A WELFARE ANALYSIS OF TAX AUDITS ACROSS THE INCOME DISTRIBUTION *

William C. Boning
Nathaniel Hendren
Ben Sprung-Keyser
Ellen Stuart

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 from comprehensive internal accounting information and audit-level enforcement logs. We begin by estimating the average initial return to all audits of U.S. 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 effect 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

*We gratefully acknowledge research assistance from Jamie Emery and Sethu Odayappan. We are 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. We thank several people at the 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 article 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.

The Quarterly Journal of Economics (2025), 63–112. https://doi.org/10.1093/qje/qjae037.
Advance Access publication on November 8, 2024.

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. *JEL codes:* H21, H26.

I. Introduction

Internal Revenue Service (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 (Johns and Slemrod 2010; DeBacker et al. 2020; Guyton et al. 2021). Evidence indicates that on average, tax audits can recoup unpaid taxes and raise revenue in the process. However, there is 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 high-income taxpayers reduce the "bang for the buck" of enforcement spending? In this article, 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] 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. We combine this data with the General Scale (GS) classification for each employee to estimate the hourly costs of all direct enforcement activities.

1. We focus on in-person audits rather than audits conducted by correspondence. Data limitations prevent a comprehensive analysis of the returns to correspondence audits. In particular, we are unable to measure the marginal costs or 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 the pattern of returns across the income distribution is qualitatively similar to the pattern we find for in-person audits.

Second, we use detailed internal IRS business unit accounting information. This data provides 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 nonauditing hours, the cost of management, and fringe benefits of employees, as well as central overhead costs such as building and technology service costs. To the best of our knowledge, this cost information has not been used in academic studies of tax enforcement. We show that this cost data is nonetheless critical for constructing 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.

We combine our measures of audit costs and audit revenue to estimate the average return to audit expenditures. We estimate that every $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 half 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. These rising costs across the income distribution are, however, more than offset by increasing revenues. We estimate that audits of taxpayers in the bottom half of the income distribution produce $0.96 in revenue for each dollar of audit cost. Audit spending in the top 1% and top 0.1% produce returns of $4.25 and $6.29, respectively.

In the second step of our analysis, we estimate the costs and revenues associated with a marginal audit expansion. 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 to assess the returns to reversing them. If audit selection decisions prioritized cutting back on audits with low rates of return, we would expect that the remaining audits would have a higher average rate of return. In fact, we see no such pattern. Revenues per hour spent auditing remained stable during this period. We show that returns remained stable because audit revenues within income groups did not change and audit rates

declined similarly across 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 results suggest that expanding audit rates from 2014 back to their 2010 levels could deliver revenues per audit similar to the average revenues we estimate.

We also consider the costs associated with marginal audit expansions. It is natural to expect some economies of scale in the nondirect costs of conducting audits. 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. 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% and top 0.1% would return $5.82 and $8.62, respectively.

In the third part of our analysis, we study how audits can 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 et al. (2018), henceforth DHTY, who use randomly selected audits as part of the IRS National Research Program (NRP) to estimate the magnitude of individual deterrence effects in the United States. We extend their analysis by using an additional decade of postaudit data and a stratified matched-control strategy, which allows us to estimate long-run individual deterrence effects and examine the heterogeneity in deterrence effects across the income distribution.

We show that audited taxpayers pay more tax in the years following an audit. The additional tax payments persist over the 14 years we observe in the data. In present discounted value, these additional tax payments raise 3.2 times as much revenue as the revenue raised from the initial audits. These deterrence effects are similar across the full income distribution, although we lose the power necessary to estimate precise effects in the top 1%. 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.

In the fourth and final step, 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 taxpayers' willingness to pay to avoid the audits divided by the net revenue raised. The MVPF captures the welfare cost imposed per dollar of government revenue raised and can be compared with other methods of raising revenue, such as changes in tax rates.

Estimating the MVPF of audits requires calculating not only the revenue raised per dollar spent on audits, but also the burden imposed on taxpayers during the audit. We draw on new evidence from a survey of recently audited taxpayers that measures the time and money that taxpayers at various income levels spent to comply with the audit. 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. Although this is a meaningful burden, it is substantially smaller than the revenue collected.

Combining our estimates of audit revenue, audit costs, and the burden of complying with an audit, 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. The declining pattern in income means that expanded audits on high-income taxpayers raises revenue at a lower welfare cost than expanded audits on low-income taxpayers. In stark contrast, tax increases of high-income earners typically have much larger MVPFs than tax increases on low-income earners.

We also show how the use of welfare weights in the MVPF framework readily incorporates concerns for horizontal equity (the desire that all taxpayers with the same level of true income pay the same amount in taxes). We calculate the fraction of audits at each income level that result in additional assessed tax liability. Such taxpayers are considered noncompliant with regard to their tax obligations. To distribute the burden of audits between compliant and noncompliant taxpayers, we use new survey data that report the average burden of audits across the income distribution separately for compliant and noncompliant taxpayers. Using these inputs, we show that nearly all of the welfare costs imposed by marginal

audits are borne by noncompliant as opposed to compliant taxpayers. Although determinations of policy optimality depend on the welfare weights placed on taxed or audited individuals, our results suggest that expanding individual audits raises revenue at a lower welfare cost than increasing tax rates.

This article relates to several strands of previous literature. A recent body of work calculates the average returns to audits (Holtzblatt and McGuire 2016, 2020; CBO 2018, 2020). There is also a robust literature on the size of the tax gap across the income distribution (Johns and Slemrod 2010; DeBacker et al. 2020; Guyton et al. 2021).[2]

There is, however, comparatively little empirical work quantifying the costs and benefits of tax enforcement across the income distribution.[3] Sarin and Summers (2019) provide an estimate of the returns from auditing taxpayers with more than $5 million in income. Their analysis is meant to be illustrative, rather than providing a full accounting of costs and revenues across the income distribution.[4] We fill this gap by providing detailed estimates of the returns to audits across the income distribution.

We also build on existing literature measuring the individual deterrence effects of audits. Our core contributions are identifying the long-run effects of audits and studying how individual deterrence effects vary across the income distribution. We build most closely on DeBacker et al. (2018) by including additional years of data and using a matching strategy to allow comparisons within fine-grained income groups. Beer et al. (2020) also examine short-run deterrence effects among self-employed U.S. taxpayers using nonrandom operational audits, finding results broadly consistent with our estimates. Our work also relates to the large body of evidence on deterrence measured outside the United States (Kleven et al. 2011; Best, Shah, and Waseem 2021; Bjørneby, Alstadsæter,

2. Alstadsæter, Johannesen, and Zucman (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 across the wealth distribution.

3. 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 Elzayn et al. 2025).

4. In Online Appendix D, we provide a more detailed comparison between our estimates and those of Sarin and Summers (2019) and Holtzblatt and McGuire (2020).

and Telle 2021; Mazzolini, Pagani, and Santoro 2022; Kasper and Alm 2022a, 2022b; Advani, Elming, and Shaw 2023; Hebous et al. 2023; Løyland et al. 2024). For example, Kleven et al. (2011) estimate deterrence effects for a single postaudit study year in Denmark, and Advani, Elming, and Shaw (2023) estimate deterrence effects in the United Kingdom up to eight years postaudit. These papers find per year deterrence effects between 20% and 35% of initial audit revenue, broadly consistent with our per year estimates. Crucially, we find that the deterrence effects of audits persist for at least 14 years in our setting.

This article 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). 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 correspond to differences in the distortionary costs of audits. Our estimates of the returns to tax audits allow for the estimation of those distortionary effects. Our discussion of the MVPF of tax audits formalizes the parallel with tax rates.

Finally, our MVPF approach connects to a theoretical literature on optimal tax administration (see 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. Comparing the MVPFs of expanded tax audits to the MVPFs of modifications to the tax schedule puts this concept into practice. 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 the article proceeds as follows. Section II provides an overview of the audit process and describes our data and sample. Section III presents the results for the average costs of audits and average revenue raised, reporting their heterogeneity across the income distribution. Section IV estimates the

returns to marginal audits, rather than average audits. Section V studies the individual deterrence effects of audits, measuring the impact on future tax revenue. Section VI analyzes the welfare consequences of audits using the MVPF framework. Section VII concludes.

