxpayer's total positive income (TPI) using different methods to allocate overhead costs (non-direct labor-related costs, organization-wide costs, and general overhead costs) across the income distribution. The baseline method shown in dark purple allocates overhead costs in proportion to direct audit wage costs. The second method shown in the mid shade of purple allocates costs in proportion to labor hours rather than total labor costs. The third method shown in light purple allocates overhead costs equally per audit.

APPENDIX FIGURE A5: Cumulative Share of Revenue Collected and Labor Costs Accrued, by Years Post-Tax Year and Income

A. Audit Revenues



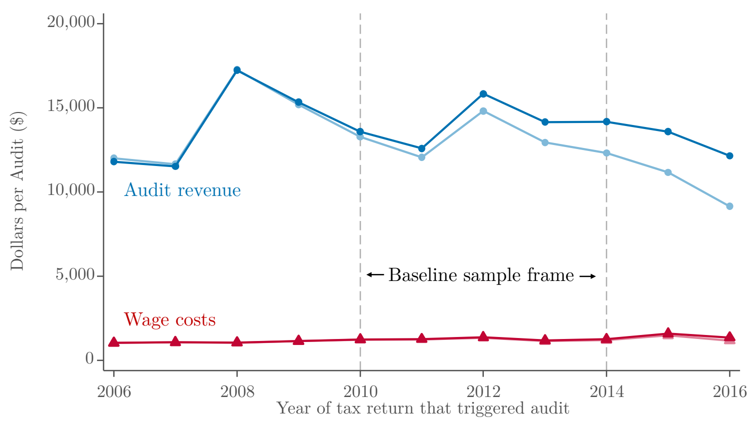
B. Labor Costs



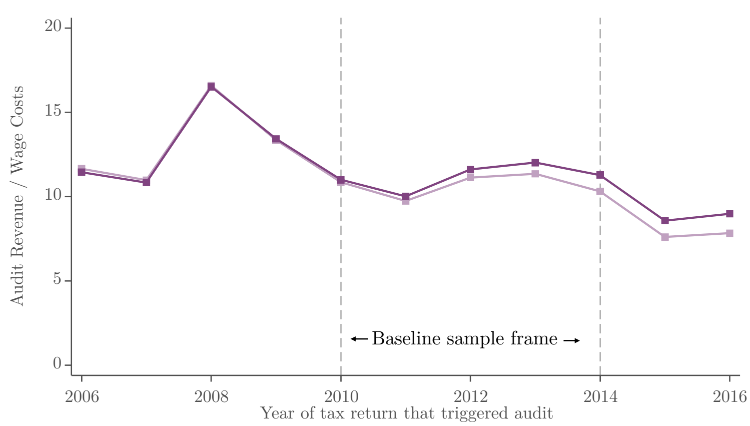
Notes: This figure presents the trajectory of revenues raised (Panel A) and costs accrued (Panel B) following audits of tax year 2003 returns. The 2003 tax year lies before our primary sample window, but shows 18 years of follow-up data. The y-axis shows the cumulative share of revenues collected and labor costs accrued relative to the total values 18 years post-tax year. The gray vertical lines indicate the 7–11 year windows observed after the tax year 2010 to 2014 returns in our primary sample were filed.

APPENDIX FIGURE A6: Average Costs, Revenue and Revenue over Costs per In-Person Audit, by Year

A. Average Costs and Revenue



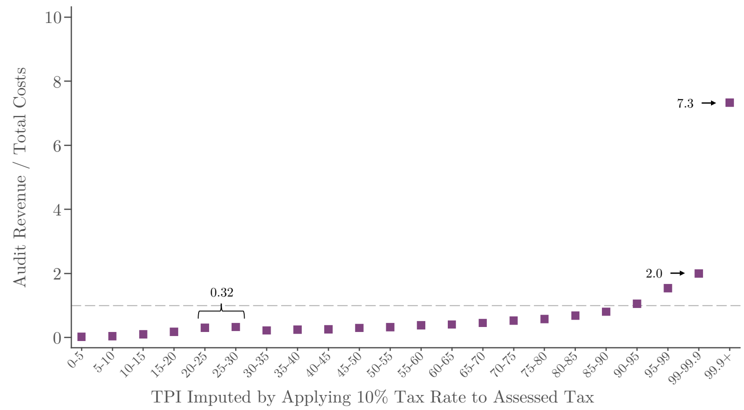
B. Average Revenue over Costs



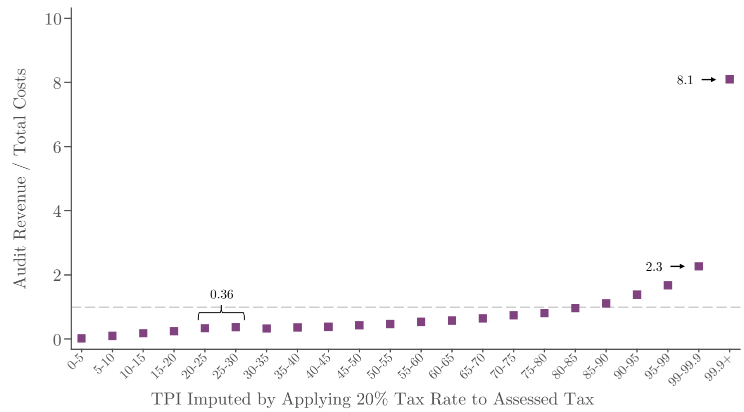
Notes: Panel A presents the average total costs and revenue raised per in-person audit by the tax year for which a return was filed. Panel B shows the ratio of the average revenue and costs per audit by tax year. The vertical gray dashed lines indicate our primary sample window. Total costs are the sum of labor costs (auditors' wages times hours spent on exam) and additional costs (labor/fringe/primary, organization-wide, and overhead/HQ costs). Total revenue is the sum of additional tax liability, penalties and interest collected. Average costs and revenues include projections of revenue collected and costs accrued outside the observed post-audit sample window for each tax year. The series plotted in the lighter shades of blue, red, and purple show the average values of revenues, costs, and revenues over costs without this projection adjustment. Revenues are discounted using a 3% discount rate since average revenues lag average costs by approximately one year. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

APPENDIX FIGURE A7: Average In-Person Audit Revenue Over Total Costs Using Imputed Post-Audit TPI

A. TPI Imputed by Applying 10% Tax Rate to Assessed Tax

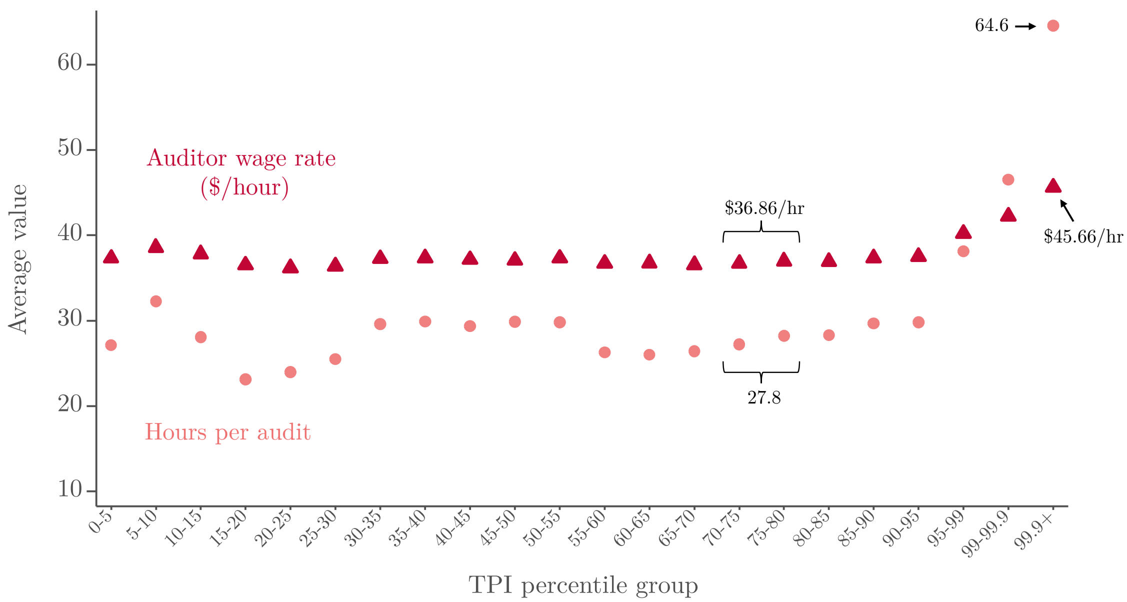


B. TPI Imputed by Applying 20% Tax Rate to Assessed Tax



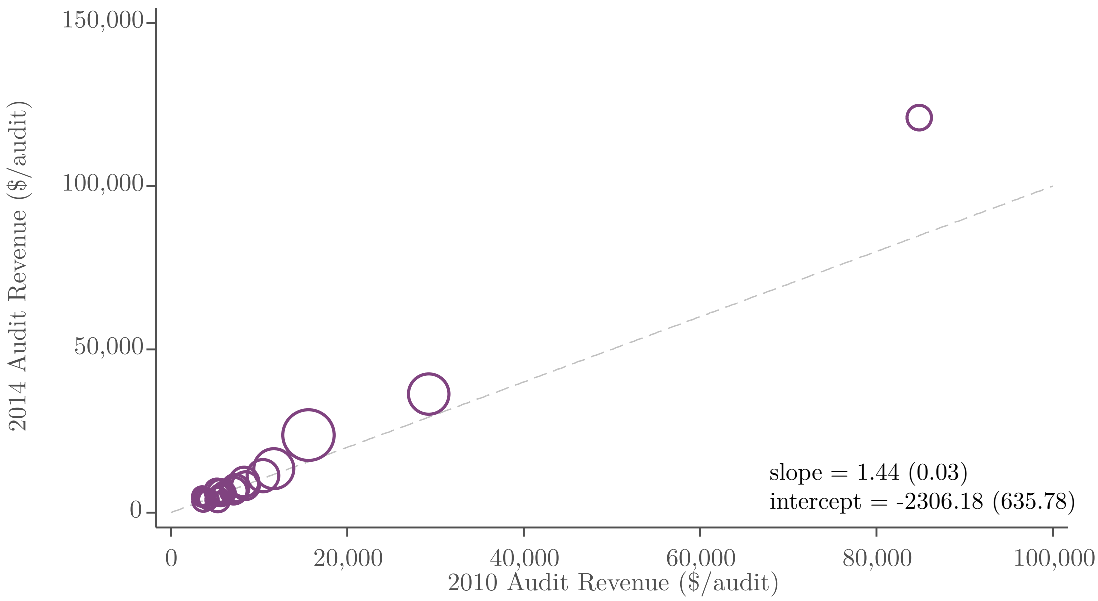
Notes: This figure repeats Figure 2 (Panel B) using imputed post-audit TPI on the horizontal axis instead of pre-audit reported TPI. We impute post-audit TPI by dividing the change in assessed tax by a 10% average marginal tax rate (Panel A) and by a 20% average marginal tax rate (Panel B) and adding the change to reported TPI.

APPENDIX FIGURE A8: Average Hours per In-Person Audit and Auditor Wage Rate, by Income Group



Notes: This figure presents the average values of the components of labor costs (auditors' wages times hours spent on exam) for in-person audits of tax returns filed for tax years 2010–2014 by the taxpayer's total positive income (TPI). The x-axis groups TPI into bins of five percentiles and splits out the top bin into the 95–99th and 99–99.9th percentiles and the top 0.1%. Average hours per audit include projected labor hours accrued after the observed 7–11 year post-audit sample window.

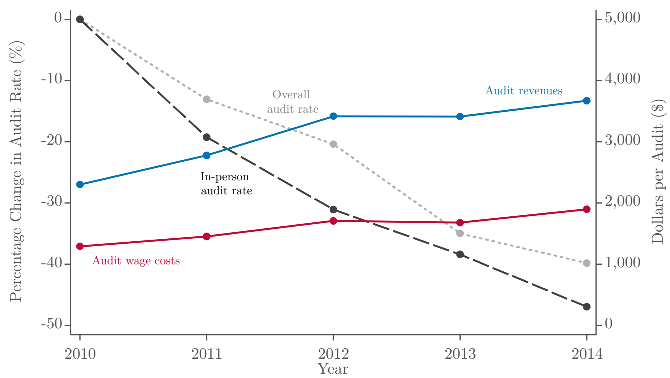
APPENDIX FIGURE A9: Marginal Revenue



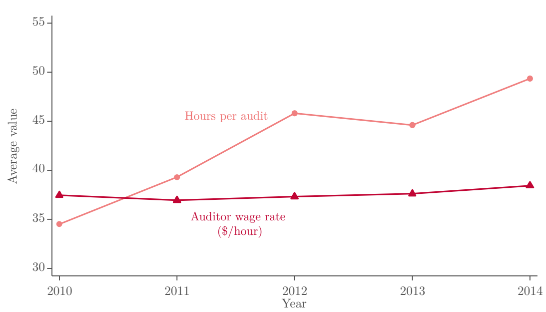
Notes: This figure shows average revenue raised per in-person audit for each 5-percentile TPI bin for audits of returns from tax year 2014 = against the values for the same TPI bins for audits of returns from tax year 2010. Average revenues include projected revenue collected and costs accrued after the observed 7 and 11 year post-audit sample window for 2014 and 2010 audits respectively. Revenues are discounted using a 3% discount rate since average revenues lag average costs by about one year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues divided by costs) to adjust revenues downwards to align the two paths.

APPENDIX FIGURE A10: Audit Probability, Revenue Collected and Wage Costs per NRP Audit, by Year

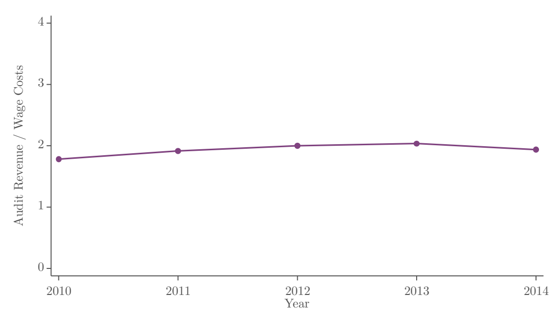
A. Audit Probability, Revenue Collected and Wage Costs



B. Components of Wage Costs



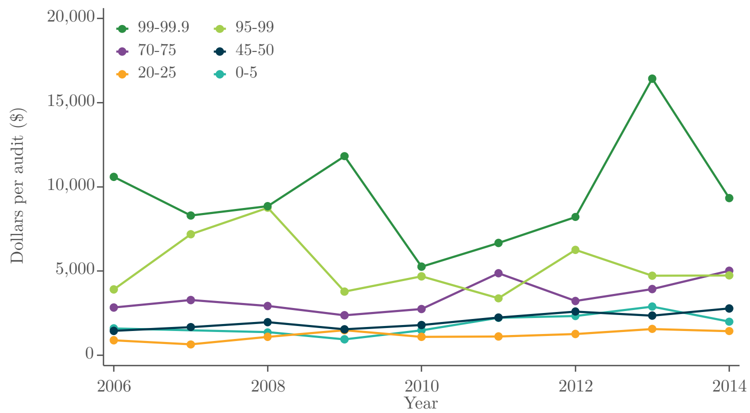
C. Audit revenue over Wage Costs



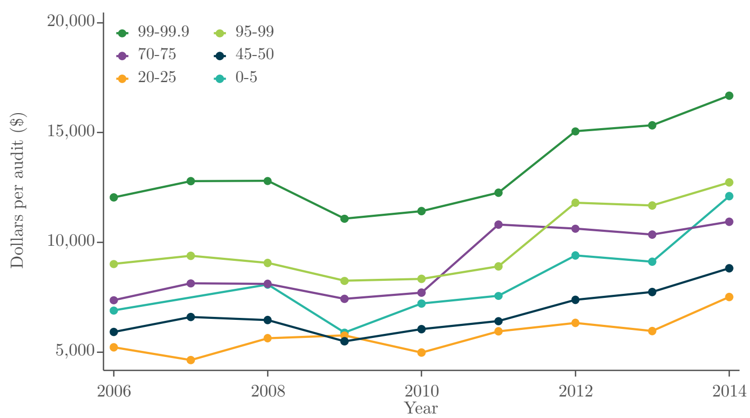
Notes: Panel A presents the percentage change in overall and in-person audit rates, total revenues raised, and direct labor costs (auditors' wages times hours spent on exam) per National Research Program (NRP) random audit for NRP study tax years in our sample frame (2010–2014). Panel B shows each component of labor costs (auditors'wages and hours worked per audit) by year. Panel C shows average revenue per audit divided by costs per audit for each NRP study tax year. Total revenue is the sum of additional tax liability, penalties and interest collected. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues divided by costs) to adjust revenues downwards to align the two paths.

