NBER WORKING PAPER SERIES

A WELFARE ANALYSIS OF TAX AUDITS ACROSS THE INCOME DISTRIBUTION

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

Working Paper 31376
http://www.nber.org/papers/w31376

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2023

We gratefully acknowledge research assistance from Jamie Emery. We are also grateful for feedback from conference and seminar participants at the National Tax Association, the Office of Tax Analysis, the 2022 and 2023 ASSA meetings, and the NBER TAPES 2022. This research was supported by Policy Impacts, which receives funding from the Spiegel Family Fund. This research was conducted while an author was an employee at the U.S. Department of the Treasury. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury. Any taxpayer data used in this research was kept in a secured Treasury or IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

A Welfare Analysis of Tax Audits Across the Income Distribution
William C. Boning, Nathaniel Hendren, Ben Sprung-Keyser, and Ellen Stuart
NBER Working Paper No. 31376
June 2023
JEL No. H0

ABSTRACT

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

William C. Boning
Office of Tax Analysis
Department of the Treasury 1500
Pennsylvania Avenue, NW
Washington, DC 20220
William.Boning@treasury.gov

Nathaniel Hendren
Harvard University
Department of Economics
Littauer Center Room 230
Cambridge, MA 02138
and NBER
nhendren@gmail.com

Ben Sprung-Keyser
Harvard University
Department of Economics
Cambridge, MA 02138
bsprungkeyser@g.harvard.edu

Ellen Stuart
University of Sydney
School of Economics
Sydney NSW 2026
Australia
ellen.stuart@sydney.edu.au

1 Introduction

IRS estimates suggest that more than $500 billion in tax obligations go unpaid each year (IRS, 2022). Those unpaid tax liabilities are concentrated among taxpayers at the top of the income distribution. For example, the top 10% of earners may owe more than 60% of all unpaid tax liabilities (DeBacker et al., 2020; Guyton et al., 2021; Johns and Slemrod, 2010).

Evidence suggests that, on average, tax audits can recoup unpaid taxes and raise revenue in the process. There is, however, very limited evidence on the returns to tax audits across the income distribution (Holtzblatt and McGuire, 2016, 2020).

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

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

We combine these measures of audit costs and audit revenue to estimate the average return to

[1] We restrict our analysis to focus on in-person audits rather than audits conducted via correspondence. This is because correspondence audits have a fundamentally different cost structure and are narrower in scope. They generally focus on a restricted set of issues, such as EITC compliance. In-person audits are generally more intensive and broader in their scope, examining a wide set of potential compliance concerns.

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

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

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

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

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

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

We show that audits lead to an increase in future taxes paid that persists over the 14 years we observe in the data. In present discounted value, these additional taxes are 3.2 times the revenue raised from their initial audit. We also show that these deterrence effects are similar across the full income distribution (although we lose the power necessary to estimate precise effects in the top 1%).[3] When we combine these deterrence figures with our estimates of the direct returns to marginal audits, we find that auditing individuals in the bottom half of the income distribution produces a return slightly above 5:1. By contrast, audits of individuals in the 90–99th percentiles produce a return of nearly 13:1.[4]

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

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

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

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

per dollar spent on tax audits, but also include the burden imposed on taxpayers during an audit. We draw upon taxpayer burden estimates from an IRS survey analyzed by Guyton and Hodge (2014).

With these inputs, we estimate that additional audits of taxpayers in the bottom half of the income distribution impose a welfare cost of $1.40 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 $1.13 per dollar of net government revenue raised. This pattern lies in stark contrast to MVPF estimates for tax changes across the income distribution, where tax increases of high-income earners typically have much larger MVPFs than tax increases on low-income earners. While determinations of policy optimality depend on the welfare weights placed on taxed or audited individuals, this suggests that increasing audit rates raises revenue at relatively low welfare cost.

Related literature As noted above, a recent body of work calculates the average returns to audits (CBO, 2018, 2020; Holtzblatt and McGuire, 2016, 2020). There is also a robust literature on the size of the tax gap across the income distribution (DeBacker et al., 2020; Guyton et al., 2021; Johns and Slemrod, 2010). There is, however, comparatively little empirical work quantifying the costs and benefits of tax enforcement across the income distribution. Sarin and Summers (2019) provide a figure for the returns to auditing taxpayers with more than $5 million in income, but their analysis is meant simply as an illustrative calculation rather than a full accounting of costs and revenues across the income distribution.[5] This paper seeks to fill this gap by providing detailed estimates on the returns to audits across the income distribution.