II. Data and Sample

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

II.A. Audit Overview

In this article we focus on in-person audits of individuals.[5] Online Appendix Figure A.I provides a flow chart that maps out the audit process. The 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. Office audits are performed by tax compliance officers and often include an interview at an IRS office. Office audits tend to involve less complex issues. Field audits are conducted by revenue agents and often include one or more interviews at the taxpayer's home or place of business. Field audits focus on more complex issues. 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 determine whether any corrections should be made to the tax liability. The exam process typically includes reviewing documents not submitted with the tax return, such as receipts that verify the validity of deductions or bank records that validate total reported income. In some cases, an examination will expand to include returns filed

5. 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 Online Appendix B.2.

in additional tax years. For 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.[6]

After the examiner recommends any corrections to the taxpayer's tax liability, the next phase of the audit process depends on whether the taxpayer agrees with the examiner's recommendations. If the taxpayer agrees, then any additional tax is assessed and becomes due. (If the examiner recommended a reduction in tax liability, a refund is issued.) If the taxpayer disagrees, the unagreed on 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 moves to tax court for a final determination.

Once a taxpayer pays the full amount of additional assessed tax, 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 them to find a payment solution. 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.

Most of the analysis in this article focuses on in-person operational 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 NRP. The NRP is designed to provide critical information on tax compliance and aid in the IRS's estimate of the tax gap.[7] To achieve those goals, the IRS assigns revenue agents to examine all areas of a randomly selected sample of returns. NRP examinations are more

6. Here we depart from prior research and IRS statistics that consider each tax year as a separate audit. 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.

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

intensive than a typical in-person audit because they examine all line items on the return, but they follow the same basic steps outlined in Online Appendix Figure A.I.

In addition to in-person audits, the IRS conducts correspondence audits. Correspondence audits are conducted by mail and tend to focus on simpler issues arising from inconsistencies across tax forms (e.g., two returns claiming the Earned Income Tax Credit [EITC] for the same child). Our primary analysis does not examine the returns to correspondence audits for two reasons. First, it is difficult to measure the marginal cost of correspondence audits. Conducting additional correspondence audits potentially requires fixed-cost investments, such as developing and deploying new algorithms to identify noncompliance. It is difficult for us to quantify the marginal costs of these expenditures. By contrast, expanding in-person audits mostly requires hiring more auditors, so it is more straightforward to measure the marginal costs. The second reason we do not focus on correspondence audits is that there are no random audits conducted solely via correspondence, which makes estimation of deterrence effects more difficult than for in-person audits.[8] That said, we show in Online Appendix C (and Online Appendix Figures A.II and A.III) that the returns to correspondence audits and in-person audits are quantitatively similar, both on average and across the income distribution.

II.B. Data

Our data are primarily drawn from three internal IRS sources. First, we use case-level data that contain detailed information on IRS enforcement activities at each stage of each audit. The data include time spent by IRS enforcement personnel and the revenue collected. Second, we use internal accounting data for the audit divisions of the IRS. These data provide information on costs other than the direct hourly wage costs for the auditors (e.g., employee benefits, management, rent, IT). Finally, Section VI presents results from a representative survey that asks audited taxpayers about the time they spent and expenditures they made during the audit process. We defer a detailed discussion of the survey to the welfare analysis in Section VI.

8. 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.

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

1. *Audit-Level Enforcement Data.* We observe audit-level 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 during the audit process. For each audit conducted, the database 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 information into wage costs using a location-specific GS hourly pay scale and the ZIP code 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.

2. *Internal IRS Accounting Data.* The cost of conducting audits goes beyond the time cost of direct enforcement activity. We use detailed internal IRS business unit accounting information from fiscal years 2011–2020 to identify and incorporate these costs.[9] These data have not 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: (i) nondirect labor-related costs, (ii) organization-wide costs, and (iii) central IRS management overhead. Nondirect labor-related costs include time spent by auditors on nonauditing activities (such as training), wage costs for management and support staff, and nonwage costs for employees such as fringe benefits and workers' compensation. Organization-wide costs include items such as building costs and information technology costs. Finally, central IRS management overhead includes costs incurred

9. 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.

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 subcomponents can be found in Online Appendix B.2.

We calculate direct wage costs separately for the exam, appeals, and collection stages of the audit.[10] For the exam stage, we estimate cost multipliers for each major component of overhead: nondirect 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 nonaudit labor costs and fringe benefits are proportional to labor costs. Online Appendix Figure A.IV shows the robustness of our results to two alternative methods of cost allocation: (i) by hours spent on the audit and (ii) equally across audits. When nondirect labor costs are allocated in proportion to labor hours, the rates of return across the income distribution are essentially unchanged. When nondirect labor costs are allocated on a per audit basis, fewer costs are assigned to high-income audits (which require more hours from higher-wage auditors). Our baseline approach is intended to be relatively conservative given that our primary results show that the return to audits increases with taxpayer income.

The costs from the exam stage make up 93% of total audit costs. For the other audit costs, we apply a similar method for direct hours during the appeals process, 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. Online Appendix B provides further details.

II.C. Sample

We focus our analysis on audits of individual tax returns filed in tax years for which (i) we observe comprehensive measures of costs and (ii) sufficient time has passed that nearly all audits

10. We use the administrative enforcement data to measure the direct labor costs for both the exam stage 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. We impute direct labor costs of collection for cases that started in the exam stage before moving to collections by deciles of income and amount owed. We report the details of the imputation method in Online Appendix B.

have been completed. This leads us to study the universe of audits initiated by reviews of individual tax returns from 2010 to 2014.

We begin with the 2010 tax year because the audits of those returns begin 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 seven years to observe any follow-up costs and revenues associated with an audit. Our baseline sample includes approximately 710,000 in-person audits.[11]

For each audit, we link the taxpayer to their Form 1040 in the tax year that triggered the audit. We use this linked 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. TPI is the primary measure used by the IRS to categorize returns by income for audit selection.[12] We deflate all income values to constant 2016 dollars using the CPI-U-RS.

III. Average Revenue and Costs per Audit

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

III.A. 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, reported in Figure I.

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

12. TPI excludes losses, which avoids concerns about the accuracy of reported losses. Consequently, TPI 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 preaudit. Below, we discuss how our results could potentially differ if one were to use postaudit income. We conclude that the general patterns of our findings are largely unchanged.

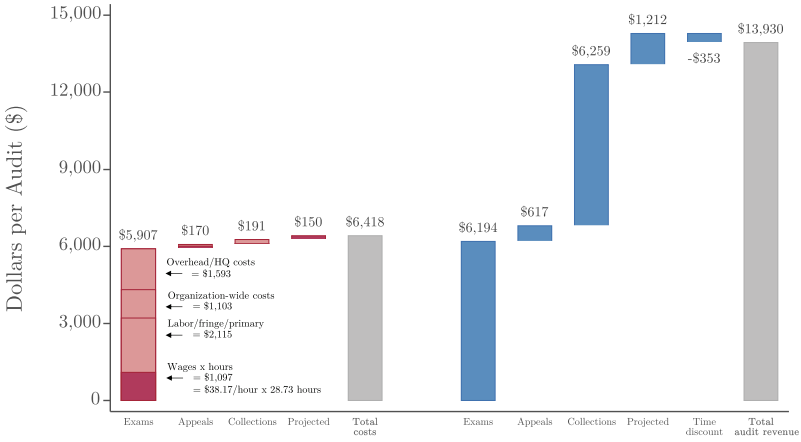


FIGURE I

Average Costs and Revenue Raised per In-Person Audit

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 7.72 times total wage costs at the exam, appeals, and collections stage, respectively, using average multiplier values from 2011 to 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 Online Appendix Figure A.I. Average costs and revenues include projected costs incurred and revenue collected after the observed 7–11-year postaudit 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 postaudit 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.

The total cost of the hours spent on the audit also includes overhead costs. The first component of overhead is additional labor costs not directly allocated toward an audit, including wage costs for nonauditing hours of auditors, fringe benefits, training costs, and manager labor costs. We allocate these costs in proportion to the 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, such as 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, which result in an additional $1,593 per audit. Taken together, the $1,097 spent on audit hours comes with an additional $4,811 in overhead costs at the exam stage, for a total of $5,907.

Our 4.39:1 estimated overhead ratio is meaningfully larger than the ratios 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 nonlabor overhead costs and labor costs for support staff and management. Our measure of overhead costs is an average cost measure and may not be the same as the overhead costs associated with marginal expenditures on audits. In Section IV, 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 fraction of taxpayers file 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: $108 in direct labor hours costs and $62 in additional costs. In addition, some cases end up in collections, which we estimate leads to an additional $191 in costs.[13]

Finally, 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.[14] To account for those potential costs, we consider the trajectory of costs accrued for returns filed in the 2003 tax year. Online Appendix Figure A.V plots these costs. We draw on tax year 2003, which lies before our primary sample frame, because it provides 18 years of follow-up data.[15] We

13. 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.

14. The IRS typically has three to six 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 six-year mark are typically part of the collections or appeals process.

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

estimate that in our 7–11-year sample frame, we capture 96%–99% of total labor hours spent on audits.[16] Our forecasts suggest that we do not observe an average of $150 in additional future spending on the audits in our sample. Incorporating these costs yields a total average cost of $6,418 for each in-person audit.