APPENDIX FIGURE A11: Average Revenue and Costs per NRP Audit, by Income and Year

A. Average Audit Revenue

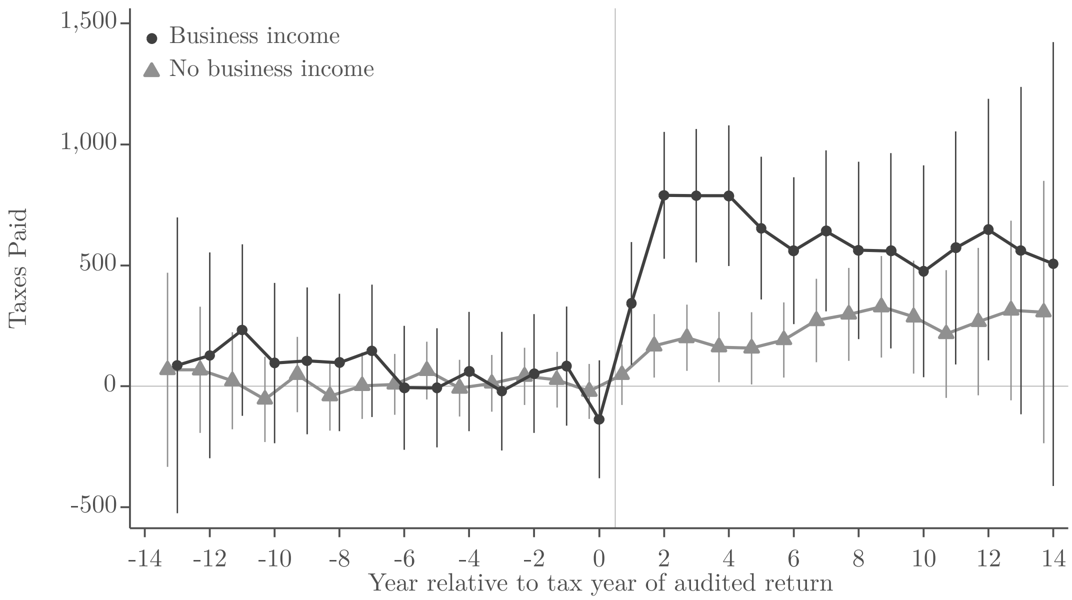


B. Average Costs



Notes: This figure presents total audit revenues raised (Panel A) and total costs accrued (Panel B) per random audit in each National Research Program (NRP) study tax year for select total positive income (TPI) percentile bins. Total revenue is the sum of additional tax liability, penalties and interest collected. Total costs are the sum of labor costs (auditors' wages times hours spent on exam) and additional costs (labor/fringe/primary, organization-wide, and overhead/HQ costs), which are allocated in proportion to direct labor costs. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11 year post-audit sample window. Revenues are discounted using a 3% discount rate because revenues lag costs by about a year on average. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues divided by costs) to adjust revenues downwards to align the two paths.

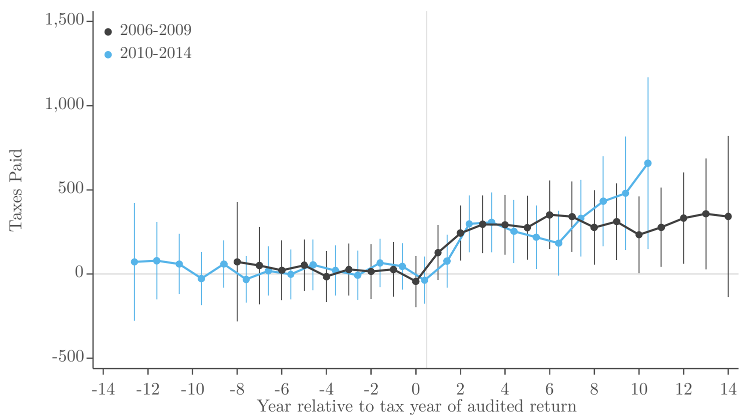
APPENDIX FIGURE A12: Within-Taxpayer Impact of Audits on Future Tax Payments, by Presence of Income



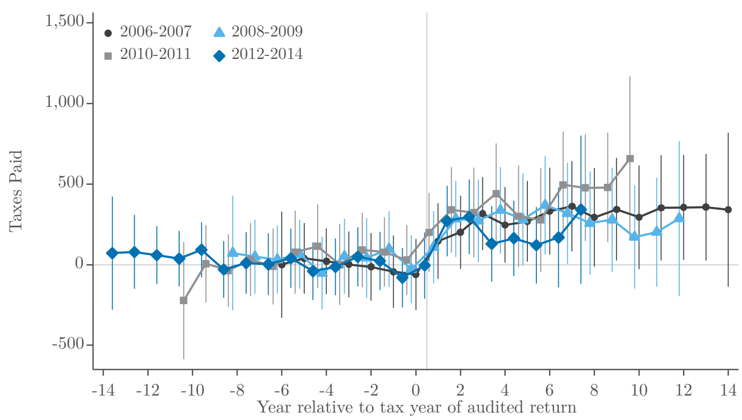
Notes: This figure presents estimates of the change in taxes paid each year post-audit for individuals selected for random audit by the National Research Program (NRP) separately for individuals with and without business income (as measured by income on Schedule C, E and F). The control group is a matched sample of individuals not selected for random audit. Collected tax revenue is winsorized at the 99th percentile of the population distribution to limit the influence of outliers. The plotted estimates show the difference in taxes paid between control and treated individuals in each year in a single difference specification.

APPENDIX FIGURE A13: Within-Taxpayer Impact of Audits on Future Tax Payments, by NRP Study Year

A. 4 year groupings

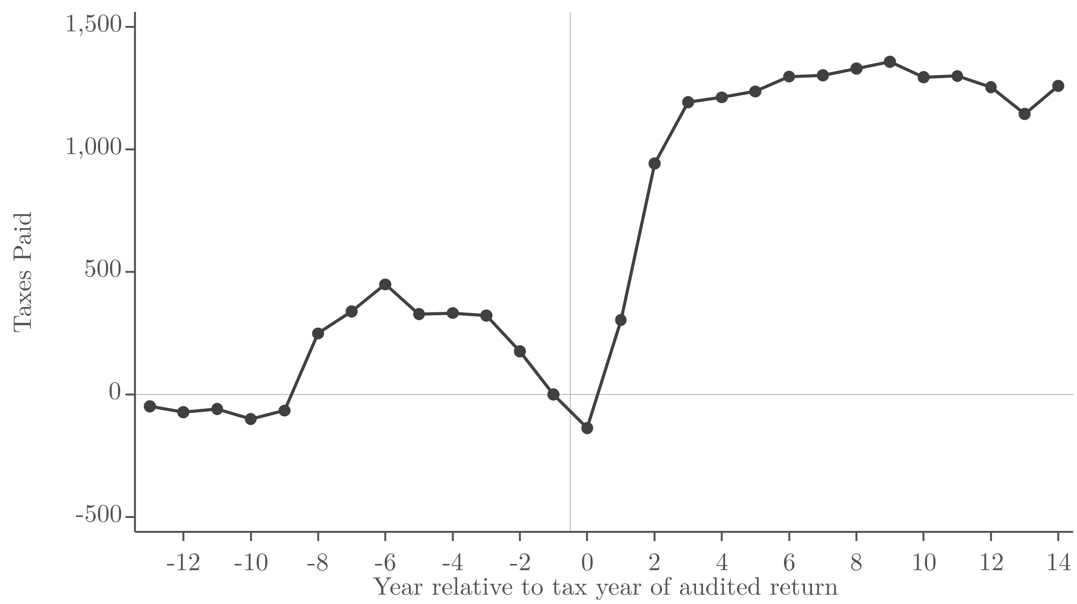


B. 2–3 year groupings



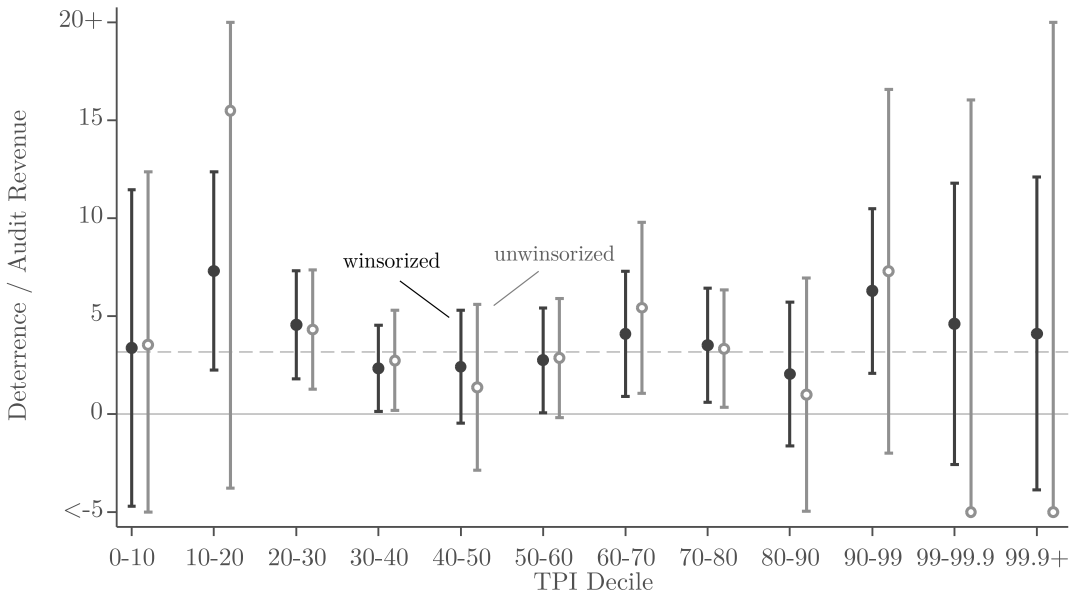
Notes: This figure presents estimates of the change in taxes paid for each year post-audit for the full population of individuals selected for random audit by the National Research Program (NRP), by groups of NRP study tax years. The control group is a matched sample of individuals not selected for random audit. Collected tax revenue is winsorized at the 99th percentile of the population distribution to limit the influence of outliers. The plotted estimates show the differences in taxes paid between control and treated individuals in each year in a single difference specification.

APPENDIX FIGURE A14: Within-Taxpayer Impact of (Non-Random) In-Person Audits on Future Tax Payments



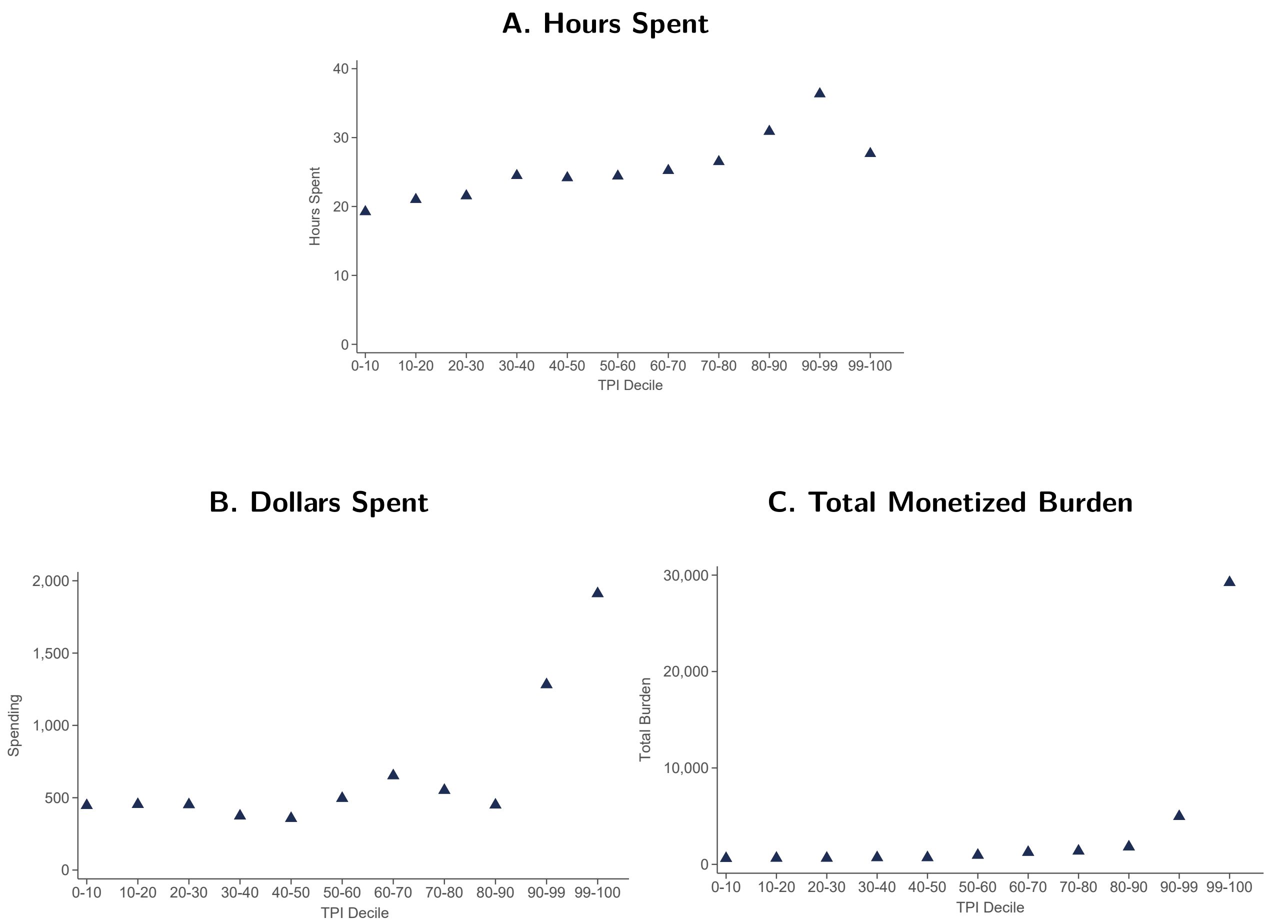
Notes: This figure presents estimates of the change in taxes paid in each year post-audit for individuals selected for an in-person audit. The control group is a matched sample of individuals not selected for audit. Coarsened matching is done based on income, lagged income, and return characteristics. The plotted estimates show the result of a difference-in-differences comparison in taxes paid. The figure compares treated and control individuals, comparing both to their respective taxes paid in the year before the audit. Collected tax revenue is winsorized at the 99th percentile of the population distribution to limit the influence of outliers.