In estimating the total revenue raised by individual audits, we extend upon an existing literature measuring the individual deterrence effects of audits. Our core contribution here is to identify the long-run effects of audits and study how deterrence effects vary across the income distribution. We build most closely on DeBacker et al. (2018) by extending our analysis to additional years of data and using a matching strategy to allow comparisons within fine-grained income groups.[6] Our work also relates to large body of evidence on deterrence measured outside the United States (Bjørneby et al., 2021; Hebous et al., 2020; Mazzolini et al., 2022; Kasper and Alm, 2022a,b; Kleven et al.,

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

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

2011; Advani et al., 2021; Best et al., 2021). For example, Kleven et al. (2011) estimate deterrence effects for a single post-audit study year in Denmark and Advani et al. (2021) estimate deterrence effects in the UK up to 8-years post-audit. These papers find yearly deterrence effects between 20% and 35% of initial audit revenue, which are relatively consistent with the magnitudes estimated in this paper, which we find to persist for 14 years.

Our paper also relates to the large literature on the distortionary effects of raising tax revenue. There is an extensive theoretical and empirical body of work quantifying the costs of raising revenue through changes in tax rates. This literature highlights the importance of analyzing heterogeneity in the distortionary cost of taxation across the income distribution (E.g. Saez 2001; Kleven and Kreiner 2006).[7] 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 across the income distribution allows for the estimation of those distortionary effects. Our discussion of the MVPF of tax audits also helps to formalize the parallel with tax rates.

Relatedly, our MVPF approach connects to a theoretical literature on optimal tax administration (see, e.g., Mayshar, 1991; Slemrod and Yitzhaki, 2002). For example, recent work by Keen and Slemrod (2017) develops a general model of optimal tax administration. They show that, at an optimum, the marginal costs and benefits of enforcement should be equated both to each other, and to the marginal costs and benefits of changes in tax rates. Our MVPF approach provides a way of empirically operationalizing this idea by comparing the MVPFs of tax audits to those of modifications to the tax schedule.[8]

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

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

[8] The MVPF approach also generalizes the framework developed in Keen and Slemrod (2017) by allowing for social preferences to differ based on whether revenue is raised via tax audit or higher tax rates.

2 Data and Sample

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

2.1 Audit Overview

In this paper, we focus on in-person audits of individuals.[9] Appendix Figure A1 provides a flow chart that maps out the audit process. The audit process begins when a tax return is selected for audit and designated as a field audit or an office audit. This determination is made based on the expected complexity of the audit. In general, office audits are conducted by Tax Compliance Officers and tend to involve a lower-complexity mix of issues. These audits often involve an interview at an IRS office. Field audits are conducted by Revenue Agents and involve more complex cases. These audits often involve an interview at the taxpayer's home or place of business.[10] In either case, the exam stage begins when an IRS examiner reviews the taxpayer's relevant documents and meets face-to-face with the taxpayer to assess whether they have additional tax liability. This process may involve reviewing documents beyond what was included in the tax return, such as receipts that verify the validity of deductions or bank account records that validate all reported income. In some cases, an examination will expand to include returns filed in additional tax years. For the purposes of our analysis, we define an "audit" to include both the evaluation of an initial tax return and the evaluation of additional returns from other tax years that are triggered by the initial examination.[11]

If the examiner determines that the taxpayer has additional tax liability, the next phase of

[9]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 (i.e. form 1040). Finally, this means that we only provide a limited discussion of correspondence audits. Correspondence audits are relatively low-complexity and are conducted primarily via mail without assignment of a Tax Compliance Officer or Revenue Agent. Our estimate of the average return to correspondence audits can be found in Appendix Figure A2. A more detailed discussion of the structure of the IRS can be found in Data Appendix A.2.

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

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

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

Once a taxpayer pays their full tax liability, the audit process is completed. If the taxpayer does not pay the full amount, the case is sent to collections. The collections process starts with the IRS sending notification letters to the taxpayer indicating that they have an unpaid balance. If the taxpayer does not respond to the notifications, the case is handled by the Automated Collection System (ACS) or a local field office. If the case is sent to ACS, ACS personnel will try to contact the taxpayer and work with the taxpayer to find a payment solution.[12] 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.[13]

Most of the analysis in this paper focuses on in-person audits which are selected for review on the basis of suspected non-compliance. In our analysis of deterrence effects, we also make use of audits conducted as part of the IRS' National Research Program (NRP). The NRP is designed to provide critical information on tax compliance and aid in the IRS' estimate of the tax gap.[14] In order to achieve those goals, the IRS selects a small random sample of tax returns and Revenue Agents examine all areas of the tax return. While the audit exam is slightly more intensive than a typical in-person audit, NRP audits follow the same basic steps outlined in Appendix Figure A1.