III.B. Average Total Revenue

How much revenue do audits generate? Figure I 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 with our cost estimates, we use 2003-based projections to estimate the 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, with the exact percentages varying by TPI bin (see Online Appendix Figure A.V). Summing across those TPI bins yields average additional revenue of $1,212 per audit.

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. The costs of the audit are, however, generally incurred before the revenue is obtained. Online Appendix Figure A.V demonstrates that average revenues lag average costs by approximately one year. We apply a 3% discount rate to align the present values of the costs incurred and revenues collected. Discounting reduces the return to an average audit by $353.[17] This adjustment results in total revenue of $13,930, which is 2.17 times higher than the total average cost.

III.C. Heterogeneity by Income

We analyze how revenues and costs vary by taxpayer income. Figure II, Panel A presents the average revenues and costs for

16. In particular, we estimate the trajectory of costs accrued in each postfiling year for each decile of TPI. We then aggregate across all years and all observed income bins to produce this average.

17. 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 use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downward to align the two paths.

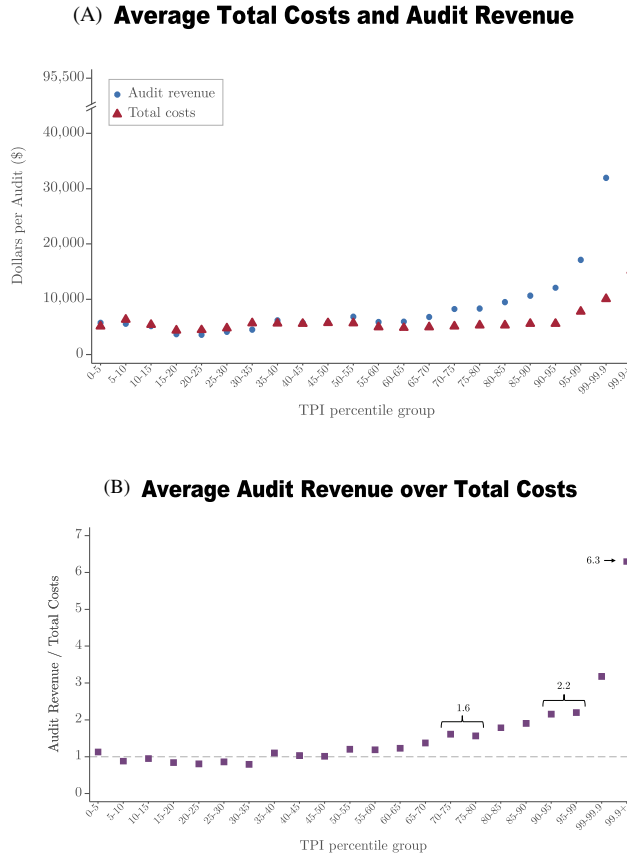


FIGURE II

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

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 × 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 postaudit 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 postaudit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downward to align the two paths.

in-person audits separately by percentiles of taxpayer TPI.[18] We split income on the horizontal axis into bins containing five percentiles, breaking out the top 5% into the 95–99th percentiles, 99–99.9th percentiles, and the top 0.1%. The red triangles plot average audit costs, and the blue dots plot average audit revenues by income bin.

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%. Online Appendix Figure A.VIII shows that this increase in costs with income 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, whereas auditing taxpayers in the top 0.1% requires 64.6 hours. Audits of high-income taxpayers are also conducted by more experienced auditors who receive higher wages. As shown in Online Appendix Figure A.VIII, audits of taxpayers at the 70–80th income percentiles are conducted by auditors earning an average of $36.86 per hour and audits of taxpayers at the top 0.1 percentile are conducted by auditors earning an average of $45.66 per hour.

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, auditing a taxpayer in the 90–99th percentiles yields $14,973, and auditing a taxpayer in the top 0.1% yields $95,491.

Figure II, 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

18. Our primary analysis uses observed TPI on the originally filed tax return. This is the measure available at the time of the audit and therefore the relevant definition of income from the perspective of the IRS when deciding whether to conduct an audit. It is also useful, however, to understand the returns to audits as a function of the corrected income of the taxpayer. Online Appendix Figure A.VII shows that we obtain qualitatively similar, but slightly accentuated, patterns when using a postaudit-assessment imputed TPI measure 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 to more than 8:1.

considerably with income.[19] On average, each dollar spent to audit a taxpayer in the bottom half of the income distribution returns $0.96 in revenue. This return rises to $1.58 in the 70–80th income percentiles, $2.18 in the 90–99th income percentiles, and $6.29 in the top 0.1%.[20]

IV. Marginal Audits

What are the returns to expanding (or contracting) the number of audits conducted? In the previous section, we calculated the average cost and average revenue associated with in-person audits. Here we explore the returns to marginal audits. The revenues and costs of marginal audits may differ from the average revenues and costs for two main reasons. First, if the audit selection process seeks to maximize revenue per dollar of audit cost, then additional audits may have diminishing marginal revenues. Second, economies of scale mean that some costs—particularly those other than direct labor costs—may not increase linearly with audit hours. We explore how both diminishing marginal revenues and economies of scale shape the returns to marginal audit expansions.

IV.A. (Lack of) Diminishing Marginal Revenue

We begin our analysis of the returns to marginal audits by studying the steep audit rate decline for returns filed between the 2010 and 2014 tax years.[21] As shown in Figure III, Panel A,

19. While our analysis focuses on in-person audits, Online Appendix Figure A.III 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%.

20. These calculations use 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 cost-intensive than audits of low-income taxpayers.

21. We end our analysis with 2014 tax returns because audits from subsequent tax years may be ongoing (see Section II). Consequently, we have a less complete picture of the returns filed in subsequent years. Online Appendix Figure A.VI 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 III.

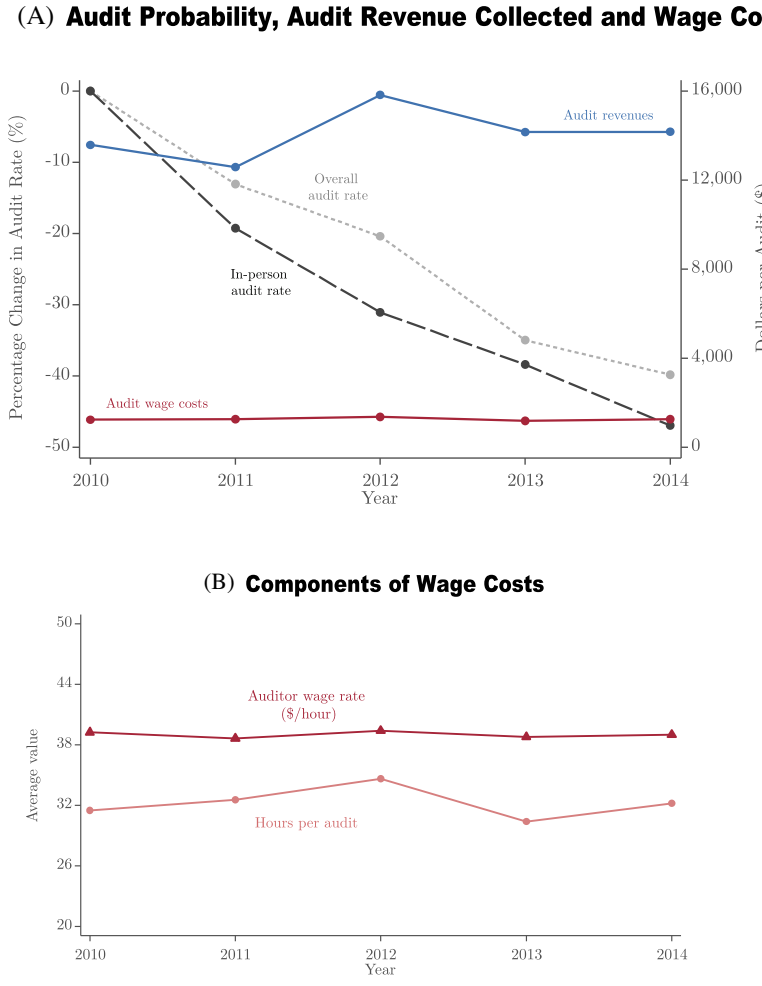


FIGURE III

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

Panel A presents the percentage change in overall and in-person audit rates, total revenues raised, and direct labor costs (auditors' wages × 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 postaudit 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 postaudit back to the tax year. We use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

audit rates declined by 40% between tax years 2010 and 2014. The IRS audited 0.92% of all returns from tax year 2010 but just 0.56% of returns from 2014.[22]

We use the reduction in audit rates to understand the revenue generated by marginal audits. We conceptualize marginal audits as those that would have been conducted at 2010 audit rates but not conducted at 2014 audit rates. This exercise sheds light on the potential returns to expanding audit expenditures back to their 2010 levels.