APPENDIX FIGURE A15: Deterrence Effect over Initial Audit Revenue (Winsorized and Unwinsorized), by Income



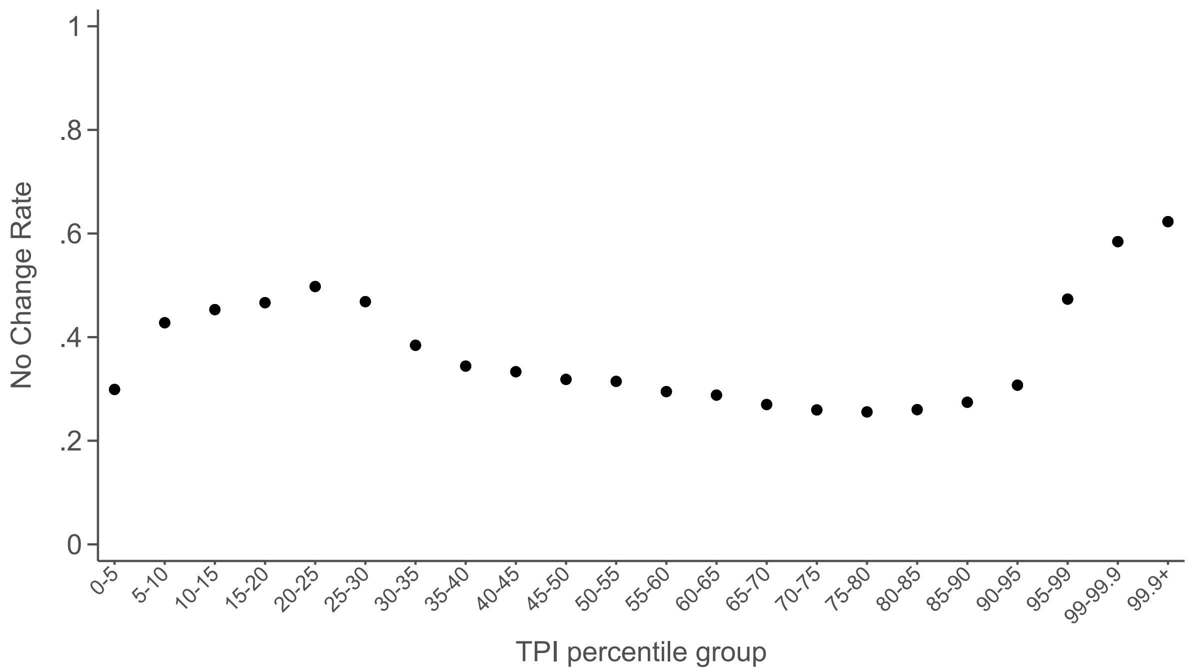
Notes: Panel A presents estimated deterrence effects by the taxpayer's Total Positive Income (TPI) decile. Deterrence effects are the net present value (NPV) of total additional taxes paid post-audit divided by the NPV of upfront revenue raised per National Research Program (NRP) random audit. Additional taxes paid are estimated using a matched differences-in-differences specification, which compares taxes paid for each year post-audit by individuals selected for random audit relative to a matched sample of individuals not selected for random audit. This specification is run separately by TPI decile. Collected tax revenue is winsorized at the 99th percentile of the population distribution to limit the influence of outliers.

APPENDIX FIGURE A16: Taxpayer Burden of Audits, by Income



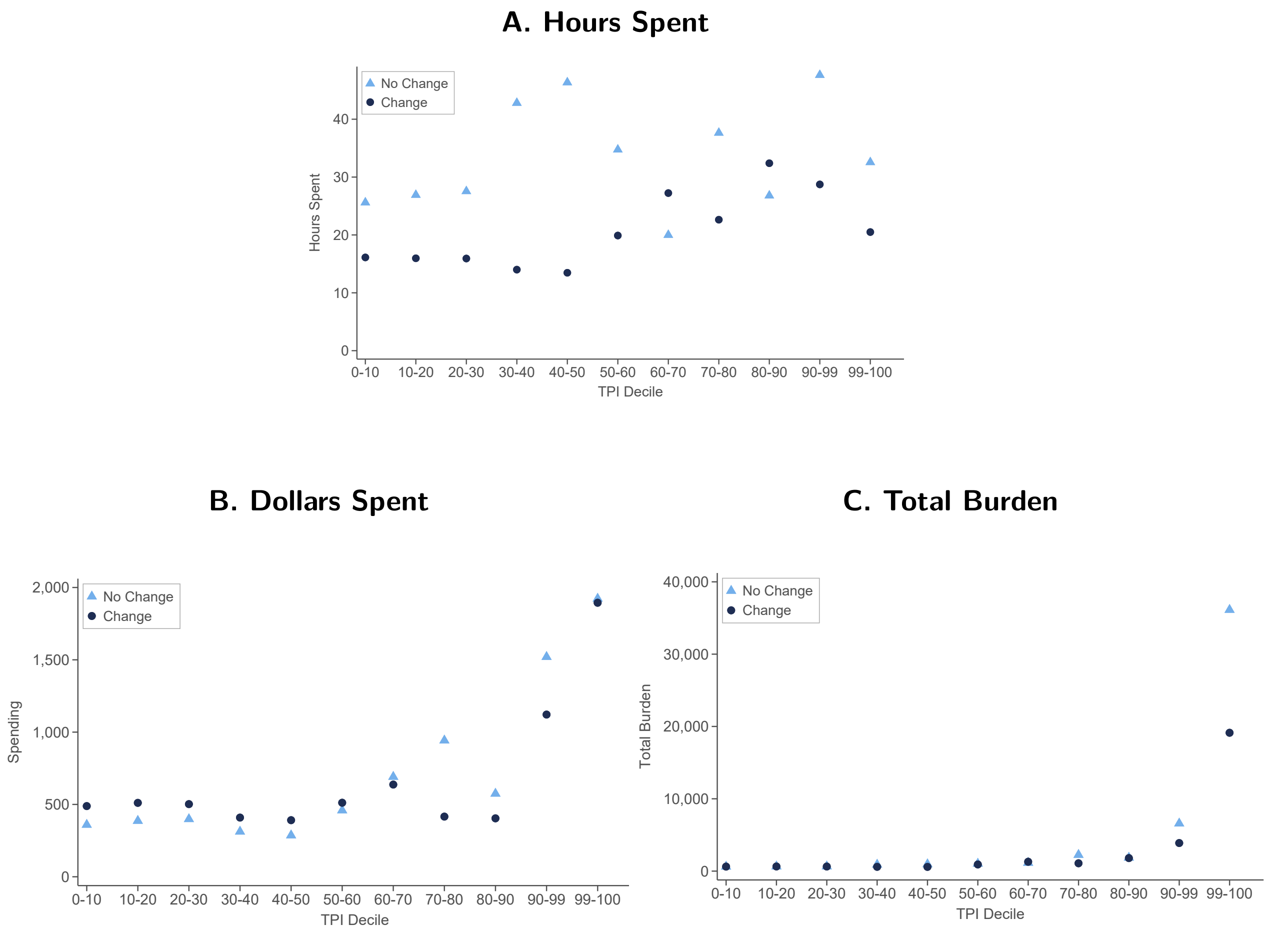
Notes: This graph presents the taxpayer burden of audits using a 2023 representative survey of audited taxpayers conducted by the IRS. Taxpayers are asked about the time and money they spent to comply with the audit. Responses are matched to taxpayer TPI using coarse bins corresponding closely to TPI percentile thresholds. The figures report the average hours spent (Panel A) and dollars spent (Panel B) for each audit. In panel C, the total monetized burden imputes an hourly wage for the taxpayer by dividing their total income by roughly 2000.

APPENDIX FIGURE A17: Fraction of Audits with No Additional Assessed Tax Liability, by Income



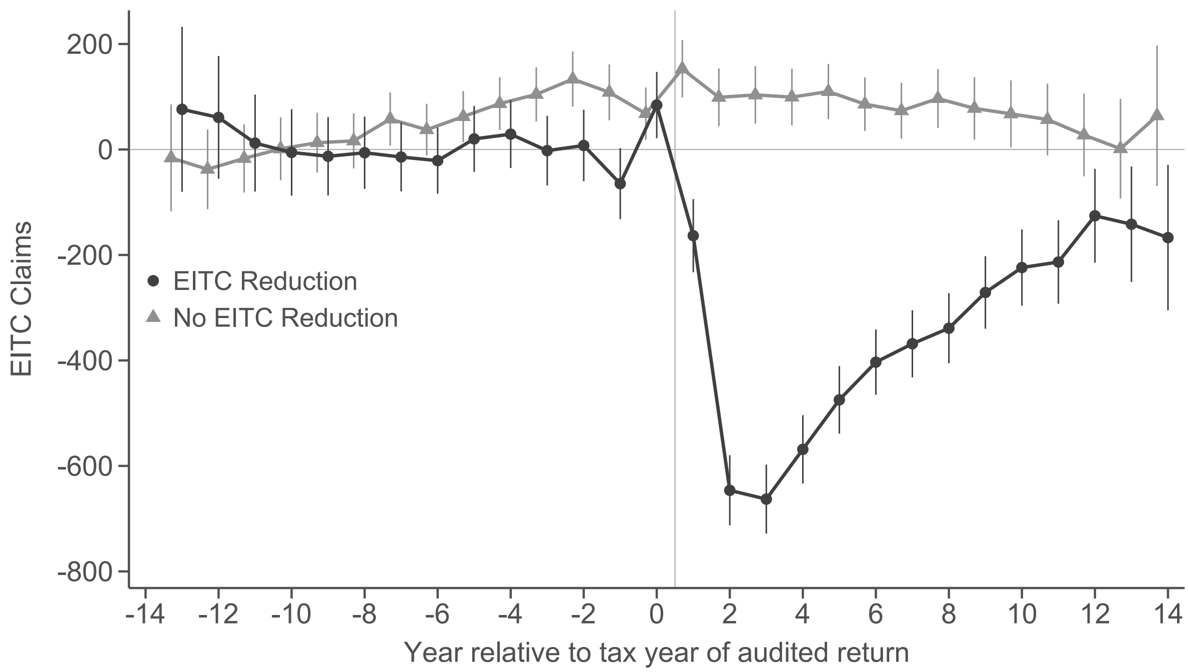
Notes: This graph presents the fraction of in-person audits with no additional assessed tax liability, separately by TPI bin.

APPENDIX FIGURE A18: Taxpayer Burden of Audits, by Income and Whether Additional Tax was Assessed



Notes: This figure repeats Appendix Figure 16 but reports the estimates separately by whether the audit resulted in additional assessed tax liability.

APPENDIX FIGURE A19: Response of EITC Amount Claimed to Random Audits, by Whether the Auditor Recommended Reducing the EITC Amount



Notes: This figure presents differences in average EITC amounts claimed between taxpayers subject to random audit by the NRP and matched controls. It restricts to randomly audited taxpayers who claim the EITC in the tax year selected for random audit and their matched controls, who also claim the EITC and have similar incomes in that tax year. The sample is further split by whether the randomly audited taxpayer's auditor recommended reducing the EITC amount claimed on the return. Each subgroup is then compared to their income-and-EITC-claim-status-matched controls. The difference for those with an auditor-recommended reduction in EITC amount is shown with circles and for those without an auditor-recommended reduction in EITC amount in triangles.

Online Appendix

A Data Appendix

This appendix discusses the data used in our analysis and the methods used to compute the costs and revenues associated with each audit. As noted in the main text, we use two types of internal IRS data as well as data from a survey on taxpayer burdens. In what follows, we detail how we handle each type of data to estimate revenues and costs.

A.1 Audit-Level Enforcement Data: Revenue and Direct Labor Cost Estimates

We use administrative, audit-level data from the IRS's internal enforcement database. The transaction-level dataset records all activities associated with a given audit (e.g. time spent by the auditor, payments collected from the taxpayer, etc). These data allow us to calculate revenues collected for each audit and direct labor costs expended on each audit.

We begin with all rows in the enforcement database where the taxpayer is an individual. [80] This includes both operational exams (in-person and correspondence) as well as random NRP audits.

Identifying a single audit. We combine examinations conducted at the same time of multiple tax returns filed by the same taxpayer into a single audit to capture scenarios where an auditor makes similar adjustments to returns for multiple years. We join examinations that start between a prior examination's start date and 90 days after the prior examination's end date. For example, if a taxpayer's 2016 return is examined from January 2018 to January 2019 and an examination begins of the same taxpayer's 2015 or 2017 tax return between January 2018 and April 2019, we sum the revenue and costs associated with that return as part of the revenues and costs of the examination of the 2016 return. By this definition, 81.9% audits are associated with one tax return, 12.5% are associated with 2 tax returns, and 5.6% are associated with 3 or more tax returns.

Date variables. We estimate the date an audit started using the transactions associated with the audit. In addition, we use five variables included in the enforcement data to estimate four year variables that help us track each audit over time, adjust for inflation and the timing of costs and revenues, etc.. We define the audit start year as the year in which transactions associated with an audit began. We define the primary year to be the tax year of the return that triggered the

[80] More than 99% of these are audits of individual income tax returns, with the remainder including income tax returns of trusts and gift or estate tax returns.

enforcement process by our definition above.[81] For example, suppose that an audit of a tax year 2014 return (which would have been filed in calendar year 2015) began in calendar year 2017. The audit start year would be 2017 and the primary year would be 2014. The labor year is the fiscal year for which the hours are recorded.[82] We use the labor year to determine the auditor's wage rate. Finally, the revenue year is the fiscal year in which payments are recorded in the transaction data.

Identifying the stage of the audit. We assign revenue and costs by the stage of the audit (i.e. exam, appeals, collections). To do so, we use the function code associated with each transaction to classify each transaction into these stages.

Assigning TPI percentiles. We assign individuals to the income distribution using the taxpayer's total positive income (TPI) in the tax year that triggered the audit (i.e., the "Primary Year" and by-year, population-level percentiles of reported TPI. For example, if the primary year of an audit was 2012, we determine that audit's TPI percentile using population-level TPI percentiles from 2012.

Estimating revenue. To estimate the revenue from a given audit, we add total enforcement revenue from taxes, penalties, and interest for the following stages:

1. Exam
2. Appeals and Counsel
3. Collections: notices (1st, 2nd, 3rd, and 4th notices)
4. Collections: ACS
5. Collections: any revenue collected while in the queue for field collections
6. Collections: field collections

Total revenue from collections is the sum of items 3 through 6 above, and total revenue from an audit is the sum of revenue from the exam, appeals, and collections stages. Before summing, all revenue variables are adjusted for inflation using the CPI-U-RS to 2016 dollars.

[81] For any audit identified as a random NRP audit, we use the NRP study year for the "primary year". This changes the value of primary year in 8,000 transaction-level rows (out of 295,000 identified by the NRP data).

[82] When available, labor year is set equal to the fiscal year of the assessment. If assessment fiscal year is missing, we use the year after the exam start year. When both assessment fiscal year and exam start year are populated, assessment fiscal year is one greater than exam start year in 68.8% of rows (and equal in 22.6% of rows). Exam start year may differ from assessment fiscal year in situations where there was a delay between the beginning of the examination and when the work took place. This leaves very few cases with a missing value for labor year. In those few cases, there are no associated hours.

Estimating direct labor costs (exam and appeals stages). Each row of data notes the hours spent by the IRS employee as well as their General Scale (GS) pay grade. We estimate direct labor costs for the exam and appeals stages by multiplying the hours in a given transaction by the matched hourly GS pay rate (described below). We note that hours and costs associated with legal counsel are listed as a separate category of transactions. Because these expenses are generally incurred during the appeals phase, we add the direct costs from any "counsel" hours to appeals to get a total direct labor cost estimate for appeals and counsel. We then use the year in which the activities and expenses were incurred (i.e. "labor year" defined above) to adjust these costs for inflation.

Estimating direct labor costs (field collections). While hours and GS grade are stored for cases that are sent directly to collections, the hours and GS grade information for collections personnel is not stored in the enforcement database for cases that originated in the exam stage.