2.2 Data

Our data is drawn from two internal IRS sources. First, we use audit-level data that contains detailed information on IRS enforcement activities. The data includes the time spent and activities performed by IRS enforcement personnel in the US. It also includes information on the revenues raised from each stage of each audit. Second, we use internal accounting data for the audit divisions

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

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

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

of the IRS. This provides information on costs other than the direct hourly wage costs for the auditors (E.g., employee benefits, management, rent, IT, etc.). We begin with a description of the audit-level enforcement data and then discuss the internal IRS accounting data. A more detailed discussion of the data can be found in Appendix A.

Audit-Level Enforcement Data. We observe data on IRS enforcement activities from fiscal years 2003-2021 (which run from October 1, 2002 to September 30, 2021). The enforcement database includes comprehensive information on the revenue collected via the audit process. For each audit conducted, it reports the tax liability assessed at each stage of the audit process, as well as the amount of tax liability collected (including penalties and interest payments). These data allow us to separately estimate the revenue raised at the exam, appeals, and collection stages. On the cost side, the database includes a detailed log of the hours spent on direct enforcement activity by IRS employees on each audit. It also includes the government pay grade (GS grade) of the associated IRS personnel. We translate this into wage costs using a location-specific GS hourly pay-scale[15] and the zipcode of the taxpayer under audit. Multiplying personnel hours by the hourly wage yields the direct labor costs accrued at each stage of the audit process.[16]

Internal IRS Accounting Data. The cost of conducting audits goes beyond the time cost of direct enforcement activity. In order to identify and incorporate these costs, we utilize detailed internal IRS business unit accounting information from fiscal years 2011-2020.[17] 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.

[15]Pay may vary within a single pay grade as employees also have a step within each paygrade that helps determines their pay. Following the approach in Holtzblatt and McGuire (2020), we use the median rate of pay (see step 5 out of 10 in Appendix A).

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

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

[17]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.

We then combine these primary categories: (1) non-direct labor-related costs, (2) organization-wide costs, and (3) central IRS management overhead. Non-direct labor-related costs include time spent by auditors on non-auditing activities (such as training), wage costs for management and support staff, and non-wage costs for employees such as fringe benefits and workers' comp. Organization-wide costs include items such as building costs and information technology costs. Finally, central IRS management overhead includes costs incurred inside the IRS but outside the primary business unit conducting individual audits. It also includes costs performed by other government agencies. A detailed accounting of these sub-components can be found in Appendix A.2.

We calculate direct wage costs separately for the exam, appeals, and collection stages of the audit. For the exam stage, we are able to estimate cost multipliers for each major component of overhead: non-direct labor-related costs, organization-wide costs, and general overhead costs. We allocate these to each audit in proportion to audit's direct wage costs.[18] In Appendix Figure A3, we show the robustness of our results to this specific method of allocating costs.[19] As noted below, these exam costs make up 93% of total audit costs. As for the other audit costs, we apply a per-notice cost for cases that went to collections, a "cost-per-dollar-raised" multiplier for cases spending time in ACS, and an overhead cost per dollar of direct labor cost for audits in field collections. Appendix A provides further details on these costs.

2.3 Sample

We focus our analysis on audits of individual tax returns filed in tax years for which (a) we observe comprehensive measures of costs and (b) sufficient time has passed that nearly all audits have been completed. This leads us to study the universe of audits of individual tax returns from 2010-2014.[20] We begin with 2010 because tax year 2010 audits take place beginning in 2011, the first year for which we have internal IRS accounting data. We end our sample with the 2014 tax year because this provides us at least 7 years of follow-up to observe the costs and revenues associated with an audit.

[18] The intuition here is that costs such as non-audit labor costs and fringe benefits should rise with the amount of expenditure on a given audit. In practice, this means that more costs are allocated to audits of higher income taxpayers due to both the greater quantity of time involved in those audits and the higher pay grade of the auditors. This assumption is intended to be relatively conservative one given that the results to follow show higher returns on audits of high-income taxpayers.

[19] In particular, we allocate costs in proportion to labor hours rather than audit costs and show the results are nearly indistinguishable. We could also allocate overhead costs on a per-audit basis, which, relative to our primary specification, would result in lower costs for audits of high-income taxpayers. This approach would only further reinforce our core conclusion regarding the higher returns to audits at the top of the income distribution.

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

With these definition, our baseline sample includes approximately 710,000 in-person audits.[21]

Appendix Figure A4 uses data from tax year 2003 to report the revenues that are obtained in each year after the audit. The figure illustrates that almost all revenue from audits is collected within 15 years. Of this, 86% of revenue is obtained within 7 years and 96% of revenue is obtained within 11 years.[22] We use these patterns in our results to impute any remaining revenue we expect to be collected from audits of 2010-2014 tax years.