The return to reversing this decline in audit rates depends, in part, on how the IRS selects returns for audit. If audit reductions between 2010 and 2014 prioritized cutting audits with lower 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, and so on—then the returns to marginal audits may not differ from the returns to average audits.

Figure III, Panel A displays the average revenue per audit and average direct wage cost per audit between 2010 and 2014. Despite the sharp decline in audit rates in this period, the revenue and cost per audit remain essentially unchanged. In 2010, revenues were 11.0 times direct wage costs. In 2014, revenues were 11.2 times direct wage costs.[23] This suggests that the types of audits cut between 2010 and 2014 had similar returns to the audits that were still conducted in 2014.[24] In other words, the return to marginal audits was the same as the return to average audits.[25]

22. The rate of in-person audits declined 47% between 2010 and 2014. In 2010, 0.21% of all returns were selected for in-person audit compared with 0.11% in 2014.

23. 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.

24. We find similar results 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.

25. This pattern 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 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. In Online Appendix D, we discuss

The similarity between average and marginal returns holds across the income distribution. In particular, we find that the reduction in audit expenditures was relatively consistent across the income distribution and that the returns to audits remained stable at each income level. Figure IV, Panel A reports the change in the total direct wage costs of audits for each 5% 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%.[26] The 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 IV, 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.[27] We find a slope close to 1 (1.05, std. err. 0.03) and intercept close to 0 (0.21, std. err. 0.34), showing that across TPI bins, there is no systematic change in the returns to audits between 2010 and 2014 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. Changing evasion rates could produce a reduction in the returns to audits. In such a scenario, changes in the audit selection process could have increased the average return to audits between 2010 and 2014, but this increase could have been masked by a compositional shift in the amount of tax evasion

the differences between our approach and the samples used for these alternative estimates.

26. 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.

27. Online Appendix Figure A.IX reports the analogous figure for 2010 and 2014 revenue per audit by TPI bin.

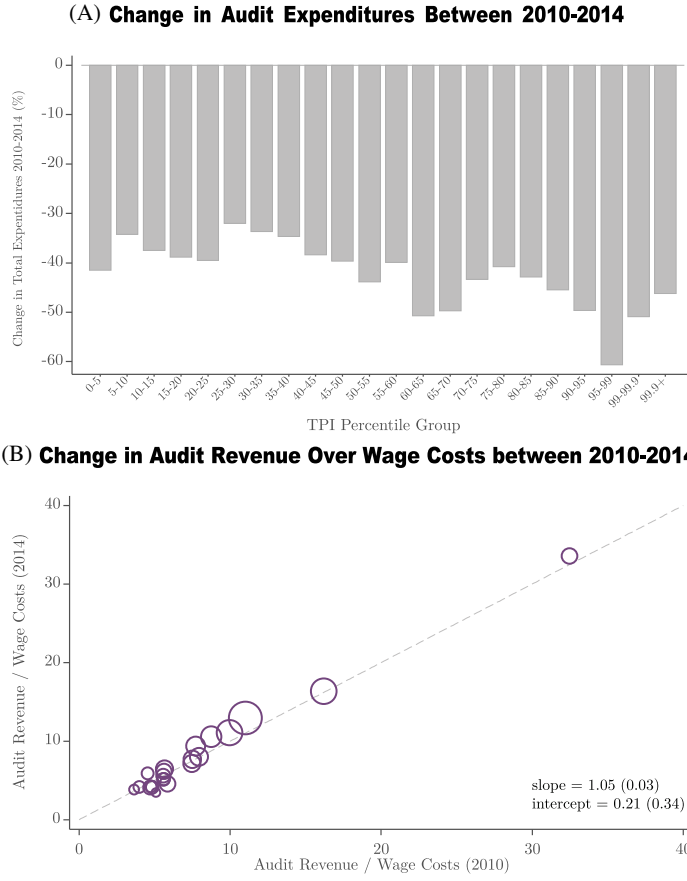


FIGURE IV

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

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 × 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 postaudit 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 postaudit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downward to align the two paths.

in the first place. We mitigate this concern by drawing on audit data from the NRP. Each year, the NRP randomly selects a sample of returns for review. NRP audits are more intensive than operational audits, examining all line items rather than only selected items, because they are designed to measure tax compliance. If tax evasion declined over time, then NRP audits would detect less evasion in later years. Online Appendix Figure A.X, 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. Online Appendix Figure A.XI repeats this analysis separately by income bin. It shows there is no systematic decline in NRP audit returns within any income group.

Together, the evidence suggests that the marginal audits cut between 2010 and 2014 produced similar revenue (per dollar of auditor wages) to the audits that remained. Although estimating the IRS's audit selection priorities is beyond the scope of this article, the evidence suggests that audit cuts between 2010 and 2014 were driven by considerations other than maximizing revenue per dollar of audit cost. Our findings are 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 focuses on ensuring broad compliance with the tax code.

Our estimates of stable returns from 2010 to 2014 suggest that expanding audit rates back to 2010 levels would yield marginal revenue close to average revenue. It is important to note, however, that these results are local to the large (∼40%) decline in audits that occurred in this period (or the 66% expansion necessary to return to previous audit levels). They may 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.[28]

28. This is clear from 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. Online Appendix Figure A.XI, Panel A shows that, for example, the average revenue per NRP audit of taxpayers in the the 99–99.9th income percentile is $9,432. In-person operational audits of

IV.B. Marginal Costs

Estimating the return to marginal audits requires accounting for differences between average costs and marginal costs. In particular, overhead costs may not scale proportionately with the number of hours spent auditing, bringing marginal costs below average costs.

Our baseline estimate is that total 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.

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. Labor/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.[29]

This estimate of the marginal cost structure for expanded audits is 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 nonlabor costs in those requests are approximately

taxpayers in this income group collect an average of $28,451, about three times as much. Lower returns from NRP audits suggest that current selection procedures have higher returns than randomly selected audits would. 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 reauditing the same taxpayers.

29. 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).

23.6% as large as total labor and fringe costs.[30] By comparison, in our baseline estimates total marginal nonlabor costs are 34.3% as large as total labor and fringe costs.

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

Combining our estimates of marginal revenues and marginal costs, we calculate the marginal return to expanding audits from 2014 to 2010 levels. We start by combining the average costs and average revenue of audits across the income distribution from Figure II. We then adjust 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. Figure V shows that 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 top 0.1%.

V. 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 indirectly by deterring future noncompliance. 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

30. See https://home.treasury.gov/system/files/266/02.-IRS-FY-2022-CJ.pdf, 127–129. 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 and other direct costs. Taking the ratio of $\frac{20,111}{85,074}$ yields 23.6%. We also note this figure is broadly consistent with internal IRS estimates of large-scale audit expansions.

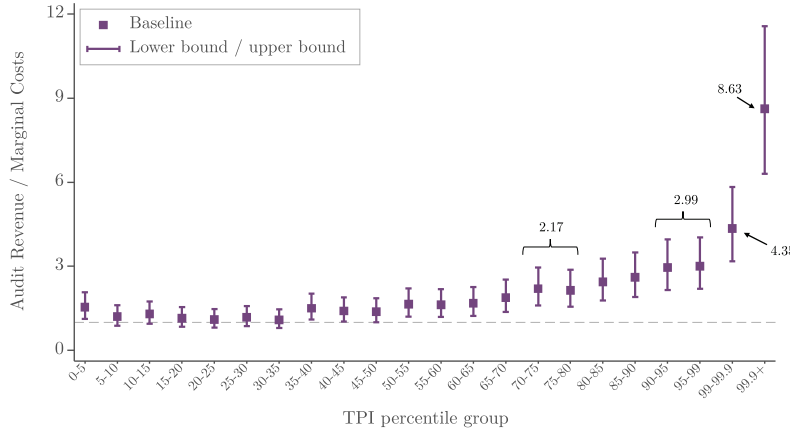


FIGURE V

Average Audit Revenue over Marginal Costs

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" nondirect labor costs as variable costs (i.e., labor, benefits, training, travel). The upper-bound case assumes all nondirect 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 $\times$ 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 postaudit 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 postaudit back to the tax year. We use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downward to align the two paths.

years.[31] Individual deterrence can be an important mechanism through which audits affect revenue.

To estimate individual deterrence effects, we draw upon random audits conducted by the NRP and build on the approach developed by 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. We construct our control group using stratified matching. The NRP divides taxpayers into strata on the basis of return characteristics, including 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 that has the closest value of TPI. Stratified matching produces a control group with similar pre-treatment income levels and income sources as the treatment group. We track income reported and taxes paid in each year after the initial audit.