To estimate the direct labor cost of field collections, we use transactions related to cases that went straight to collections (and therefore have associated hours and GS grade information) from the enforcement database. We estimate direct labor costs for these cases in the same way as exam and appeals cases in our main dataset. We create a 10-by-10 index, where one axis plots deciles of total positive income (TPI; the lowest decile restricted to zero) and the other axis plotting deciles of amount assessed. Each cell contains the average cost estimate associated with that combination of TPI-decile and amount assessed-decile. For each audit in our primary dataset, we determine the relevant decile along both dimensions and apply the corresponding cost estimate.

General Schedule (GS) pay rates. We determine the relevant hourly GS pay rate using the year and location in which the labor activities took place. We proceed as follows. First, we determine the relevant zip code for each individual in our dataset. We determine zip code for a given tax year first by using the modal zip code for an individual's third-party information reporting in that tax year. If there is no zip code found that way, we use the zip code from the individual's 1040. If there is still no zip code, we use the modal third-party information reported zip code from the previous tax year, then the previous tax year's 1040, going back 5 tax years. If there is still no zip code, we apply the average location adjustment from the matched zip codes (matched as described below).

We create a mapping of zip code to FIPS codes, and then FIPS codes to the localities provided in the historical GS pay scale data. After creating this mapping, we merge location-specific hourly pay rates to the transactions the enforcement data by year, location, and GS grade. We use the GS pay scale hourly rate for Step 5. If we had a matched zip code in the enforcement data but no associated GS pay rate for that zip code, we applied the Rest of US (RUS) rate for that GS grade and year. [83]

We then take an average of the difference between the assigned payrates and the base GS payrates to determine our average location adjustment.[84] We use this average location adjustment for cases where there was no matched zip code.

A.2 Accounting Data: Non-Direct Labor Cost Estimates

Our second data source is internal line-item accounting data from the IRS. These data enable us to include all potential costs associated with audits beyond the direct labor hours spent by the auditors.

As background, the IRS has four large operating divisions that deal with taxpayers: (1) Small Business/Self-Employed (SB/SE), (2) Large Business and International (LB&I), (3) Wage and Investment (W&I), and (4) Tax Exempt and Goverment Entities (TE/GE). These operating divisions are responsible for different populations of taxpayers (Internal Revenue Manual 1.1.1).

SB/SE is the operating division responsible for audits of individual tax returns and therefore the operating division relevant for this project.[85] There are three organizations within SB/SE: collection, examination, and operations support (Internal Revenue Manual 1.4.40.2). The examination organization is responsible for both field examination and correspondence examination (Internal Revenue Manual 1.1.16.5).

The internal accounting data we use is organized by enforcement function (e.g., field exam versus correspondence exam) rather than by operating division (e.g., SB/SE versus LB&I). We use these data to construct cost measures for five different enforcement functions: Field Exam (i.e. in-person exams), Correspondence Exam, Field Collections, Collections Notices, and the Automated Collection System (ACS). We then use these numbers to calculate the average cost of each audit by assuming that total costs associated with the audit are proportional to the labor costs associated with the audit. Where possible, we construct our multipliers using SB/SE-specific cost information.

In-person exams. To calculate the cost multipliers for in-person exams, we use internal IRS

[83] The historical GS pay scale data is available at `https://www.opm.gov/policy-data-oversight/pay-leave/salaries-wages/`.

[84] This average is weighted by number of hours.

[85] SB/SE audits both individuals and small businesses such as partnerships, S-Corporations, and C-Corporations with assets under $10 million. The same revenue agents conduct audits of complex individual returns and business returns. C-Corporations with assets greater than or equal to $10 million are handled by LB&I. W&I conducts pre-refund examinations of EITC returns, which are not included in our data. For more information on the SB/SE operating division, visit `https://www.irs.gov/about-irs/small-business-self-employed-division-at-a-glance` .

accounting data for the Field Exam organization from 2011–2020. These data include line-item level cost information for all Field Exam operations as well as total costs for Field Exams conducted by SB/SE. We create our in-person exam multipliers by calculating the appropriate ratios of these line-items to total costs using the information available for all Field Exam costs and applying those ratios to the total Field Exam costs for SB/SE. We describe this process in detail below.

First, we determine the total costs to the government associated with in-person exams. This is similar to the total costs given in the internal accounting data with two slight differences. First, the total costs given in the Field Exam accounting data exclude "imputed" costs for support services provided by other government agencies and not directly part of the IRS's budget. For example, the Department of Agriculture runs payroll for several government agencies, and the Bureau of the Fiscal Services processes payments for the IRS. While these costs are not relevant for internal IRS budgeting purposes, they contribute to the total fiscal cost of audits. We observe imputed costs and include them in the total costs of in-person exams. Second, the internal accounting data include the cost of appeals in their total costs for in-person exams. We separate exams and appeals in the labor input process and construct separate measures of the direct labor costs of appeals. We exclude these costs from our calculations for in-person exam and use them to construct a separate cost multiplier for the appeals process. We next determine the total costs associated with audits for the SB/SE operating division. We do not observe imputed costs and appeals by operating division, so we assume that the fraction of the budget that consists of imputed costs is the same for each operating division. In practice, this means we can apply the ratios from the overall costs for Field Exams to the total costs from SB/SE. Table A1 shows the total costs to the government for SB/SE field exams.

After determining the total costs associated with audits for the SB/SE operating division, we split these costs into three broad buckets: (1) non-direct labor-related costs, (2) organization-wide costs, and (3) other overhead. We use the line-item level data for all field exams to assign costs to categories. Non-direct labor-related costs include time spent by auditors on tasks other than examining returns (e.g. training), fringe benefits, management labor costs, training costs, and other primary labor-related costs. Organization-wide costs include space/rent and information technology costs incurred by auditors. Finally, overhead costs are allocations of accounting costs from the central IRS management that overseas the audit programs.[86] We only observe these line-item costs totaled across all of field exam, including LB&I and TE/GE field exam in addition to

[86]Table A2 lists the specific line-items included in each of these broad categories.

Appendix Table A1: Total Costs for Field Exam ($million)

Year	Total Field Exam Costs	Imputed %	Appeals %	Total Gov. Costs	SBSE Total Gov. Costs	SBSE Direct Labor Costs
2011	3,224	8.34%	3.44%	3,382	1,722	314
2012	3,201	8.35%	3.45%	3,358	1,645	325
2013	3,158	9.39%	3.33%	3,349	1,689	319
2014	3,091	8.97%	3.22%	3,268	1,652	284
2015	2,955	8.14%	0.29%	3,188	1,547	294
2016	2,787	9.78%	0.40%	3,048	1,635	265
2017	2,697	8.86%	0.27%	2,928	1,549	256
2018	2,650	11.91%	0.25%	2,959	1,558	234
2019	2,568	11.85%	0.23%	2,867	1,523	200
2020	2,475	10.74%	0.19%	2,736	1,363	132

SB/SE field exam. We again assume that these line-items are a constant fraction of the total budget for each operating division within field exam. In Table A3, we calculate what percentage of total relevant costs comes from each of these three categories and apply these percentages to the total relevant costs estimated for SB/SE to estimate the value of each of these categories of costs for SB/SE. This gives us the costs for SB/SE that will be used in the numerator of our cost multipliers for field exam.

The last piece of information we need to calculate our multipliers is an estimate of total direct labor costs from SB/SE Field Exam. This enables us to calculate the ratio of the different types of costs discussed in Tables A2 and A3 to the direct labor costs we estimate using the administrative enforcement data.

To calculate total direct labor costs for SB/SE, we proceed analogously to our measure of direct labor costs for individual audits measured above, but we now pull all transactions from the administrative enforcement data associated with in-person exams and with the SB/SE operating division (i.e. including businesses). We calculate total direct labor costs as described above for our audit-level dataset. For transactions related to businesses instead of individuals, we apply an average location adjustment rather than a zip-code specific hourly pay rate.[87] The total direct labor costs from SB/SE in-person exams is given in the last column of Table A1.

We calculate three multipliers for in-person exams using the three broad cost categories described above: (1) non-direct labor-related costs per-dollar of direct labor costs,[88] (2) organization-wide costs per dollar of direct labor costs, and (3) other overhead costs per dollar of direct labor costs (see Table A4). We sum these three values for an overall multiplier.

Because we focus on audits of returns for tax years 2010–2014 (which are filed in early 2011–2015), we use the average of the 2011–2015 fiscal year values of these multipliers. The overall value is 4.39. That is, for every dollar of direct labor costs spent on in-person exams, we include $4.39 of non-direct labor costs, organization-wide costs, and other overhead costs. Non-direct labor costs account for nearly half of the additional costs.

[87] To check our direct labor cost estimates for all of SB/SE field exam, we compared the underlying counts of hours with headcount estimates of the total number of active in-the-field Revenue Agents and Tax Compliance Officers for 2016–2018. The estimates imply that active field examiners average about 70% of a 40-hour work week directly on exams. This value is consistent with conversations we had with individuals who work in the SB/SE field exam unit. We do not include 2019 and 2020 in this exercise. There was an influx of new examiners in 2019. As a result, averages are likely depressed by training time. The pandemic led IRS to pause operations in 2020; any hours estimates would not be reflective of usual IRS operations.

[88] When calculating the multiplier for non-direct labor-related costs, we subtract our direct labor costs estimate form the numerator.

Appendix Table A2: Categories for Field Exam Costs

Non-Direct Labor-Related Costs	Organization-Wide Costs	Other Overhead
Labor	Rent /Building	Internal BU
Benefits	IT	Imputed Cost
Services/Supplies	Printing/Postage	
Traveling	TAS	
Training	TE/GE	
Enforcement	Depreciation	
Printing	Appeals	
Moving Expense	Other Finance	
ADP Operations	Workers Comp	
Space & Housing	UCFE	
Rent	HCO	
Equip.-Non-ADP	WISK Other	
Postage	Corporate S&F	
Communications	Comm & Liaison	
Misc. Revenue	SB/SE	
Misc. Expense	LB&I	
	Corr. Exam Support	

Appendix Table A3: Cost Breakdown for Field Exam

Year	Non-Direct Labor Costs	Organization-Wide Costs	General Overhead Costs
2011	58.39%	21.80%	19.80%
2012	56.63%	22.14%	21.22%
2013	52.74%	18.24%	29.02%
2014	52.66%	19.00%	28.34%
2015	51.52%	13.10%	35.39%
2016	51.39%	12.64%	35.96%
2017	52.90%	11.47%	35.63%
2018	50.32%	11.20%	38.48%
2019	50.78%	11.63%	37.59%
2020	53.70%	12.01%	34.29%

Appendix Table A4: Cost Multipliers per Direct Labor Dollar for In-Person Exam

Year	Non-Direct Labor Cost Multiplier	Organization-Wide Cost Multiplier	General Overhead Cost Multiplier	Overall Cost Multiplier
2011	2.21	1.20	1.09	4.49
2012	1.87	1.12	1.07	4.06
2013	1.79	0.96	1.54	4.29
2014	2.06	1.10	1.65	4.81
2015	1.71	0.69	1.86	4.27
2016	2.17	0.78	2.22	5.16
2017	2.20	0.69	2.15	5.05
2018	2.35	0.74	2.56	5.65
2019	2.88	0.89	2.87	6.64
2020	4.54	1.24	3.54	9.32
2011-2015 Average	**1.93**	**1.02**	**1.44**	**4.39**

Correspondence exams. To calculate the costs for correspondence exams, we use internal IRS accounting data for the Correspondence Exam organization from 2011–2020. We obtain data on costs for all correspondence exams within SB/SE, not just those of individuals. We create our correspondence exam multipliers by calculating the appropriate ratios with the detailed data for all correspondence exams and applying those ratios to the total costs for SB/SE. We describe this process in detail below.

First, we determine the total costs of conducting correspondence audits by adding in the imputed costs that are incurred by other government agencies on behalf of Correspondence Exam activities. We next determine the total relevant costs for correspondence exam for the SB/SE operating division. Because imputed costs by operating division are not included in the data, we assume that the fraction of costs that are imputed costs is the same across operating divisions within correspondence audits. This means that we can apply the ratios from the overall costs for Correspondence Exams to the total costs from SB/SE. This is shown in Table A5.

Table A8 also shows how we apply the relevant ratios to the total costs for SB/SE to estimate the division of costs between our three categories.[89]

[89]Before we break these relevant total costs down into our three cost subcategories, we need to subtract the

Appendix Table A5: Total Costs for Correspondence Exam ($million)

Year	Total Corr. Exam Costs	Imputed %	Total Gov. Costs	SBSE Total Gov. Costs	SBSE Direct Labor Costs
2011	430.0	12.88%	485.4	248.1	16.7
2012	476.9	12.42%	536.1	252.7	16.5
2013	444.8	15.90%	515.5	249.5	16.0
2014	463.0	14.68%	531.0	257.3	12.6
2015	392.7	15.18%	452.3	224.2	11.8
2016	377.8	16.16%	438.9	223.3	11.5
2017	357.8	14.87%	411.0	171.8	10.0
2018	350.8	19.86%	420.5	189.7	9.7
2019	347.3	19.83%	416.2	179.1	8.6
2020	362.3	18.40%	428.9	182.9	5.7

Appendix Table A6: Cost Breakdown for Correspondence Exam

Year	Non-Direct Labor Costs	Organization-Wide Costs	General Overhead Costs
2011	49.34%	32.35%	18.32%
2012	44.59%	29.96%	25.45%
2013	43.75%	12.56%	43.69%
2014	41.01%	11.64%	47.36%
2015	44.37%	17.84%	37.78%
2016	42.36%	15.32%	42.32%
2017	44.35%	17.10%	38.54%
2018	41.47%	15.47%	43.05%
2019	42.45%	16.49%	41.06%
2020	42.55%	16.68%	40.76%

We calculate three multipliers for correspondence exams using the three broad cost categories described above: (1) non-direct labor-related costs per-dollar of direct labor costs,[90] (2) organization-wide costs per dollar of direct labor costs, and (3) other overhead costs per dollar of direct labor costs (see Table A7). We sum these three values for an overall multiplier.

We use the average of the 2011–2015 values of these multipliers. The overall value is 16.07. That is, for every dollar of direct labor costs spent on in-person exams, we include $16.07 of non-direct labor costs, organization-wide costs, and other overhead costs. For correspondence exams, non-direct labor costs account for about 40% of these additional costs.

Appeals. Both field and correspondence exams recommendations can be appealed, which sends the case to the IRS's Independent Office of Appeals. To calculate the cost of the appeals stage of audits, we take the line-item costs for Appeals from the Field Exam data and divide by direct labor costs for appeals and counsel from the administrative enforcement data.[91] Neither the internal

component of the Correspondence Exam efforts that were conducted in support of field exam audits. These costs were included in the costs of Field Exam, because they are costs incurred as a result of the Field Exam program, and therefore should be excluded from the total costs for Correspondence Exam. We assume these costs in support of field exam efforts are incurred proportionally across the line items and subtract them from our cost categories accordingly.

[90] When calculating the multiplier for non-direct labor-related costs, we subtract our direct labor costs estimate form the numerator.

[91] There is not a corresponding Appeals line-item from the Correspondence exam data, perhaps because appeals of Correspondence exams are even rarer than appeals of in-person exams.