Our approach builds on DHTY in two main ways. First, we use data on revenue collected through 2021 to substantially increase the duration of observed postaudit period. We analyze 14 years of postaudit revenue as compared with 6 years in DHTY. (We also examine twice as many random audits by examining the full set of NRP audits conducted from 2006 to 2014.)

Second, our event-study framework incorporates year-since-audit fixed effects. As we detail shortly, the omission of these controls in DHTY results in a mismatched comparison between treated and control taxpayers. That comparison produces a spurious negative deterrence effect at the top of the income distribution.[32]

31. 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 and Sandmo 1972). In this scenario, increasing audit rates causes even those who are not audited to change their behavior. We do not quantify the role of general deterrence in this article but return to a discussion of general versus individual deterrence in our discussion of the welfare effects of audits.

32. 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 for taxpayers overall.

Figure VI, 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 taxpayers is statistically indistinguishable from zero in the years before the NRP audit. In the years after the audit, a clear gap emerges. Starting in year 2, we find a statistically significant deterrence effect.[33] The yearly impact on taxes paid is around $300 per audited taxpayer, which is approximately 30% of the $1,026 in revenue collected by the initial audit.[34] The 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.[35] There is no clear change in the magnitude of the effect over the full 14-year window. 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.[36] This

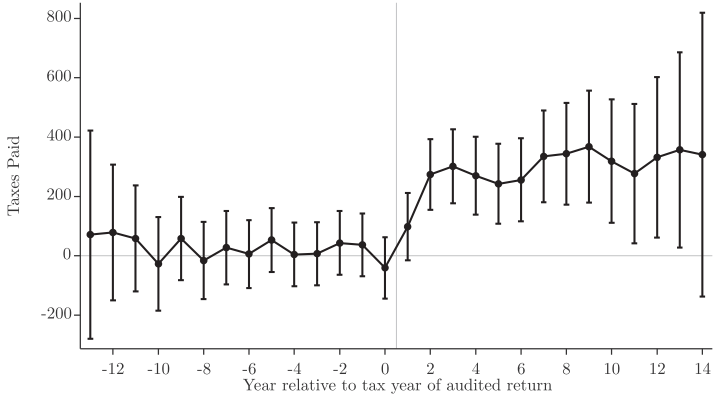
33. The presence of a smaller treatment effect in year 1 is consistent with the fact that the audit process itself is not always conducted within a year of the initial tax filing. As a result, taxpayers may file their year 1 taxes before the NRP process is complete.

34. The magnitude of this effect is consistent with the previous literature, which finds annual deterrence effects ranging from 20% to 35%. Advani, Elming, and Shaw (2023) find a slight decay in point estimates and increase in standard errors over time such that eight years postaudit they are unable to reject the null hypothesis of no effect. The slight decay over time is consistent with our findings for audits of taxpayers with business income (Online Appendix Figure A.VII). In contrast to Advani, Elming, and Shaw (2023), however, we continue to find statistically significant effects after eight years.

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

36. The sign of this deterrence effect is not obvious ex ante. Slemrod, Blumenthal, and Christian (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.

(A) **Within-taxpayer impact of audits on future tax payments**



(B) **Deterrence effect over initial audit revenue, by income**

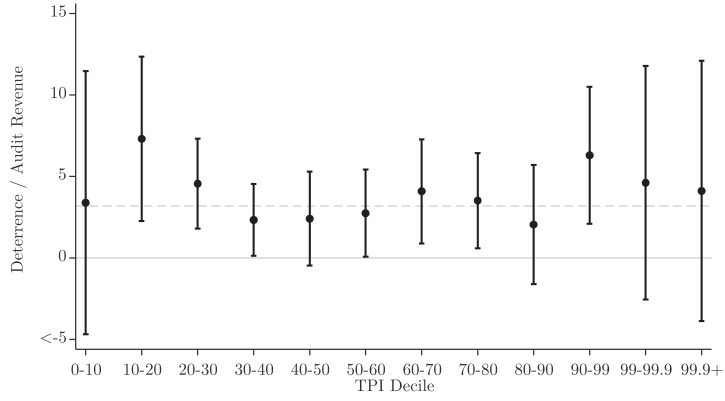


FIGURE VI

Estimated Deterrence Effects

Panel A presents estimates of the change in taxes paid each year postaudit 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-differences 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 postaudit 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.

effect is roughly 3.2 times the revenue collected from the initial NRP audit.

This result is robust along several dimensions. Online Appendix Figure A.X shows that these results are consistent across NRP audit waves, suggesting that the effects are stable over time and that the unbalanced nature of our panel does not introduce bias. Online Appendix Figure A.XII explores how deterrence effects differ for taxpayers with and without business income (as measured by income on Schedules C, E, or F). For taxpayers without business income, the future revenue collected is roughly 3.81 times initial NRP revenue. Consistent with previous literature, we find slightly less persistent deterrence effects for taxpayers with business income, with a ratio of 2.81. Finally, Online Appendix Figure A.XIV studies deterrence effects among a sample of individuals chosen for typical in-person audits, rather than those chosen at random for an NRP audit. Using a matched difference-in-differences design, we obtain a multiplier of approximately 2.5, consistent with the results for NRP audits. These results alleviate concerns that the results from the NRP sample do not generalize to typical in-person audits.

We next use our baseline sample of NRP audits to explore how deterrence effects vary across the income distribution. Table I reports the results of the treatment-control difference-in-differences specification separately by bins of TPI.[37] We display both raw estimates; following DHTY, we winsorize collected tax revenue at the 99th percentile of the population distribution to limit the influence of outliers. Broadly, we find significant deterrence effects across the income distribution. Figure VI, Panel B summarizes these results by plotting the ratio of the estimated present discounted value of deterrence effects relative to the upfront revenue collected from the NRP audit. 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.[38] While we lack the power to estimate precise

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

38. Online Appendix Figure A.XV repeats the exercise found in Figure VI, Panel B 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.

TABLE I

DETERRENCE EFFECTS VERSUS MECHANICAL AUDIT REVENUE

	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	1,354.76 [−1,883.91, 4,593.42]	3.38 [−4.70, 11.46]	426.20	1,510.64 [−2,244.12, 5,265.40]	3.54 [−5.27, 12.35]
10–20	444.89	3,248.50 [999.72, 5,497.27]	7.30 [2.25, 12.36]	462.36	7,161.34 [−1,748.29, 16,070.97]	15.49 [−3.78, 34.76]
20–30	640.78	2,920.62 [1,153.00, 4,688.24]	4.56 [1.80, 7.32]	663.15	2,861.65 [847.11, 4,876.20]	4.32 [1.28, 7.35]
30–40	764.16	1,782.24 [102.27, 3,462.20]	2.33 [0.13, 4.53]	785.02	2,145.27 [134.08, 4,156.45]	2.73 [0.17, 5.29]
40–50	869.07	2,095.34 [−404.53, 4,595.20]	2.41 [−0.47, 5.29]	901.44	1,225.10 [−2,600.50, 5,050.69]	1.36 [−2.88, 5.60]
50–60	1,000.79	2,745.07 [62.29, 5,427.85]	2.74 [0.06, 5.42]	1,061.92	3,036.64 [−192.24, 6,265.51]	2.86 [−0.18, 5.90]

TABLE I
CONTINUED

	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
60–70	1,105.77	4,517.27 [987.27, 8,047.26]	4.09 [0.89, 7.28]	1,183.69	6,421.21 [1,251.31, 11,591.12]	5.42 [1.06, 9.79]
70–80	1,174.29	4,119.87 [698.47, 7,541.26]	3.51 [0.59, 6.42]	1,250.96	4,168.65 [416.83, 7,920.47]	3.33 [0.33, 6.33]
80–90	1,302.59	2,661.52 [−2,108.13, 7,431.17]	2.04 [−1.62, 5.70]	1,435.94	1,412.17 [−7,132.02, 9,956.36]	0.98 [−4.97, 6.93]
90–99	1,757.80	11,054.78 [3,664.31, 18,445.26]	6.29 [2.08, 10.49]	2,199.03	16,031.50 [−4,391.24, 36,454.23]	7.29 [−2.00, 16.58]
99–99.9	3,027.57	13,945.91 [−7,774.43, 35,666.24]	4.61 [−2.57, 11.78]	6,663.11	−53,104.94 [−2.13e+05, 106,903.60]	−7.97 [−31.98, 16.04]
99.9+	3,567.96	14,653.04 [−13,843.91, 43,149.98]	4.11 [−3.88, 12.09]	13,576.52	−1,597,725.00 [−3.92e+06, 728,385.40]	−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 postaudit 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. Ninety-five percent confidence intervals are reported in square parentheses.

effects in the top 1%, the point estimates remain similar at the 99–99.9th percentiles and in the top 0.1%.[39]

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, introducing potential bias due to mean reversion: filers classified as high income in 2006 may have regressed further to the mean at a fixed future year of income measurement 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. The inclusion of years-relative-to-audit fixed effects in the DHTY specification recovers estimates broadly consistent with our results.

Next, we estimate the marginal return to IRS audits including our estimates of individual deterrence effects. Figure VII, Panel A takes our estimates of the initial return to operational audits (found in Figure V) and multiplies these by 3.2, our overall deterrence multiplier.[40]

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

40. 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 assumption is consistent with our results discussed above in Online Appendix Figure A.XIV 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. 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 reason to believe that their audit was triggered by content of their previous tax filings. Those subject to in-person audits may have larger

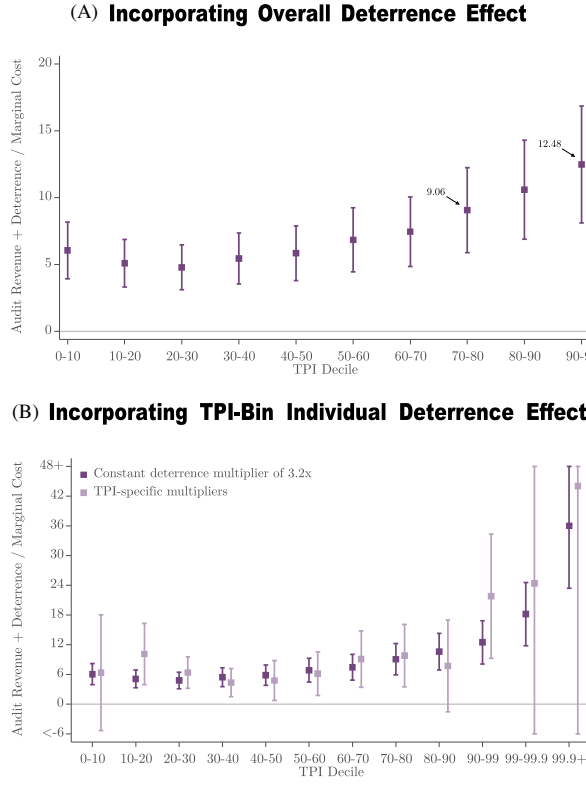


FIGURE VII

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

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 VI). The series plotted in darker purple uses the overall average deterrence multiplier from the single difference specification shown in Figure VI, Panel A. The series plotted in lighter purple uses the TPI bin-specific multipliers from a difference-in-differences specification shown in Figure VI, Panel B. 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 postaudit 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 nondirect 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 the IRS) from the total costs used to construct the average overhead multiplier.

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.[41]

Figure VII, Panel B shows how these estimates vary when using deterrence multipliers calculated in 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 6.29× for taxpayers at the 90–99th percentiles, we get a total return of 21.8:1. The same approach yields 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%.[42] Because of the greater precision of the estimates, we use the overall deterrence multiplier of 3.2, as in Panel A, in the welfare analysis that follows. Our results are similar when using TPI-bin-specific deterrence effects.

VI. Welfare Analysis of Marginal Audits

We find that audits, particularly those 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 trade-off between greater tax enforcement and alternative policies that raise revenue, such as higher tax rates?

In this section, we consider the welfare consequences of tax audits. In particular, we derive and estimate the marginal value of public funds (MVPF) of 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 the MVPF estimates to examine the desirability of expanded audits relative to other potential methods of raising revenue, such as changes in tax rates.

behavioral responses because they infer that similar filing practices will result in additional audits in the future.

41. If we assume that our estimated 3.2× 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.

42. We omit the top 1% from our primary results because we do not have the necessary power to precisely estimate the deterrence multiplier.

A key advantage of the MVPF framework is that it provides a transparent way to incorporate welfare weights that vary across the population.[43] 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 of the costs are borne by noncompliant taxpayers. 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 on audited individuals per dollar of revenue raised by the government. Formally, 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 nondistortionary tax would have an MVPF of 1. The revenue raised by the government is exactly equal to the individual beneficiary's willingness to pay to avoid the tax. If the MVPF of additional tax audits were 1.1, then those audits impose \$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 affected by a marginal audit. We discuss these considerations in further detail when comparing audits to other methods of raising revenue.

Online Appendix E provides a formal model derivation of the MVPF of individual tax audits in a wide class of dynamic models with audits, evasion, and deterrence effects. In the main text we focus on the intuition behind the MVPF components we derive from this model.

First, consider the numerator of the MVPF, which captures the willingness to pay to avoid the audit. There are two terms

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

in the numerator. 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 Online Appendix E, 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.

Second, the individual cost of being audited 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. This term 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 forgoing an additional audit. The denominator 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:

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

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, $\frac{R}{C}$ is the revenue raised per dollar of marginal spending and $\frac{B}{C}$ is the taxpayer burden per dollar of government cost. Intuitively, a higher value of $\frac{R}{C}$ pushes the MVPF of audits closer to 1. If the cost to conduct the audit is small compared with the revenue collected, the policy is closer to a nondistortionary tax. By contrast, if $\frac{R}{C}$ is low, then there are large distortionary costs associated with the audit, and so the MVPF is higher.

To estimate the MVPF of tax audits across the income distribution, we use the empirical estimates of revenues and costs constructed above. The remaining input into the MVPF is the taxpayer burden of audits, B.

VI.A. Burden of Audits

We provide new insights into the burden of audits using information from a 2023 survey conducted by the IRS.[44] The survey asked about the hours that taxpayers spent responding to the audit and any financial expenses 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.[45] While the survey measures capture only time and monetary costs taxpayers bear because of the audit, and not psychic burdens like stress, the survey provides a sense of important costs imposed on taxpayers.

We obtain the average value of these survey-derived burden estimates separately by TPI bin.[46] In each TPI bin we also obtain separate burden estimates for compliant and noncompliant taxpayers (i.e., those with and without assessed increases in tax liability due to the audit). This latter split is useful in the analysis to follow because it allows us to consider the possibility that welfare weights differ between those two groups.

Online Appendix Figure A.XVI, Panel A reports the average hours that taxpayers in each income group spent complying with audits. On average, individuals spend about 30 hours to comply with an in-person audit. The time burden of an audit varies from around 20 hours at the bottom of the income distribution to slightly below 40 hours in the 90th–99th TPI percentiles. Individuals in the top 1% spend just under 30 hours on average. Panel B reports the money spent by taxpayers across the income distribution to respond to the audit (e.g., spending on accountants and lawyers). Taxpayers in the bottom 90% of the income distribution

44. The survey examined a representative sample of taxpayers who faced an audit on their returns filed in 2019, 2020, and/or 2021 and whose audit was closed by 2022.

45. This procedure is roughly equivalent to dividing total year earnings by 2,000 (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. That work reports the average burden across all audited taxpayers and models the relationship between burden and income with a quadratic function.

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

spend an average of around $500, and taxpayers in the top 10% spend around $1,500–$2,000. Panel C uses a taxpayer-specific imputed wage to create a measure of the total monetized burden of an audit. Combining the monetary and time costs yields total audit burdens that rise from around $650 in the bottom half of the income distribution to $30,000 in the top 1%.[47]

When considering the welfare effects of expanded audits, one may be particularly concerned about the burden on taxpayers who are found to be compliant (those without additional assessed tax liability). In particular one may place a differential welfare weight on compliant versus noncompliant individuals. Many audited taxpayers are found to be compliant: the "no change" rate varies from around 40% in the bottom half of the income distribution to upward of 60% in the top 1% of the distribution.[48] In our analysis, we separately estimate the welfare cost of audits imposed on compliant and noncompliant taxpayers. Such an analysis requires burden information separately by compliance status. Online Appendix Figure A.XVIII shows the estimated burden imposed on individuals by compliance status. Broadly, we find similar but slightly higher burdens for compliant taxpayers.

VI.B. Marginal Value of Public Funds

We calculate the MVPF of audits across the income distribution. Figure VIII, Panel A reports the components of the MVPF of expanding audits of taxpayers in the 20–30th TPI percentiles. Above, we estimate that at the 20–30th percentiles an additional dollar spent on audits of such taxpayers generates $1.14 in upfront revenue and $3.63 in future revenue from deterrence effects, for total revenue of $4.78. 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). The ratio of willingness to

47. 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.

48. We note that assessed tax liability is an imperfect measure of compliance. The extent of noncompliance detected could potentially be a function of the intensity of the audit and the individual conducting the audit.