Appendix Table A7: Cost Multipliers per Direct Labor Dollar for Correspondence Exam

Year	Non-Direct Labor Cost Multiplier	Organization-Wide Cost Multiplier	General Overhead Cost Multiplier	Overall Cost Multiplier
2011	6.34	4.81	2.73	13.88
2012	5.85	4.60	3.91	14.36
2013	5.82	1.96	6.81	14.60
2014	7.40	2.38	9.70	19.49
2015	7.44	3.39	7.18	18.01
2016	7.23	2.98	8.22	18.43
2017	6.62	2.94	6.62	16.17
2018	7.08	3.02	8.39	18.49
2019	7.82	3.43	8.53	19.78
2020	12.61	5.34	13.04	30.98
2011-2015 Average	**6.57**	**3.43**	**6.07**	**16.07**

accounting data nor the administrative enforcement data splits appeals costs by operating division. Consequently, we use the internal accounting data for all field exams and all transactions associated with appeals and counsel from the administrative enforcement data to estimate the direct labor costs. We use the average from 2011–2014, which is 0.57.[92]

Collections. Not everyone pays what they owe, even after the amount owed is not in dispute. The IRS collections process begins with notification letters to the taxpayer indicating that they have an unpaid balance due. If the taxpayer does not respond to the notifications, the case will be handled by the Automated Collection System (ACS) or by a local field office (Field Collections). If the case is sent to ACS, ACS personnel will try to contact the taxpayer by correspondence and by phone to work with the taxpayer to find a payment solution. If ACS is unsuccessful at resolving the unpaid balance, the case is sent directly to a local IRS field office in which a Revenue Officer will work with directly with the taxpayer to attempt to resolve the balance due.

To calculate the costs of these functions, we use data from 2016 (the earliest year for which information on Collections is available) from the internal IRS accounting data for Notices, ACS, and Field Collections.

Per-collections notice. The cost multiplier we use for notices is a cost-per-notice multiplier. These values are highlighted in Table A8. The average cost is $10.97 for sending a first notice, $9.13 for a second notice, and $17.70 for a "final" notice.

We identify someone as having received a notice (for notices 1–4) if they have positive revenue associated with their audit from any of the parts of the collections process as marked in Table A9. We apply the average per-notice rate for "final" notices to the third and fourth notices.

[92]We do not use the values from 2015 here due to an apparent change in the structure of the appeals line item reported in the internal IRS accounting data in 2015. The drop is extensive enough that our estimated direct labor costs from the administrative enforcement data (which do not experience the same drop) are an order of magnitude bigger than the costs listed in the internal IRS accounting data.

Appendix Table A8: Average Cost per Notice ($)

Year	Average Cost per Notice		
	1st Notice	2nd Notice	Final Notice
2016	10.97	9.13	17.7
2017	13.22	10.25	16.32
2018	15.68	12.5	17.97
2019	12.7	12.6	16.3
2020	15.61	16.4	20.69

Appendix Table A9: Per-Notice Average Rate Applied if Positive Revenue Found from These Stages

If positive revenue from:	Notice 1 Average Rate Applied	Notice 2 Average Rate Applied	Notice 3 Average Rate Applied	Notice 4 Average Rate Applied
Notice 1	X			
Notice 2	X	X		
Notice 3	X	X	X	
Notice 4	X	X	X	X
ACS	X	X	X	X
Collections queue	X	X	X	X
Field collections	X	X	X	X

ACS. We estimate the cost multiplier from an audit going through ACS as a "cost-per-dollar raised." We use the 2016 value of cost-per-dollar raised ($0.0513), as shown in Table A10.

Field collections. We apply our estimated cost multiplier from in-person exams to estimate the total costs associated with field collections. This is because we are not able to separately calculate direct labor costs that result from time in field collections (as opposed to direct labor costs associated with the entire collections process). We assume that, on average, the ratio of non-direct labor costs

Appendix Table A10: ACS Multiplier (Cost per Dollar of Revenue)

Year	Cost per Revenue Dollar
2016	$\boxed{0.0513}$
2017	0.0551
2018	0.0567
2019	0.0529
2020	0.0518

to direct labor costs is the same between in-person exams and in-person collections. This assumption is simpler and, in our view, more plausible than the various assumptions needed for our best attempt to calculate a separate, field collections-specific cost multiplier (which yielded a cost multiplier of 2.19).

B Correspondence Audits

In this section, we present results for the average returns to correspondence audits. Correspondence audits are primarily conducted by mail and are not assigned to a particular auditor. Correspondence audits are generally less complex than in-person audits and rely more heavily on algorithms to identify noncompliance. Appendix Figure A2 presents the cost per audit and revenue raised in each stage of the audit process, using the same format as Figure 1 for in-person audits. Labor costs are a much smaller fraction of the total costs of a correspondence audit than of an in-person audit. However, the overall ratio of revenue to cost for correspondence audits after accounting for non-labor costs is 2.13, similar to the ratio for in-person audits.

Appendix Figure A3 shows how these patterns vary across the income distribution. Panel A shows correspondence audit rates across the income distribution. The high rates of audits in the 20–30th income percentiles are EITC correspondence audits that typically ask for verification of relationship and residency status (e.g. forms such as birth certificates, school paperwork showing home address, etc.). Panel B shows how the hours spent and wages of auditors vary across the income distribution. The hours spent per audit are significantly lower for correspondence audits than for in-person audits. As with in-person audits, correspondence audits of higher-income taxpayers require higher-paid auditors and more of auditors' time. Panel C shows how revenues and costs per correspondence audit vary across the income distribution. Total costs per audit are broadly increasing in taxpayer income; but as with in-person audits, revenues increase faster than costs. Panel D shows the ratio of revenue to total costs per audit. This ratio rises from below 1 for audits at the bottom of the income distribution to 3.4 for audits of individuals in the 99–99.9th percentile of the TPI distribution. It then increases to 11.7 in the top 0.1% of the income distribution.

To summarize, non-labor costs comprise a much larger share of the costs of correspondence audits than of in-person audits, but the average return to correspondence audits is similar to the return to in-person audits, and returns to both types of audits increase with taxpayer income.

C Relation to Existing Estimates

In this appendix, we discuss the relationship of our estimates to two key estimates from existing literature on the returns to audits: Holtzblatt and McGuire (2020) and Sarin and Summers (2019).

Holtzblatt and McGuire (2020) (referred to as HM2020 hereafter) estimate the revenues and costs associated with IRS operational audits in the US. They estimate that, for in-person audits, revenue collected divided by costs was 3.3 and 2.8 in 2010 and 2017, respectively. It is not possible to conduct an exact comparison between that paper and this one because of differences between the two papers' samples. For example, HM2020 include not only individual audits, but also audits of corporations. They examine audits from tax years 2010 and 2017 that were completed before March 31st, 2012 and 2019. They exclude returns with EITC and also remove any "outlier" returns, defined as those in the top 0.5% of taxes collected.

One key difference that leads to higher average returns in HM2020 as compared to our work is that the cost estimates in HM2020 only include the direct costs of hours spent on enforcement activity and do not include additional labor and non-labor costs. In particular, HM2020 calculate the cost of direct enforcement activity by multiplying hours spent auditing by a wage rate. They then incorporate an estimate of additional employee benefit costs on top of those hourly wage costs. As they note, this does not include labor costs for management and support staff. It appears to omit the labor costs of time spent by enforcement personnel on non-audit tasks and does not include non-labor costs such as rent and IT costs. In our cost estimates, which are based on internal IRS accounting information, we find that non-labor costs contribute a substantial fraction of total costs. We estimate that, on average, total costs are 4.39 times larger than the direct labor cost of enforcement activities, and we estimate that more than half of that 4.39 figure is the result of non-labor costs.[93] The inclusion of these non-labor costs in our average cost estimates contributes to the divergence between our estimates and HM2020.[94]

Sarin and Summers (2019) discuss the returns to auditing very high income taxpayers and argue that auditing taxpayers with more than $5 million in earnings can produce a return of 18:1. In particular, they conduct a back-of-the-envelope calculation, drawing upon hourly audit adjustment estimates from George (2019) and dividing by estimates of average auditor costs. While the broad trajectory of our results are consistent with their findings, the approach used in our paper differs from the approach in theirs. First, George's estimates quantify recommended audit adjustments rather than audit revenue collected. A meaningful portion of assessed tax obligations are overturned on appeal and never collected, and so these adjustments often far exceed audit revenue. This means that the observed return on audit expenditures should fall below the ratio of audit assessments to auditor costs. Second, if the calculation in George (2019) is focused on the hours necessary to produce audit adjustments, these calculations may omit the cost of auditor hours following the initial recommendation and appeals and collections activity. While there is no formal confirmation of this hypothesis in George (2019), we find the average hours associated with an audit of the top 0.1% are meaningfully in excess of the average hours reported in that work.[95] Finally, our analysis of very high income taxpayers focuses on those in the top 0.1%. This threshold falls below the $5 million threshold in Sarin and Summers (2019) and so if the pattern of increasing retrurns with income continues to hold we should expect a lower return for the top 0.1% than for those with at

[93] HM2020 argue that non-labor costs are likely to be small because 94% of the IRS enforcement budget is attributable to personnel compensation. That calculation, based on Table 28 on the 2018 IRS Data Book, appears to only include costs associated with the "enforcement" line-item and, therefore, omits the costs associated with the "operational support" line-item. The lion's share of non-labor costs can be found under the operational support line item.

[94] Our marginal cost estimates in Section 4 fall closer to the estimates in HM2020 because, while they still include the labor costs associated with support staff and non-auditing hours, they do not include the portion of non-labor costs we estimate to be fixed.

[95] Hours per audit differ because of different definitions of audit, but the total revenue estimates in George (2019) resemble the revenue estimates in our work.

least \$5 million of income. Those differences help to explain the discrepancy between 18:1 figure in Sarin and Summers (2019) and the 6:1 average return we find for the top 0.1% when deterrence effects are not included.[96]

Our estimates are comparable to recent budget scores produced by the Congressional Budget Office. For example, the CBO estimated (with no deterrence effects) that \$46 billion in additional audit enforcement focused on high-income taxpayers would raise \$180 billion in revenue, a revenue to cost ratio around 3.9. That 3.9 figure aligns fairly closely with our estimates of the returns to marginal audits of taxpayers of high income taxpayers absent deterrence effects. We find returns of 3.2:1 in the 99–99.9th percentile and 6.2:1 in the top 0.1%.

D MVPF of Tax Evasion

In this Appendix, we provide a class of structural models that motivate our MVPF formula in Equation (1). Our modeling approach builds on the large literature on tax evasion (e.g. Allingham et al. (1972); Keen and Slemrod (2017)), but extends to a dynamic context that allows for audits today to change tax payments and evasion behavior in the future. In order to incorporate dynamics while still keeping the model relatively tractable, we economize on other features of the model. For example, we assume quasilinear utility, we do not allow the probability of audits to depend on past behavior, and we do not allow for strategic interactions between evasion levels and probability of audits. We show that, with these assumptions, we can derive our exact formula for the MVPF in Equation (1). We then discuss how relaxing each of these assumptions leads to potential modifications to the MVPF. Finally, we discuss how one can use the MVPF combined with social welfare weights to compare the desirability of raising revenue through audits to other potential policies.

D.1 Setup

We consider an individual, i, who has a utility function over consumption, c_t, and earnings, y_t, in each period indexed by t. Earnings are taxed at $T(y_t)$ so that in the absence of any evasion consumption would be equal to $y_t - T(y_t)$. However, individuals have the opportunity to evade e_t dollars of their tax liability. This increases consumption by e_t in the event they are not audited. We let a_t denote an indicator for being audited in period t, and $\alpha_t = (a_1, ..., a_{t-1})$ denote the individual's audit history up through period t. We assume that, when audited, the individual must

[96] Coincidentally, when deterrence effects are included, the returns we find at the top of the income distribution approach or exceed 18:1.

repay the evaded amount, e_t, plus a penalty $\phi^{\alpha_t}(e_t)$ that depends not only on the evaded amount but also on the individual's audit history, α_t.

Utility in each period is given by $u_i(c_t, y_t) = c_t - \psi_i(y_t) - \mathbb{I}\{a_t = 1\} B_i$, where $\psi_i(\circ)$ measures individual i's disutility of earning income and B_i measures their disutility of being audited (alternatively, the "taxpayer burden" of being audited). As noted above, we assume for simplicity that evasion has no psychic cost. This means the expected PDV of utility is given by

$$U_i = \mathrm{E}\left[\sum_{t=1}^{\infty} \beta^{t-1}\Big(c_t - \psi_i(y_t) - \mathbb{I}\{a_t = 1\} B_i\Big)\right]$$

where the expectation is taken with respect to the probability that $a_t = 1$ in each period, which we denote $p_t = \Pr\{a_t = 1 | e_t\}$. We assume for simplicity this probability is exogenous to evasion choices and income, and that the probability of future audit does not depend on past audits. As we discuss further below in Section D.4.3, our formulas are unchanged when allowing audit probabilities to depend on evasion levels, $p_t(e_t)$.[97]

In each period, the budget constraint is given by

$$\begin{aligned} c_t &\leq y_t - T(y_t) + e_t \quad \text{if } a_t = 0 \\ c_t &\leq y_t - T(y_t) - \phi^{\alpha_t}(e_t) \quad \text{if } a_t = 1 \end{aligned}$$

We note that, while we model ϕ^{α_t} as the true penalty, it would be straightforward to extend the model to allow for misperceptions of the penalty by interpreting ϕ^{α_t} as the perceived penalty.

Individuals make two choices: earnings and evasion. The additive separability in the model implies that the choice of earnings y_t is independent of audits and/or evasion: The optimal choice of income, y_t^*, satisfies $\psi_i'(y_t^*) = 1 - T'(y_t^*)$ in each period (we suppress the i subscript on y_t^* but note this choice differs across individuals). After plugging the budget constraints into the objective function, we see that the choice of evasion in each period solves:

$$\max_{e_t} (1 - p_t(e_t)) e_t - p_t(e_t) \phi^{\alpha_t}(e_t) \tag{3}$$

Intuitively, individuals maximize the expected money they keep from the government net of penalties. In other words, they minimize the expected taxes they pay inclusive of expected penalties that they pay in the event they are audited. We let e_{t,α_t}^* denote the solution to this maximization program in each period t after realizing audit history, α_t. This is given by:

$$e_{t,\alpha_t}^* = \left(\frac{\partial \phi^{\alpha_t}}{\partial e}\right)^{-1}\left(\frac{1-p_t}{p_t}\right)$$

[97]In Section D.4.3, we also discuss a small modification below that enables the probability of a future audit to be increasing in the presence of past audits.

Note that this equation shows how audits can impact future evasion behavior: if being audited increases the marginal penalty from future evasion (e.g. because it is no longer possible to claim that misreporting was not willful), individuals may choose to reduce their evasion behavior in the future.

We can then plug in the choice of evasion into the utility function to write the indirect expected ex ante utility as:

$$\begin{aligned} V_i\left(\{p_t\},\{\phi^{\alpha_t}(\circ)\},T(\circ)\right) &= \mathrm{E}\left[\sum_{t=1}^{\infty}\beta^{t-1}\left(y_t^*-T\left(y_t^*\right)-\psi_i\left(y_t^*\right)+e_{t,\alpha_t}^*-\mathbb{I}\{a_t=1\}\left(e_{t,\alpha_t}^*+\phi^{\alpha_t}\left(e_{t,\alpha_t}^*\right)+B_i\right)\right)\right] \\ &= \sum_{t=1}^{\infty}\beta^{t-1}\Pr\{\alpha_t\}\left(y_t^*-T\left(y_t^*\right)-\psi_i\left(y_t^*\right)+e_{t,\alpha_t}^*-p_t\left(e_{t,\alpha_t}^*+\phi^{\alpha_t}\left(e_{t,\alpha_t}^*\right)+B_i\right)\right) \end{aligned}$$

where $\Pr\{\alpha_t\}$ is the probability of a particular audit sequence, α_t. The *ex ante* expected utility experienced by the individual, V, is a function of the audit probabilities, p_t, and the penalty functions, $\phi^{\alpha_t}(\circ)$, and the tax schedule, $T(\circ)$.