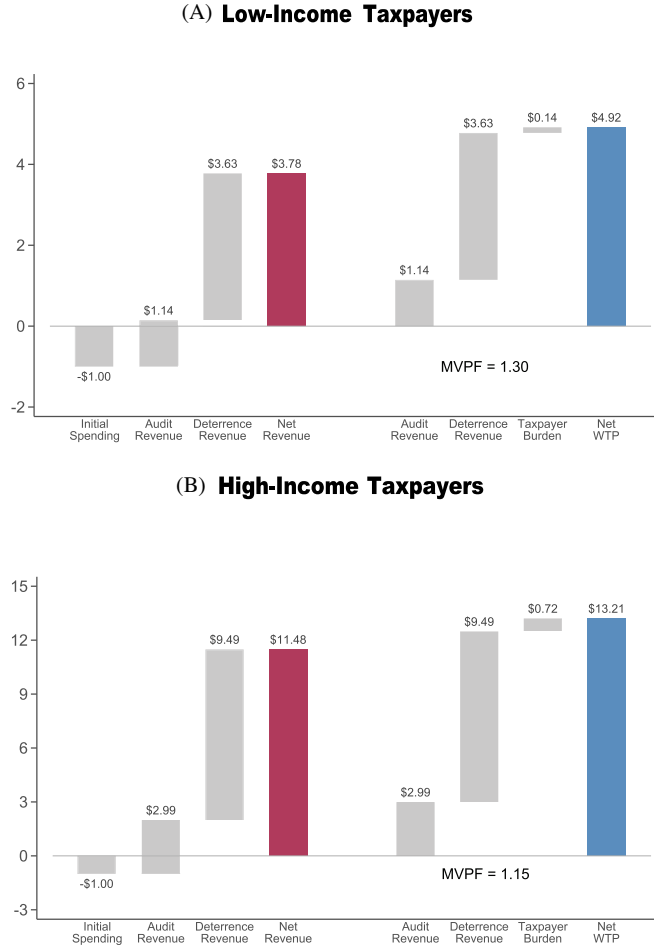


FIGURE VIII

The Marginal Value of Public Funds of Marginal Tax Audits

This figure presents the components of the MVPF of expanding audits among those 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 VII). The deterrence effect is calculated by multiplying direct revenue by the relevant deterrence multiplier shown in Figure VII, 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.

pay to net revenue yields an MVPF of 1.30. Figure VIII, Panel B repeats this calculation for expanding audits in the 90–99th TPI percentiles. We find an MVPF of 1.15, which is lower than the MVPF of audits at the 20–30th TPI percentiles.

Would it be welfare enhancing to audit more high-income individuals and fewer low-income individuals? The answer 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.[49] The left-hand side of equation (2) measures the social welfare cost from raising revenue through policy A and the right-hand side measures the social welfare gain from expending resources through policy B. Our MVPF estimates imply 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 1.13 times the welfare weight placed on audited low-income individuals.

VI.C. Comparison to Tax and Transfer Policies

One main benefit of the MVPF approach is that it enables welfare comparisons of policies across broad policy domains. For example, we consider how the MVPF of tax audits differs from the

49. See Saez and Stantcheva (2016) for a detailed discussion of the social marginal utility of income. Formally, let η_i 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 policy maker's social welfare function by η_i. The social welfare effect 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, $\eta^{audit} = \frac{\Sigma_i \eta_i (R_i + B_i)}{\Sigma_i (R_i + B_i)}$. Here, R_i and B_i is the financial and burden cost of the audit of individual i. 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 Online Appendix E.6

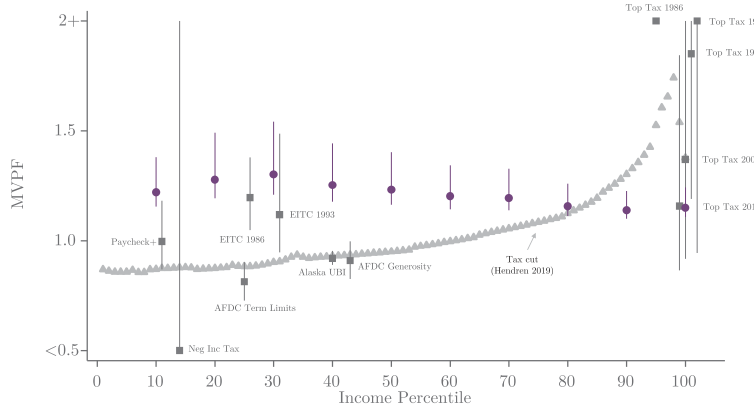


FIGURE IX

The MVPFs of Revenue-Raising and Transfer Policies, by Income

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 VII.

MVPFs 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 fiscal externality that reduces government revenue, so tax rate increases usually have MVPFs greater than one.[50]

Figure IX compares the MVPF of tax audits to a wide range of tax and transfer policies in the United States. The horizontal axis reports the average quantile in the income distribution of each policy's beneficiaries. The black squares report MVPF

50. By contrast, the deadweight loss associated with tax audits is driven by the compliance burden and administrative cost of conducting the audits.

estimates of tax and transfer policies analyzed in Hendren and Sprung-Keyser (2020), who draw on existing causal estimates to compute the MVPF of policies such as the 1993 EITC expansion and the 1986 top tax rate reduction. The gray triangles report estimates from Hendren (2020), who provides a microsimulation of 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 as estimated above.

Broadly speaking, the MVPFs of tax and transfer policies increase with income. The MVPFs of taxes and transfers targeting low-income individuals are near 1, 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, this 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.[51] Such a pattern could only be an optimum if

51. Our 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. To explore this possibility, Online Appendix Figure A.XIX examines audits of individuals who claimed the EITC. It examines the effect of the initial audit on future EITC claiming behavior separately for those individuals whose initial EITC claim was corrected downward (noncompliant) versus those whose initial EITC claim was not corrected

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 nonaudited taxpayers.

Thus far, our discussion has only considered a single average welfare weight placed on audited individuals at a given income level. The MVPF framework can also be used to examine the role of differential welfare weights on compliant and noncompliant taxpayers. In particular, we can use the separate burden estimates for compliant and noncompliant taxpayers reported in Online Appendix Figure A.XVIII. Combined with our measure of revenue collected for compliant and noncompliant taxpayers ($R = 0$ for the compliant), we can decompose the total MVPF into the incidence on those with and without additional assessed tax liability. This approach allows us to incorporate concerns for horizontal equity.

Consider, for example, the MVPF of expanding audits for taxpayers in the 90–99th income percentiles. The overall MVPF of 1.15 indicates that raising one dollar of revenue incurs total private welfare costs of \$1.15. When we decompose the \$1.15 welfare cost into the revenue raised from and burdens borne by noncompliant and compliant taxpayers in the 90th–99th income percentile, we find that raising \$1 from audits imposes private welfare costs of \$1.08 on noncompliant taxpayers and private welfare costs of \$0.07 on compliant taxpayers. If the social planner places a low welfare weight on noncompliant taxpayers, this suggests

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. However, we caution that this analysis does not rule out the potential for general deterrence effects on those not audited.

that audits of high-income taxpayers are a particularly efficient means of raising revenue.[52]

This basic conclusion holds even if the IRS survey data understate the true burden of audits on high-income taxpayers. Suppose that the social planner puts no weight on the private welfare costs that audits impose on noncompliant taxpayers. How large would the burden on compliant taxpayers need to be to overturn the conclusion that expanded audits raise revenue at lower welfare cost than increases in tax rates on top earners? As we detail in Online Appendix E.6, overturning that conclusion would require an average burden in excess of $300,000 on each compliant taxpayer that is audited. This hypothetical $300,000 burden is more than 10 times the audit burden derived from IRS survey data.

VII. Conclusion

In this article, we conduct a detailed analysis of the returns to tax audits across the income distribution. The average IRS audit of a tax return filed between 2010 and 2014 produces $2.17 in revenue for each dollar spent on the audit. Audit returns vary substantially across the income distribution. Audits of higher-income taxpayers take more auditor time and cost the IRS more than audits of lower-income taxpayers, but such audits 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 income distribution to $2.18 in the 90–99th percentiles and $6.29 in the top 0.1%.

We examine the return to marginal audits, exploiting the sharp decline in audit rates for returns filed between 2010 and 2014. We find that the revenue generated by 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

52. Suppose that the social planner placed no welfare weight on noncompliant taxpayers and placed the same weight on compliant taxpayers subject to audit and on the average. Audits of the top 1% then impose a social welfare cost of just $0.07 per dollar raised, far below the welfare cost of tax rates of around $1.50 found in the literature.

audits. We find that audits generate a persistent increase in tax revenue collected from those who are audited. Measured over 14 years, the future revenue collected is 3.2 times the return to the initial audit. 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.

Last, we use the MVPF framework to consider the welfare implications of our 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. This declining pattern of MVPFs in income suggests that current audit levels across the income distribution are inconsistent with canonical social preferences.