D.2 Willingness to Pay for Expanded Audits

We now can ask: what is the welfare impact of expanding audits? We model this as an increase in the audit probability in the first period by dp_1. Individuals are willing to pay $\frac{dV}{dp_1}$ in order to avoid an audit. To see how changing p_1 affects V_i, it is helpful to write V_i by expanding out the first period probability of audit. We have:

$$V_i = p_1\left[y_1^*-T\left(y_1^*\right)-\psi\left(y_1^*\right)-\phi^{\alpha_1}\left(e_{1,\alpha_1}^*\right)-B+\beta V_i^1\right]+\left(1-p_1\right)\left[y_1^*-T\left(y_1^*\right)-\psi\left(y_1^*\right)+e_{1,\alpha_1}^*+\beta V_i^0\right]$$

where $\phi^{\alpha_1}(\circ)$ is the penalty in the first period (before there is any audit history for the individual) and e_{1,α_1}^* is the choice of evasion in the first period. The first term in brackets is the utility if audited and the second term is the utility if not audited. The term V_i^1 is the PDV of future utility in subsequent periods if $a_1=1$ and V_i^0 is the PDV of future utility in subsequent periods if $a_1=0$. The envelope theorem implies that the impact of increasing p_1 affects utility through both the first period utility and the impact on future utility:

$$\begin{aligned} -\frac{dV_i}{dp_1} &= e_1^{a_t}+\phi^{\alpha_1}\left(e_{1,\alpha_1}^*\right)+B_i+\left(V_i^0-V_i^1\right) \\ &= R^{mech}+B+\left(V_i^0-V_i^1\right) \end{aligned}$$

The first period utility impact of the additional audits is given by the level of evasion plus the penalty and the taxpayer burden. To calculate the impact of the audit on future periods, note that we can write $V_i^0 - V_i^1$ as the present discounted future revenue collected by the government from reduced evasion:

$$\begin{aligned} V_i^0 - V_i^1 &= \sum_{t=2}^{\infty} \beta^{t-1} \left(E\left[e^*_{t,\alpha_t} - p_t \left(e^*_{t,\alpha_t} + \phi^{\alpha_t}\left(e^*_{t,\alpha_t}\right)\right) | a_1 = 0 \right] - E\left[e^*_{t,\alpha_t} - p_t \left(e^*_{t,\alpha_t} + \phi^{\alpha_t}\left(e^*_{t,\alpha_t}\right)\right) | a_1 = 1 \right] \right) \\ &\equiv R_i^{future} \end{aligned}$$

so that $V_i^0 - V_i^1$ is the causal effect of the audit in period 1 on the PDV of the change in tax revenue paid to the government in the future as a result of being in the audited ($\alpha_1 = 1$) versus non-audited ($\alpha_1 = 0$) state of the world in period 1. Combining, the willingness to pay to avoid an expansion of audits in period 1 is given by:

$$-\frac{dV}{dp_1} = R_i^{mech} + R_i^{future} + B_i \tag{4}$$

which is the sum of the mechanical revenue collected by the audit, the future PDV revenue collected as a result of within-person deterrence from the audit, and the taxpayer burden of the audit.[98]

Equation (4) provides a measurement of a given individual's willingness to pay to avoid the audit. Those with higher financial impacts or non-financial burden from the audit have a higher willingness to pay to avoid the audit. With a slight abuse of notation, we let R^{mech},R^{future},and B without i subscripts denote the average values of these variables among those being audited.

D.3 Government Revenue and the MVPF of Expanded Audits

Let G denote the PDV of government revenue:

$$G = \sum_{t=1}^{\infty} \beta^{t-1} \mathrm{E}\left[T\left(y_t^*\right) - e^*_{t,\alpha_t} + \mathbb{I}\left\{a_t = 1\right\} \left[e^*_{t,\alpha_t} + \phi^{\alpha_t}\left(e^*_{t,\alpha_t}\right) - C \right] - F \right]$$

where C is the marginal cost of an audit and F is the fixed costs of audits. We assume the government and individuals have the same expectations and discount factor.

The effect of expanding audits in period 1 is given by the sum of the revenue collected in the first period, R^{mech} and the revenue collected in future periods R^{future} minus the marginal cost of the audits in period 1:

$$\frac{dG}{dp_1} = R^{mech} + R^{future} - C \tag{5}$$

[98] We assume that the probability of audit p_t is independent of the choice of income that an individual has. This latter assumption can easily be relaxed by assuming that income choices, y_t, are affected by the probability of the audit. The envelope theorem implies that these will not enter the willingness to pay to avoid the expanded audits. They could, however, generate an additional positive or negative revenue to the government from the audit.

Let $R = R^{mech} + R^{future}$ denote the total PDV of government revenue collected as a result of the audit. Combining the willingness to pay by audited individuals to avoid an audit and the revenue raised by audits, the MVPF of individual tax audits can be expressed as:

$$MVPF = \frac{R+B}{R-C}$$

which is precisely our formula in equation (1).

Specific versus General Deterrence This MVPF expression does not contain a term for the deterrence effect of the increased probability, p_t, on evasion in the first period, e^*_{1,α_1}. The envelope theorem implies that the change in evasion in response to the increase in threat of audit in period 1 does not enter the willingness to pay term. For the government cost, note that the net impact on the government budget from changes in e^*_{1,α_1} is given by:

$$-\frac{de^*_{1,\alpha_1}}{dp_t}\left[(1-p_1) - p_1\frac{\partial\phi^{\alpha_1}}{\partial e}\left(e^*_{1,\alpha_1}\right)\right] = 0 \tag{6}$$

Because individuals maximize their expected income from evasion, they also minimize the government revenue from evasion. Recall from the individuals' optimization problem above that $\frac{\partial\phi^{\alpha_t}}{\partial e}\left(e^*_{t,\alpha_t}\right) = \frac{1-p_t}{p_t}$. This means that small changes in evasion behavior in response to changes in the audit probability do not affect government revenue. Expanding audits leads people to evade less ($\frac{de}{dp} < 0$), but the loss in penalties perfectly offsets the gain in tax revenue from reduced evasion. As a result, general deterrence (i.e. $\frac{de^*_{1,\alpha_1}}{dp_t}$) does not enter either the numerator or denominator of the MVPF. We show below that, under alternative assumptions about taxpayer behavior, one can obtain general deterrence effects that change government revenue. In these cases, general deterrence should be included in the net cost (denominator) of the MVPF but the envelope theorem continues to imply that it does not enter the numerator.

At this point, one might be puzzled why the willingness to pay to avoid an audit includes individual deterrence, R^{future}, but not a term related to general deterrence, $-\frac{de^*_{1,\alpha_1}}{dp_t}$. We noted above that the envelope theorem applies to $-\frac{de^*_{1,\alpha_1}}{dp_t}$, so why doesn't the envelope theorem also apply to individual deterrence? The key distinction is that the individual deterrence term includes not only a behavioral response but also the mechanical revenue collected from changes to the future constraints faced by an audited taxpayer. When individuals are audited, their future audit penalties, ϕ^{α_t}, from continuing their evasion increase.[99] The impact on revenue, R^{future}, is then the sum of the mechanical and behavioral components from these increased penalties. To see this, consider

[99]See Section D.4.3 below for a discussion of the case when probabilities of future audit increase.

just the revenue collected in period 2 as a result of expanded audits in period 1. Let R_2^{future} denote this period 2 revenue and let e_2^0 and e_2^1 denote the amount evaded in period $t = 2$ conditional on having been not audited or audited in period 1. The impact of the expanded audit in period 1 on revenue in period 2 is given by:

$$R_2^{future} = e_2^0 - p_2\left(e_2^0 + \phi^0\left(e_2^0\right)\right) - \left(e_2^1 - p_2\left(e_2^1 + \phi^1\left(e_2^1\right)\right)\right) \tag{7}$$

Now, note that we can write the change in penalty revenue as:

$$\phi^0\left(e_2^0\right) - \phi^1\left(e_2^1\right) = \phi^0\left(e_2^0\right) - \phi^1\left(e_2^0\right) + \phi^1\left(e_2^0\right) - \phi^1\left(e_2^1\right)$$

where $\phi^0\left(e_2^0\right) - \phi^1\left(e_2^0\right)$ is the mechanical effect on audit revenue of increased penalties after an audit holding evasion in period 2 fixed and $\phi^1\left(e_2^0\right) - \phi^1\left(e_2^1\right)$ is the impact of the change in evasion on the penalty revenue. Plugging this back into R_2^{future} we obtain

$$R_2^{future} = \underbrace{\phi^0\left(e_2^0\right) - \phi^1\left(e_2^1\right)}_{Mechanical} + \underbrace{\left(e_2^0 - e_2^1\right)\left(1 - p_2 - \frac{\phi^1\left(e_2^0\right) - \phi^1\left(e_2^1\right)}{e_2^0 - e_2^1}\right)}_{Behavioral}.$$

This equation shows that the impact of an additional audit in period 1 on revenue in period 2 is the sum of two components: the mechanical revenue obtained from higher penalties at the level of evasion chosen by someone who had not been audited and the impact of the change in evasion, $e_2^0 - e_2^1$,

In contrast, the impact of general deterrence on government revenue is given by Equation (6) above, which equals zero due to the envelope theorem from individuals maximizing utility when choosing their evasion. Intuitively, the general deterrence term is the behavioral component that comes along with the mechanical revenue component given by the revenue collected by the audits in period 1. The individual deterrence effect is, in contrast, the sum of both the mechancial and behavioral revenue components in future periods.

There are therefore two differences between specific and general deterrence. First, individual deterrence contains a mechanical revenue effect due to the higher penalties faced by taxpayers if they are caught misreporting again. In contrast, general deterrence does not have a mechanical revenue component: it is defined as solely the behavioral response component to the increase in p. Second, even though the change in audits is marginal, an audit leads to a non-marginal change in the future constraints faced by the individual (i.e. strictly higher penalties). The threat of future penalties may lead taxpayers to the corner solution where they report truthfully. Because these increases in penalties can be large, it is important when measuring individual deterrence to include the impact of the behavioral response to the policy on the government budget. In our model, the quasilinearity of preferences means that we can instead measure the welfare cost (mechanical + behavioral components) from individual deterrence as the net impact on future tax revenue.

D.4 Robustness and Extensions

The baseline model above makes a series of modeling assumptions. In this subsection, we discuss a few of these key assumptions and how relaxing them changes the MVPF. We begin by showing how one can use a Baily-Chetty style adjustment to our formulas to account for the fact that individuals may be risk averse when facing audits. Next, we discuss issues related to general deterrence and how various forms of general deterrence can enter the MVPF. Finally, we discuss the role of endogenous audit probabilities.

D.4.1 Risk Aversion

Let $u(c)$ denote utility over realized consumption. In the presence of risk aversion, V_i now becomes:

$$V_i = p_1 \left[u\left(y_1^* - T\left(y_1^*\right) - \psi\left(y_1^*\right) - \phi^{\alpha_1}\left(e^*_{1,\alpha_1}\right) - B\right) + \beta V_i^1\right] + (1-p_1)\left[u\left(y_1^* - T\left(y_1^*\right) - \psi\left(y_1^*\right) + e^*_{1,\alpha_1}\right) + \beta V_i^0\right]$$

So, when we consider the derivative with respect to p_1 we end up with

$$-\frac{dV_i}{dp_1} = u\left(c_{noaudit}\right) - u\left(c_{audit}\right) + \left(V_i^0 - V_i^1\right)$$

where

$$\begin{aligned} c_{audit} &= y_1^* - T\left(y_1^*\right) - \psi\left(y_1^*\right) - \phi^{\alpha_1}\left(e^*_{1,\alpha_1}\right) - B \\ c_{noaudit} &= y_1^* - T\left(y_1^*\right) - \psi\left(y_1^*\right) + e^*_{1,\alpha_1} \end{aligned}$$

Following Baily (1978), we can take a second order Taylor expansion of the utility function around a fixed level of consumption to approximate the difference in the levels of utility with the marginal utility of consumption yielding

$$u(c) \approx u(c_0) + u'\left(c_0\right)\left(c - c_0\right) + \frac{1}{2}u''\left(c_0\right)\left(c - c_0\right)^2$$

where u' and u'' are the values of the first and second derivative of the utility function at a particular point of consumption. We take the point of marginal utility expansion to be the one normalizing willingness to pay on the left-hand side of the equation. This leaves us with

$$\begin{aligned} u\left(c_{audit}\right) - u\left(c_{noaudit}\right) &\approx u'\left(c_{noaudit}\right)\Delta c - \frac{1}{2}u''\left(c_{noaudit}\right)\left[\left(c_{audit} - c_{noaudit}\right)^2\right] \\ &= u'\left(c_{noaudit}\right)\Delta c\left(1 + \frac{1}{2}\sigma\frac{\Delta c}{c}\right) \end{aligned}$$

where $\sigma = \frac{-u''(c_{noaudit})c_{noaudit}}{u'(c_{noaudit})}$ is the coefficient of relative risk aversion evaluated at the level of no audit consumption and $\frac{\Delta c}{c} = \frac{c_{audit} - c_{noaudit}}{c_{audit}}$ is the change in consumption relative to the consumption in the absence of an audit. This equation shows that willingness to pay to avoid the audit is higher than the full monetary difference in consumption between audited and non-audited states of the world. We note that a similar adjustment applies to $V_i^0 - V_i^1$, so that one would also adjust upwards the willingness to pay from individual deterrence.

The advantage of this expression is that it provides guidance on the magnitude of how risk aversion affects the MVPF (recall the net cost to the government is unchanged): the willingness to pay is adjusted upwards by the coefficient of relative risk aversion times the percent change in consumption as a result of the audit. The ideal dataset would track consumption of audited versus non-audited individuals, in the same way such formulas have been applied in the unemployment insurance context. To start, consider the top of the income distribution. Audits of taxpayers in the top 10% of the income distribution collect revenue around \$30,000, and we know that roughly a third of that amount is the persistent increase in reported tax liability. Hence, it would be natural to think of audits as leading to a drop in consumption of \$10,000, which is at most 5% of average income. Combining this with a CRRA of 3 would suggest a willingness to pay that is 15% higher to avoid the audit. This would increase the MVPF from 1.15 to 1.3. Even under these conservative assumptions, the MVPF of expanding audits of top earners is less than the estimated MVPF of 1.5 or more of increasing tax rates on the same group.

We can consider a similar exercise at the bottom of the income distribution. Here, perhaps the upfront revenue paid of \$5,000 translates into a consumption drop of around \$2,000, which would be a roughly 10% drop for someone consuming \$20,000. With a coefficient of relative risk aversion of 3, this would suggest we need to revise upwards the MVPF for audits of low income taxpayers by 30%. While future work should seek to measure the consumption impact of audits across the income distribution, these stylized calculations suggest that including risk aversion would not overturn our core conclusion that the MVPF of audits is declining across the income distribution.

D.4.2 General Deterrence

Our baseline formula does not include general deterrence effects. This is because when individuals choose evasion levels to maximize income after taxes and expected penalties, they also minimize government revenue from taxes and penalties so that the envelope theorem from taxpayer maximization also implies that general deterrence responses from audits have no first order effect on government revenue.

In this section, we discuss and relax the modeling assumptions that generate this result. We show that this result is dependent on the combination of (a) no risk aversion and (b) no "non-financial" costs of evasion. When either of these assumptions does not hold, then changes in evasion in response to the increased threat of audit can have first order fiscal benefits/costs to the government. These fiscal effects should be included in the denominator of the MVPF. However, they continue to be omitted from the numerator of the MVPF due to the envelope theorem. This means that including general deterrence effects is likely to reduce the MVPF of tax audits, which makes expanded audits a more efficient method of raising revenue than is implied by our baseline measures.

We discuss the cases of risk aversion and non-financial costs of evasion in turn below.