There is ample room for future work to build on 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 returns filed in 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. Although our work captures audits of individual income generated by businesses, future work should further examine the returns to auditing businesses themselves.

U.S. Department of the Treasury, United States
Massachusetts Institute of Technology and Policy Impacts, United States
University of Pennsylvania and Policy Impacts, United States
University of Sydney, Australia and Policy Impacts, United States

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

Data Availability

The data underlying this article are available in the Harvard Dataverse, https://doi.org/10.7910/DVN/SVQASQ (Boning et al. 2024).

References

Arun, Advani, William Elming, and Jonathan Shaw, "The Dynamic Effects of Tax Audits," *Review of Economics and Statistics*, 105 (2023), 545–561. https://doi.org/10.1162/rest_a_01101.

Allingham, Michael G., and Agnar Sandmo, "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 3 (1972), 323–338. https://doi.org/10.1016/0047-2727(72)90010-2.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman, "Tax Evasion and Inequality," *American Economic Review*, 109 (2019), 2073–2103. https://doi.org/10.1257/aer.20172043.

Beer, Sebastian, Matthias Kasper, Erich Kirchler, and Brian Erard, "Do Audits Deter or Provoke Future Tax Noncompliance? Evidence on Self-Employed Taxpayers," *CESifo Economic Studies*, 66 (2020), 248–264. https://doi.org/10.1093/cesifo/ifz018.

Best, Michael, Jawad Shah, and Mazhar Waseem, "Detection without Deterrence: Long-run Effects of Tax Audit on Firm Behavior," Mimeo, University of Manchester, (2021).

Bjørneby, Marie, Annette Alstadsæter, and Kjetil Telle, "Limits to Third-Party Reporting: Evidence from a Randomized Field Experiment in Norway," *Journal of Public Economics*, 203 (2021), 104512. https://doi.org/10.1016/j.jpubeco.2021.104512.

Boning, William C., Nathaniel Hendren, Ben Sprung-Keyser, and Ellen Stuart, "Replication Data for: 'A Welfare Analysis of Tax Audits across the Income Distribution'," (2024), Harvard Dataverse. https://doi.org/10.7910/DVN/SVQASQ.

CBO, "Options for Reducing the Deficit: 2019 to 2028," *Congressional Budget Office*, 2018.

———, "Trends in the Internal Revenue Service's Funding and Enforcement," *Congressional Budget Office*, (2020).

DeBacker, Jason, Bradley Heim, Anh Tran, and Alexander Yuskavage, "Once Bitten, Twice Shy? The Lasting Impact of Enforcement on Tax Compliance," *Journal of Law and Economics*, 61 (2018), 1–35. https://doi.org/10.1086/697683.

———, "Tax Noncompliance and Measures of Income Inequality," *Tax Notes*, (2020). https://www.taxnotes.com/tax-notes-state/compliance/tax-noncompliance-and-measures-income-inequality/2020/02/17/2c5p7.

Elzayn, Hadi, Evelyn Smith, Thomas Hertz, Arun Ramesh, Robin Fisher, Daniel E. Ho, and Jacob Goldin, "Measuring and Mitigating Racial Disparities in Tax Audits," *Quarterly Journal of Economics*, 140 (2025), 113–163. https://doi.org/10.1093/qje/qjae027.

Guyton, John, and Ronald Hodge, "The Compliance Costs of IRS Post-Filing Processes," paper presented at the *An IRS-TPC Research Conference: Advancing Tax Administration*, 2014.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman, "Tax Evasion at the Top of the Income Distribution: Theory and Evidence," NBER Working Paper no. 28542, (2021). https://doi.org/10.3386/w28542.

Hebous, Shafik, Jia Zhiyang, Løyland Knut, Thoresen Thor Olav, and Øvrum Arnstein, "Do Audits Improve Future tax Compliance in the Absence of Penalties? Evidence from Random Audits in Norway," *Journal of Economic Behavior & Organization*, 207 (2023), 305–326. https://doi.org/10.1016/j.jebo.2023.01.001.

Hendren, Nathaniel, "Measuring Economic Efficiency Using Inverse-Optimum Weights," *Journal of Public Economics*, 187 (2020), 104198. https://doi.org/10.1016/j.jpubeco.2020.104198.

Hendren, Nathaniel, and Ben Sprung-Keyser, "A Unified Welfare Analysis of Government Policies," *Quarterly Journal of Economics*, 135 (2020), 1209–1318. https://doi.org/10.1093/qje/qjaa006.

Hodge, Ronald H., Alan H. Plumley, Kyle Richison, Getaneh Yismaw, Nicole Misek, Matt Olson, and H. Sanith Wijesinghe, "Estimating Marginal Revenue/Cost Curves for Correspondence Audits," *IRS Research Bulletin*, (2015).

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

———, "Effects of Recent Reductions in the Internal Revenue Service's Appropriations on Returns on Investment," *IRS Research Bulletin: Proceedings of the 2020 IRS/TPC Research Conference*, (2020).

IRS, "Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2014–2016," Technical Report, Applied Analytics & Statistics, Internal Revenue Service, (2022).

Johns, Andrew, and Joel Slemrod, "The Distribution of Income Tax Noncompliance," *National Tax Journal*, 63 (2010), 397–418. https://doi.org/10.17310/ntj.2010.3.01.

Kaplow, Louis, "Optimal Taxation with Costly Enforcement and Evasion," *Journal of Public Economics*, 43 (1990), 221–236. https://doi.org/10.1016/0047-2727(90)90031-C.

Kasper, Matthias, and James Alm, "Audits, Audit Effectiveness, and Post-Audit Tax Compliance," *Journal of Economic Behavior & Organization*, 195 (2022a), 87–102. https://doi.org/10.1016/j.jebo.2022.01.003.

———, "Does the Bomb Crater Effect Really Exist? Evidence from the Laboratory," *FinanzArchiv* 78, (2022b), 87–111. https://doi.org/10.1628/fa-2021-0014.

Keen, Michael, and Joel Slemrod, "Optimal Tax Administration," *Journal of Public Economics*, 152 (2017), 133–142. https://doi.org/10.1016/j.jpubeco.2017.04.006.

Kleven, Henrik J., Martin B. Knudsen, Claus T. Kreiner, Søren Pedersen, and Emmanuel Saez, "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark," *Econometrica* 79 (2011), 651–692. https://doi.org/10.3982/ECTA9113.

Kleven, Henrik J., and Claus T. Kreiner, "The Marginal Cost of Public Funds: Hours of Work versus Labor Force Participation," *Journal of Public Economics*, 90 (2006), 1955–1973. https://doi.org/10.1016/j.jpubeco.2006.03.006.

Løyland, Knut, Oddbjørn Raaum, Torsvik Gaute, and Øvrum Arenstein, "Evaluating Compliance Gains of Expanding Tax Enforcement," *Economica*, 91 (2024), 142−162. https://doi.org/10.1111/ecca.12499.

Mayshar, Joram, "On Measures of Excess Burden and Their Application," *Journal of Public Economics*, 43 (1990), 263–289. https://doi.org/10.1016/0047-2727(90)90001-X.

———, "Taxation with Costly Administration," *Scandinavian Journal of Economics*, 93 (1991) 75–88. https://doi.org/10.2307/3440422.

Mazzolini, Gabriele, Laura Pagani, and Alessandro Santoro, "The Deterrence Effect of Real-World Operational Tax Audits on Self-Employed Taxpayers: Evidence from Italy," *International Tax and Public Finance*, 29 (2022), 1014–1046. https://doi.org/10.1007/s10797-021-09707-9.

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68 (2001), 205–229. https://doi.org/10.1111/1467-937X.00166.

Saez, Emmanuel, and Stefanie Stantcheva, "Generalized Social Marginal Welfare Weights for Optimal Tax Theory," *American Economic Review*, 106 (2016), 24–45. https://doi.org/10.1257/aer.20141362.

Sarin, Natasha, and Lawrence H. Summers, "Shrinking the Tax Gap: Approaches and Revenue Potential," NBER Working Paper no. 26475, 2019. https://doi.org/10.3386/w26475.

———, "Understanding the Revenue Potential of Tax Compliance Investment," NBER Working Paper no. 27571, (2020). https://doi.org/10.3386/w27571.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian, "Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota," *Journal of Public Economics*, 79 (2001), 455–483. https://doi.org/10.1016/S0047-2727(99)00107-3.

Slemrod, Joel, and Shlomo Yitzhaki, "Tax Avoidance, Evasion, and Administration," in *Handbook of Public Economics*, vol. 3, Alan J. Auerbach and Martin Feldstein, eds. (Amsterdam: North-Holland, 2002), 1423–1470. https://doi.org/10.1016/S1573-4420(02)80026-X.