Risk Aversion Suppose individuals are risk averse. Let $u'_{1,0}$ be the marginal utility of consumption in period 1 if not audited and $u'_{1,1}$ be the marginal utility of consumption in period 1 if audited. Note that, as in Allingham et al. (1972), we expect $u'_{1,1} > u'_{1,0}$ because if caught the individual must repay the evaded amount plus penalties. Individuals choose evasion in period 1, e_{1,α_1}, to maximize expected utility, leading to the first order condition:

$$-\frac{de^*_{1,\alpha_1}}{dp_t}\left[(1-p_1)\,u'_{1,0} - p_1\frac{\partial\phi^{\alpha_1}}{\partial e}\left(e^*_{1,\alpha_1}\right)u'_{1,1}\right] = 0$$

which can be re-arranged to yield:

$$\left[(1-p_1) - p_1\frac{\partial\phi^{\alpha_1}}{\partial e}\left(e^*_{1,\alpha_1}\right)\right] \quad = \quad p_1\frac{\partial\phi^{\alpha_1}}{\partial e}\left(e^*_{1,\alpha_1}\right)\left(\frac{u'_{1,1}}{u'_{1,0}} - 1\right)$$

With this re-arrangement, the left-hand side is the lost government revenue from additional evasion and the right-hand side is the wedge from risk aversion (which is positive as long as the audit leads to an increase in the marginal utility of consumption). This is similar to Equation (6) but also accounts for higher marginal utility of income in the audited state of the world. As a result, a marginal increase in evasion has a first order effect on government revenue. In the model, this general deterrence effect on government revenue is given by $-\frac{de^*_{1,\alpha_1}}{dp_t}\left[(1-p_1) - p_1\frac{\partial\phi^{\alpha_1}}{\partial e}\left(e^*_{1,\alpha_1}\right)\right]$ or equivalently by $-\frac{de^*_{1,\alpha_1}}{dp_t}p_1\frac{\partial\phi^{\alpha_1}}{\partial e}\left(e^*_{1,\alpha_1}\right)\left(\frac{u'_{1,1}}{u'_{1,0}} - 1\right)$. This additional term would enter as an additional fiscal externality in the denominator of the MVPF. However, the change in evasion in response to the increased probability of audit would not enter the individual's willingness to pay due to the envelope theorem. In this specification, general deterrence effects would lead to a lower MVPF than in our baseline estimate that does not allow for risk aversion, making audits a more welfare efficient method of raising revenue.

Non-Financial Costs Next, suppose that we have our baseline model but now there are non-financial costs $C(e_t)$ of evasion that are increasing in the size of evasion, $C'(e_t) > 0$. These could include either a moral cost to individuals from dishonesty or other non-financial penalties such as jail time. The individual now chooses evasion to maximize:

$$\max_{e_t} (1 - p_t) e_t - p_t \phi^{\alpha_t} (e_t) - C(e_t)$$

which generates a first order condition for the optimal choice e^*_{1,α_1}:

$$-\frac{de^*_{1,\alpha_1}}{dp_t} \left[(1 - p_1) - p_1 \frac{\partial \phi^{\alpha_1}}{\partial e} \left(e^*_{1,\alpha_1} \right) - C' \left(e^*_{1,\alpha_1} \right) \right] = 0$$

which can be re-written as:

$$\left[(1 - p_1) - p_1 \frac{\partial \phi^{\alpha_1}}{\partial e} \left(e^*_{1,\alpha_1} \right) \right] = C' \left(e^*_{1,\alpha_1} \right)$$

Note that the left-hand side is the marginal impact of general deterrence, $-\frac{de^*_{1,\alpha_1}}{dp_t}$, on government revenue, as in equation (6), and the right-hand side is the slope of the non-financial costs with respect to the size of the evasion. As with risk aversion, a reduction in evasion raises government revenue and should be incorporated into the denominator of the MVPF. Here again, however, the general deterrence response, $\frac{de^*_{1,\alpha_1}}{dp_t}$, does not affect the individual's willingness to pay to avoid the audit due to the envelope theorem.

To summarize, under the benchmark case of risk neutrality and purely financial costs from audits, general deterrence effects have no impacts on government revenue. When either of these assumptions is relaxed, general deterrence responses can have impacts on government revenue. Future work should aim to estimate these general deterrence effects and include them in the denominator (but not the numerator) of the MVPF.

D.4.3 Endogenous Audit Probability

Our baseline parameterization assumes that the probability of an audit is exogenous and independent of past audits. Here, we relax both of these assumptions, allowing the probability of an audit to depend on the extent of past audits and the amount evaded. We begin with the case where past audits increase the probability of future audits.

Past Audits Increase Probability of Future Audits Our baseline specification assumes the core reason that people reduce evasion after being audited is that they fear higher penalties in the future. This modeling choice is motivated by the sharply increasing penalties for repeated and willful (as opposed to unintentional) misreporting. Nonetheless, an alternative potential reason that we find significant individual deterrence effects is that people anticipate a higher future audit probability after being audited (e.g. they expect that that are being watched more closely). Here, we assess how our MVPF calculations would differ if the individual deterrence effect is driven by changes in future audit probabilities as opposed to changes in future penalties. The key insight from this specification is that one needs to not only include the additional revenue from individual deterrence but also the greater future audit burdens and higher administrative costs to perform the additional audits that arise due to the higher future audit probabilities. Under the assumption that the impact of audit probabilities on evasion is perfectly linear, the MVPF reduces to the static MVPF excluding individual deterrence. This is because any future revenue gained from changes in evasion (individual deterrence) is offset by reduced revenue from future audits of that individual.

To allow future audit probabilities to depend on past audits, let $p_t(\alpha_t)$ denote the audit probability in period t given audit history α_t. Let $p_t^1(\alpha_t)$ denote the probability of a future audit in period t conditional on an audit in period 1 and $p_t^0(\alpha_t)$ denote the probability of a future audit in period t conditional on no audit in period 1. The willingness to pay to avoid the audit now has a similar but slightly different form (abstracting from individual heterogeneity by dropping i subscripts):

$$-\frac{dV}{dp_1} = R^{mech} + \left(V^0 - V^1\right) + B\left(1 + \sum_{t>1} \beta^{t-1} E\left[p_t^1(\alpha_t) - p_t^0(\alpha_t)\right]\right) \tag{8}$$

conditional where the burden term, B, is now multiplied by the increase in the number of future audits so that we capture the PDV of burdens experienced by an audited taxpayer inclusive of greater future audits. Here, the term $V^0 - V^1$ is again given by the future revenue collected from audits:

$$\begin{aligned} V_i^0 - V_i^1 &= \sum_{t=2}^{\infty} \beta^{t-1} \left(E\left[e_{t,\alpha_t}^* - p_t^0(\alpha_t)\left(e_{t,\alpha_t}^* + \phi^{\alpha_t}\left(e_{t,\alpha_t}^*\right)\right) | a_1 = 0\right] - E\left[e_{t,\alpha_t}^* - p_t^1(\alpha_t)\left(e_{t,\alpha_t}^* + \phi^{\alpha_t}\left(e_{t,\alpha_t}^*\right)\right) | a_1 = 1\right]\right) \\ &\equiv R_i^{future} \end{aligned}$$

which now includes the revenue collected from the higher rate of future audits and higher penalties from future audits. To isolate the role of the probability of future audits, it is helpful to assume that the penalties do not vary with past audits (we note this is inconsistent with IRS practice, but helpful for understanding the general properties of the MVPF of audits). Consider again the future revenue to the government in period 2. Let p_2^0 denote the probability of being audited in period 2 if the individual is not audited in period 1 and p_2^1 the corresponding probability for those audited in period 1. The impact of the period 1 audit on revenue collected in period 2 (as measured in equation (7) above) now becomes

$$R_2^{future} = e_2^0 - p_2^0\left(e_2^0 + \phi\left(e_2^0\right)\right) - \left(e_2^1 - p_2^1\left(e_2^1 + \phi\left(e_2^1\right)\right)\right)$$

where we remove superscripts for ϕ to denote the fact it does not depend on past audits. Rearranging,

$$R_2^{future} = e_2^0 - e_2^1 - \left(p_2^0 - p_2^1\right)\left(e_2^0 + \phi\left(e_2^0\right)\right) + p_2^1\left(e_2^1 + \phi\left(e_2^1\right) - e_2^0 + \phi\left(e_2^0\right)\right)$$

so that the revenue collected in period 2 normalized by the change in audit probabilities is given by

$$\frac{R_2^{future}}{p_2^1 - p_2^0} = e_2^0 + \phi\left(e_2^0\right) + \frac{e_2^1 - e_2^0}{p_2^1 - p_2^0}\left(1 - p_2^1\frac{e_2^1 + \phi\left(e_2^1\right) - e_2^0 + \phi\left(e_2^0\right)}{e_2^1 - e_2^0}\right)$$

The revenue collected in period 2 is the sum of the mechanical revenue, $e_2^0 + \phi\left(e_2^0\right)$, plus the impact of the behavioral response, $e_2^1 - e_2^0$, relative to the increased audit probabilities, $p_2^1 - p_2^0$, on net government revenue. To first order (i.e. $p_2^1 \approx p_2^0$), this latter term is zero because the choice of evasion minimizes government revenue — any gains from evasion, e_2, are weighed against the penalties from evasion $p_2\left(e_2 + \phi\left(e_2\right)\right)$, so that $e_2^1 \approx p_2^1\left(e_2^1 + \phi\left(e_2^1\right)\right)$ and $e_2^0 \approx p_2^1\left(e_2^0 + \phi\left(e_2^0\right)\right)$. To first order , the revenue from future audits is simply the mechanical revenue generated from the greater number of audits in period 2, $e_2^0 + \phi\left(e_2^0\right)$. Any decrease in evasion in period 2, e_2^0 leads to an offsetting reduction in revenue that is collected euring the audit.

If we now make the additional assumption that evasion is constant over time, $e_t^0 = e_1^0$, then $R^{mech} = \frac{R_t^{future}}{p_t^1 - p_t^0}$. Hence,

$$\begin{aligned} R^{future} = V^0 - V^1 &= \sum_{t>1} \beta^{t-1} R_t^{future} \\ &= R^{mech} \sum_{t>1} \beta^{t-1} E\left[p_t^1\left(\alpha_t\right) - p_t^0\left(\alpha_t\right)\right] \end{aligned}$$

which will become useful below when forming the MVPF.

Turning to government costs, an increase in audits today leads to more audits conducted in the future due to the dependence of p_t on past audits, α_t. The total cost is

$$\frac{dG}{dp_1} = R^{mech} + R^{future} - C\left(1 + \sum_{t>1} \beta^{t-1} E\left[p_t^1\left(\alpha_t\right) - p_t^0\left(\alpha_t\right)\right]\right) \tag{9}$$

which includes the additional costs from the greater probability of future audits. Combining the willingness to pay and net costs and using the fact that can divide through by $1 + \sum_{t>1} \beta^{t-1} E\left[p_t^1\left(\alpha_t\right) - p_t^0\left(\alpha_t\right)\right]$

, we arrive at an MVPF formula given by:

$$MVPF = \frac{R^{mech} + B}{R^{mech} - C}$$

This shows that to first order the MVPF only depends on the mechanical revenue generated by the audit. The individual deterrence effect increases future revenue but it is future audits generating that revenue. When evasion is constant over time, this means that the future revenue comes with the same proportional increase in government costs from those future audits. Hence, the key lesson here is that if individual deterrence is driven by the increase in future audit probabilities, then we need to account for the cost of those future audits when also measuring the benefits in terms of collected revenue. In contrast, if individual deterrence is driven by the threat of higher penalties, then the individual deterrence revenue is collected without additional cost to the government (which is related to the classic insight of Becker (1968) and Allingham et al. (1972) that higher penalties are more efficient than expanded audits, a point we return to in Section D.7 below). It is unlikely, however, that in our setting the deterrence effect is driven by increased audit probabilities. For our individual deterrence effects to be driven solely by increased audit probabilities, it would need to be the case that audited taxpayers face on average 3 additional audits in the 14 years after their initial audit; in practice, they face on average less than 1 additional audit. In contrast, in practice taxpayers can face serious financial and criminal penalties for repeated noncompliance.

Evasion Affects Audit Probabilities The previous analyses continued to maintain the assumption that the audit probability is independent of the level of evasion. Here, we relax this assumption and show that this largely does not affect the results. As in the previous case, we need to accurately measure the extent to which an audit today leads to more future audits so that we accurately measure audit burdens and costs; but if individual deterrence effects are driven by the increase in the slope of the audit probability with respect to evasion and do not lead to more future audits, our baseline formula is valid for measuring the MVPF of expanded audits.

To see this, we assume that evasion in period, e_t, increases the chance of being audited in period t. To capture this in a fairly general way, we let the audit probability depend on both the history of past audits and the current level of evasion, $p_t^{\alpha_t}(e_t)$, and assume this is continuously differentiable and increasing in evaded income. Given any audit history α_t, the generalization of equation (3) shows that optimal level of evasion e^*_{t,α_t} is given by the first order condition :

$$1 - p_t^{\alpha_t}\left(e^*_{t,\alpha_t}\right) = \frac{\partial p_t^{\alpha_t}\left(e^*_{t,\alpha_t}\right)}{\partial e_t}\left(e^*_{t,\alpha_t} + \phi^{\alpha_t}\left(e^*_{t,\alpha_t}\right)\right) + p_t^{\alpha_t}\left(e^*_{t,\alpha_t}\right)\frac{\partial \phi^{\alpha_t}}{\partial e}\left(e^*_{t,\alpha_t}\right) \quad (10)$$

Evasion has benefits proportional to the probability of not being caught, $1-p_t^{\alpha_t}\left(e_{t,\alpha_t}^*\right)$, but has costs in terms of higher probability of being caught, $\frac{\partial p_t^{\alpha_t}\left(e_{t,\alpha_t}^*\right)}{\partial e_t}$ times the cost of being caught, $e_{t,\alpha_t}^*+\phi^{\alpha_t}\left(e_{t,\alpha_t}^*\right)$, plus the impact of the higher evasion on the fines paid, $p_t^{\alpha_t}\left(e_{t,\alpha_t}^*\right)\frac{\partial\phi^{\alpha_t}}{\partial e}\left(e_{t,\alpha_t}^*\right)$. With this formulation, individual deterrence can occur both because of higher future fines from evasion and from higher future audit probabilities. As noted in the previous section, these future burden and government costs need to be included if current audits lead to more future audits. However, aside from this modification, all of the formulas for the willingness to pay for the audit and cost to the government continue to hold when simply replacing p_t with the equilibrium level of $p_t^{\alpha_t}\left(e_{t,\alpha_t}^*\right)$.[100] For example, suppose the individual deterrence effect arises not because of higher future fines but rather because the IRS can more easily audit someone again and uncover their evasion — in other words, suppose $\frac{\partial p}{\partial e}$ is higher in future periods after an audit. This can lead to a reduction in evasion in future periods (i.e. is a rationale for individual deterrence). It need not lead to an increase in the actual number of future audits, but if it does this would need to be incorporated into the MVPF formula as in Equations (9) and (8).

Income Affects Audit Probability Lastly, suppose income choices affect audit probabilities. To do so, let the audit probability be given by $p\left(y;\mu\right)$ where y is the individual's choice of income and μ is a variational parameter describing potential changes to the audit function. For example, if $p_1\left(y\right)$ is the status quo audit function, we could write $p_1\left(y;\nu\right)=p_1(y)+\mu h\left(y\right)$ for a continuously differentiable positive function, h, and consider a small change in μ starting at $\mu=0$ (assuming the range of $p(y)$ is strictly in the interior of $(0,1)$). Formally, let $y_{i,1}^*\left(\mu\right)$ and $e_{i,1}^*\left(\mu\right)$ denote an individual's choices of income and evasion when faced with the probability of audit given by variation ν. We note that the willingness to pay to avoid the change in audit policy, $d\mu$, is given by:

$$-\frac{dV_i}{d\mu} = \frac{\partial p_1\left(y_1^*,e_1^*\right)}{\partial\mu}\left[e_1^{a_t}+\phi^{\alpha_1}\left(e_{1,\alpha_1}^*\right)+B_i+\left(V_i^0-V_i^1\right)\right]$$

so that

$$\frac{-\frac{dV_i}{d\mu}}{\frac{\partial p_1\left(y_1^*,e_1^*\right)}{\partial\mu}}=R^{mech}+B+\left(V_i^0-V_i^1\right)$$

In other words, the willingness to pay to avoid an additional audit is unchanged when we allow audit probabilities to be a function of income. This is a standard consequence of the envelope theorem.

[100] In order for dV/dp_1 to be well-defined, one needs to add the natural assumption that the audit expansion increases p_1 for all e_1 (otherwise it wouldn't be clear what is meant by dp_1). This assumption is natural as we imagine expanded audits but not necessarily improvements in the audit technology (i.e. improvements in the slope of $p_t\left(e_t\right)$).

However, the net cost to the government of a change in audit policy now includes fiscal externalities from changes in income in response to changes in the audit policy. It is straightforward to see that equation (5) now not only includes the mechanical and future revenue associated with the audit, but also includes the change in income choices, $\frac{\partial y}{\partial \mu}$, weighted by the marginal tax rate, $T'(y_1^*)$. For example, with only a single individual, we have

$$\frac{dG}{d\mu} = \frac{\frac{\partial y_1}{\partial \nu}}{\frac{\partial p_1}{\partial \nu}} T'(y_1^*) + R^{mech} + R^{future} - C$$

where $\frac{\frac{\partial y_1}{\partial \mu}}{\frac{\partial p}{\partial \mu}}$ is the impact of the policy on income choices, $\frac{\partial y_1}{\partial \mu}$, normalized by the net change in audit probability, $\frac{\partial p_1}{\partial \mu}$. If increasing audits on the rich causes people to no longer be rich, we need to estimate this behavioral response and multiply this change in behavior by the marginal tax rate on income, $T'(y_1^*)$. Hence, the MVPF of audits on the rich could have a higher MVPF when incorporating this feature.

Summary In sum, our core MVPF derivation relies on some assumptions — namely no risk aversion, no non-financial costs of the audit, and exogenous audit probabilities —- that can be relaxed if one is able to estimate the additional relevant empirical parameters. The presence of risk aversion would increase the MVPF (relative to what we measure) by a proportion equal to the percentage impact of the audit on consumption scaled by the coefficient of relative risk aversion. The general deterrence effect would be measured as the causal effect on government revenue of an additional audit due to changes in the *ex ante* evasion behavior of people regardless of whether they are audited (de_1/dp_1 in our model). The fiscal externality from these responses are zero under our baseline assumptions, but are nonzero in the presence of risk aversion or non-financial costs of audits. In this case, expanded audits could create general deterrence benefits to the government that would further increase the denominator (and not affect the numerator), leading to a lower MVPF associated with audits. Finally, expanding audits on the rich (poor) could lead to a decrease (increase) in income, which in principle would reduce (increase) the net revenue raised from audits and thus raise (lower) the MVPF. While our baseline results do not provide evidence for the idea that individuals who are audited choose to lower their incomes, we note that ex ante responses could differ from our within-person responses after being audited. The estimation of these additional components of the MVPF are important directions for future work.

D.5 Comparing the MVPF of Audits to the MVPF of Tax Rate Changes

As noted in Hendren and Sprung-Keyser (2020), for any two small policy changes, increasing spending on policy 1 financed by raising revenue from policy 2 will increase social welfare if and only if $\eta_1 MVPF_1 > \eta_2 MVPF_2$, where η_j is the social marginal utilities of income of the beneficiaries of policy j (i.e. giving \$1 to these beneficiaries raises social welfare by η_j). Therefore, we can evaluate the relative desirability of expanding audits versus increasing taxes as a method of raising revenue by comparing the MVPF of expanded audits to the MVPF of tax changes. To do so, it is natural to extend the model above to allow for heterogeneity in income choices and think about these MVPFs separately across the income distribution. Incorporating heterogeneity in incomes into the model is easily introduced by allowing the disutility of earnings to vary across individuals, which we index by θ, $\psi(y_t;\theta)$. The distribution of types in the population in turn generates an income distribution.

We can also compute the MVPF using the formula above conditional on income. If audit rates depend on income, individuals may change their incomes to reduce the probability of audit. For example, increasing the audit probability on top earners could cause people to reduce their reported taxable income, thus increasing the effective cost of the audit (by $T'(y_t)\frac{dy_t}{dp_t}$), and subsequently increasing the MVPF of the audit. This could be captured by including this term in the denominator of the MVPF. For our calculations, we do not make any such adjustment both because we do not have an empirical estimate of this potential behavioral response and also because in practice taking actions to reduce one's income in hopes of preventing an audit can increase the likelihood of the audit. In addition, in our event studies we find no evidence that audits cause reductions in future incomes.

As a result, we can write the MVPF of expanded audits around a given point of the income distribution as

$$MVPF^{Audit}(y) = \frac{R(y) + B(y)}{R(y) - C(y)}$$

where $R(y)$ is the average revenue per audit of those with incomes near y, $C(y)$ is the marginal cost of audits for those with incomes near y, and $B(y)$ is the average taxpayer burden of audits for those with incomes near y. The purple circles in Figure 9 report these estimates of the MVPFs of tax audits by decile of the income distribution.

We can now compare the MVPF of tax audits to the MVPF of changes in the income tax schedule across the income distribution. In the environment above, Hendren (2020) shows that the

MVPF of a tax change targeted to a particular region of the income distribution is given by

$$MVPF^{Tax}(y) = \frac{1}{1 + FE(y)}$$

where $FE(y)$ is the impact of the behavioral response to a small tax cut targeted to those earning near y on the government budget. Under quasilinear utility, this is given by

$$FE(y) = \frac{T'(y)}{1 - T'(y)} \kappa(y) \epsilon^c(y)$$

where $\epsilon^c(y)$ is the compensated elasticity of taxable income with respect to the marginal tax rate, $T'(y)$ is the marginal tax rate at y, and $\kappa(y)$ is the local Pareto parameter of the income distribution. The triangles in Figure 9 shows the shape of $MVPF^{Tax}(y)$ as constructed in Hendren (2020).

D.6 Heterogeneous Welfare Weights and Comparisons Across Policies

Our key finding is that the MVPF for audits of top earners is lower than the MVPF for of tax cuts for income earners, which suggests expanded audits would raise revenue at lower welfare cost than tax rate increases. Formally, raising revenue from expanded audits and using it to finance lower taxes raises welfare if and only if

$$\eta^{audits} MVPF^{audits} < \eta^{tax} MVPF^{tax}$$

where η^{audits} is the incidence-weighted average social welfare weight for those who are audited and η^{tax} is the incidence-weighted average welfare weight of those facing the tax change. Formally, if R_i is the revenue raised from individual i and B_i is the "burden" of the audit for individual i, then $R_i + B_i$ is individual i's willingness to pay for the audit. So, if η_i is individual i's social welfare weight (i.e. social welfare goes up by η_i if we give \$1 to individual i), we have

$$\eta^{audits} = \frac{\sum_i (R_i + B_i) \eta_i}{\sum_i (R_i + B_i)}$$

For example, it could be natural to imagine that the social planner places different weight on compliant and noncompliant taxpayers. Let η_d denote the welfare weight for noncompliant taxpayers and η_h the welfare weight for compliant taxpayers. Let d denote the fraction of noncompliant taxpayers amongst the set of marginally audited individuals. We can then write

$$\eta^{audits} = \frac{d\left(B_d + \frac{R}{d}\right)\eta_d + (1-d) B_h \eta_h}{B + R}$$

where R is the revenue per audit, so that R/d is the revenue per audit of a noncompliant taxpayer (assume no revenue from compliant taxpayers), and $B = dB_d + (1-d) B_h$ is the average taxpayer burden among those audited.

If the social planner places little weight on noncompliant taxpayers, $\eta_d \approx 0$, this will reduce the welfare weight of the audited group to the extent that it contains noncompliant taxpayers (e.g. $d = 1$). For this reason, we reason in the main text that $\eta^{audits} < \eta^{tax}$ so that the fact that $MVPF^{audits} < MVPF^{tax}$ suggests it is more efficient to raise revenue from audits than taxes at the top of the income distribution. A potential countervailing force that is not immediately captured in our model is the fact that auditing compliant taxpayers could impose significant psychological and financial burdens on those taxpayers. In our baseline approach, we quantify the taxpayer burden using the average reported time cost of complying with an audit, but with additional psychological costs B_h could be high.

We can use our results to ask how high B_h would need to be to overturn the conclusion that expanding audits raise revenue at lower welfare cost than raising tax rates. To do so, recall that expanding audits raises revenue at lower welfare cost than raising tax rates if and only if $\eta^{audits} MVPF^{audits} < \eta^{tax} MVPF^{tax}$. The equation above provides η^{audits}. For tax rate increases the welfare weight is just the population welfare weight at the relevant income level (e.g. average welfare weight of top earners). This is the average across compliant and noncompliant taxpayers. Let p denote the fraction of noncompliant taxpayers in the total population (note $p < d$ when audits are targeted more towards noncompliant taxpayers). Then, $\eta^{tax} = p\eta_d + (1-p)\,\eta_h$. So, we have that $\eta^{audits} MVPF^{audits} < \eta^{tax} MVPF^{tax}$ if and only if:

$$\frac{\frac{(1-d)B^h \eta_h + \left(R + dB^d\right)\eta_d}{R+B}}{\eta^h(1-p) + \eta^d p} < \frac{MVPF^{tax}}{MVPF^{audits}} = \frac{MVPF^{tax}}{(R+B)\,/\,(R-C)}$$

where B is the average per audit burden and R is the average per-audit revenue.

If the social planner puts little welfare weight on noncompliant taxpayers, $\eta^d \approx 0$, then audit expansions are preferred to tax rate increases if and only if

$$\frac{\frac{(1-d)B^h}{R+B}\eta_h}{\eta^h(1-p)} < \frac{MVPF^{tax}}{(R+B)\,/\,(R-C)}$$

$R + B$ cancels on both sides so that we are left with

$$\frac{(1-d)B^h}{(1-p)} < MVPF^{tax}\,(R-C)$$

or

$$B^h < MVPF^{tax}\,(R-C)\,\frac{1-p}{1-d}$$

In other words, the planner wishes to increase audits as long as B^h is less than the MVPF of the tax multiplied by the net revenue of the tax multiplied by the ratio of the fraction of compliant taxpayers overall relative to compliant taxpayers among those audited. Taking estimates from the literature that find the MVPF of increasing top tax rates is around 1.5, assuming most taxpayers are compliant, e.g. $p \approx 0.9$, letting d be the fraction of audited taxpayers that see a change in their tax liability (which is around 60%), and using our estimate of the net revenue raised from an audit of a taxpayer in the top 0.1% of the income distribution of around \$103,000 yields

$$B^h < 1.5 * \$103,000 * \frac{0.9}{0.4} = \$348,000$$

In other words, unless those who are audited and found to owe no additional tax are willing to pay more than \$300,000 to avoid the audit, the social planner would wish to raise revenue by expanding audits rather than by raising tax rates on top earners.

D.7 Audit Rates versus Audit Penalties

A large literature on considers the optimal mix of audit rates and audit penalties (Becker, 1968; Allingham et al., 1972), and how that mix varies across the income distribution (Border and Sobel, 1987). While optimal audit penalties are beyond the scope of our empirical analysis, here we illustrate how one can use the MVPF framework to think about the welfare impact of changes in audit penalties relative to increases in audit probabilities. In particular, we first show how one can replicate the classic Becker (1963) and Allingham and Sandmo (1972) result that it is optimal to have high penalties and low audit probabilities when audits are costly. We then discuss briefly how the MVPF provides an empirical path forward for generalizing this conclusion to settings where assumptions made in those models may not hold.

To gain intuition for the MVPF of changes in audit penalties, we can return to the model above and suppose that the government is considering modifying the penalties people must pay in some period. To be specific, suppose the government is considering raising the penalties paid in period 1 to add an additional penalty ν per dollar of money that is evaded. In other words, suppose penalties in period 1 are now given by $\phi^{\alpha_1}(e) = \phi^{\alpha_1}(e) + \nu e$, where e is the amount evaded and ν parameterizes an increase in the evasion penalty rate. With this modification, the FOC for evasion now becomes

$$\frac{\partial \phi^{\alpha_t}}{\partial e}\left(e^*_{t,\alpha_t}\right) = \frac{1-p_t}{p_t} - \nu$$

Assuming penalties are convex, then the increase in ν will reduce evasion, $de^*_{t,\alpha_t}/d\nu < 0$. The question we ask is: what is the welfare impact of changes in penalties relative to changes audit probabilities.

To assess this, note that the social welfare impact of increasing audit penalties can be written as $\eta^{penalty} MVPF^{penalty}$, where $\eta^{penalty}$ is the average social marginal utility of income of those who are subjected to the higher penalties and $MVPF^{penalty}$ is the MVPF of higher penalties, given by the ratio of people's willingness to pay to avoid the penalties relative to the net revenue raised by those penalties (which includes both changes in penalty revenue and also changes in voluntary tax revenue paid from deterrence effects). Raising revenue by raising penalties instead of raising audit rates will be preferred whenever

$$\eta^{audits} MVPF^{audits} > \eta^{penalty} MVPF^{penalty} \tag{11}$$

In Allingham and Sandmo (1972) and Becker (1968), there is a representative agent so that there is no difference in social welfare weights, $\eta^{audits} = \eta^{penalty}$. The key question is then: what is the MVPF of raising penalties? To start, note that an increase in penalties in period 1, $d\nu$, will generate a willingness to pay to avoid those increased penalties of $e^*_{1,\alpha_1} d\nu$. The revenue generated from the increased penalties in period 1 will be the sum of the mechanical revenue generated, $e^*_{1,\alpha_1} d\nu$, plus the impact of the behavioral response to the penalties on the government budget (i.e. the general deterrence effect). However, because penalties are entirely financial, individuals' evasion choices minimize government revenue. Hence, the behavioral response to the change in ν has no effect on government revenue. As a result, the cost to the government is simply the mechanical revenue, meaning the MVPF is given by

$$\begin{aligned} MVPF^{penalty} &= \frac{e^*_{1,\alpha_1} d\nu}{e^*_{1,\alpha_1} d\nu + \frac{de^*_{1,\alpha_1}}{d\nu} \left[(1-p_1) - p_1 \frac{\partial \phi^{\alpha_1}}{\partial e} \left(e^*_{1,\alpha_1} \right) \right]} \\ &= \frac{e^*_{1,\alpha_1} d\nu}{e^*_{1,\alpha_1} d\nu} \\ &= 1 \end{aligned}$$

In contrast, recall that the MVPF of expanded audit rates is always greater than 1: individuals are willing to pay the mechanical revenue to avoid the audit, but the revenue to the government is lower due to the net cost of expanded audits. Hence, $MVPF^{penalty} = 1 < MVPF^{audits}$. This is the classic logic of Becker (1968): raising penalties raises revenue more efficiently than raising audit rates.

This conclusion rests on strong assumptions about how taxpayers respond to changes in audit rates and penalties and about the social planner's preferences. As noted above in Section D.4.2, the MVPF framework can incorporate cases where general deterrence has first order impacts on government revenue – given empirical estimates of the deterrence effects of raising audit rates and raising penalties, one can construct more general MVPF measures of these policies that allow for general deterrence to affect government revenue. In addition, the MVPF framework is not limited to a representative agent and can place differential welfare weights on audited individuals — social planners may find it undesirable to give very low consumption to those being fined. Given estimates of the willingness to pay to avoid the higher penalties combined with the net revenue raised by the penalties, one can compare the relative desirability of expanded audits versus higher penalties using Equation 11. In this sense, the MVPF framework provides a path for replacing the theoretical assumptions with empirical elasticities on how audit design affects behavior and revenue.