For each audit, we link the taxpayer to their Form 1040 in the tax year that initiated the audit. We use this information to study the heterogeneity of revenue and costs as a function of the individual's income. Our income measure, total positive income (TPI), is the sum of the various positive income items reported on the return. This is the primary measure used by the IRS to categorize returns by income.[23] For all income values, we deflate to constant 2016 dollars using the CPI-U-RS.

3 Average Revenue and Costs per Audit

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

Average total costs. Starting with the exam stage, the average in-person audit takes roughly 28.7 hours and is conducted by auditors earning about $38.17 per hour. This corresponds to an average cost of $1,097 due to the direct labor hours of auditors conducting exams.

In order to measure the total cost of these hours spent on the audit, we then need to allocate the overhead cost associated with that auditing work. The first component of overhead is additional labor costs not directly allocated toward an audit. This includes wage costs for non-auditing hours, fringe benefits, training costs, and manager labor costs. We allocate these costs in proportion to wage costs on each audit, and find they are approximately $2,115 per audit, roughly double the direct labor hours costs. Next, we incorporate overhead associated with organization-wide costs. This includes costs like building rent and information technology. When we allocate those

[21] We also study the universe of 4.2 million correspondence audits between 2010 and 2014 and 63,174 NRP audits between 2006-2014 (as discussed in Section 5, we expand to tax years 2006-2014 to improve precision and the length of the post audit period for estimating deterrence effects on future tax revenue).

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

[23] This measure excludes losses, which avoids concerns about the accuracy of reported losses. Consequently, this approach also treats an individual as high-income if they have both high levels of positive income and large losses. Such an individual is likely to be high-income over the long-run, even if they have a low adjusted gross income (AGI) in a given year.

proportionally, they lead to an additional cost of $1,103 per audit in the exam stage. Finally, we incorporate the overhead costs associated with central IRS management. Those result in an additional $1,593 per audit. When combined, this suggests the $1,097 in direct cost of hourly labor is associated with an additional $4,811 in overhead costs, for a total of $5,907.

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

After the exam stage, some taxpayers appeal the results of their audit. We estimate that the appeals and tax court stage increases the average cost of an audit by $170. This combines $108 in direct labor hours costs and $62 in additional costs.[24] Additionally, many cases end up in collections, which we estimate leads to an additional $191 in costs.[25] Finally, as noted above, we only observe costs that are accrued in a 7-11 year window after the year a tax return is filed. While nearly all costs accrue within this window, there may be some that are incurred afterward.[26] In order to account for those potential costs, we plot the trajectory of costs accrued for returns filed in the 2003 tax year. Tax year 2003 lies before our primary sample frame, but provides us 17 years of follow up data.[27] We estimate that in our 7-11 year sample frame we capture 96-99% of total labor hours spent on audits.[28] The results of this projection exercise can be seen in Appendix Figure A4 panel B. Our forecast suggests that we do not observe $150 in additional future spending on each audit. When we incorporate this estimate, we get a total average cost of $6,418 for each in-person audit.

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

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

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

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

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

Average total revenue. How much revenue do audits generate? Figure 1 shows that in-person audits collect an average of $6,194 during the exam stage, $617 during the appeals stage, and an additional $6,259 through collections. These estimates incorporate all revenue collected through 2021. As is the case with our cost estimates, we use a projection exercise to estimate the magnitude of revenue collected outside our observed 7-11 year window. We estimate that 86-96% of all revenue is collected within 7-11 years after filing. This corresponds to additional revenue of $1,212 per audit. The results of the projection can be seen in Appendix Figure A4 panel A.[29]

Putting together results from each stage of the audit, we estimate average total revenue is $14,283 per audit and average costs are $6,418. However, the costs of the audit are, on average, incurred before the revenue is obtained. Appendix Figure A4 demonstrates that average revenues lag average costs by approximately 1 year. As a result, when we apply a 3% discount rate to align the time periods of cost and revenue, the return to an average audit is reduced by $353.[30] This implies total revenue of $13,930, which is 2.17 times higher than the cost.

Heterogeneity by Income Next, we analyze how revenues and costs vary when conducting audits across the income distribution. Figure 2A presents the average revenues and costs for in-person audits separately by percentiles of taxpayer total positive income (TPI). We split income on the horizontal axis into groupings of 5 percentiles, and we break out the top 5% into the 95–99th percentiles, 99–99.9th percentiles, and the top 0.1%. The red series plots the average costs and the blue series plots average revenues associated with each audit.

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 percentile, $6,863 in the 90–99th percentiles and $15,170 in the top 0.1%. Appendix Figure A5 shows that this is primarily because audits of higher income taxpayers are more time intensive than audits of lower income taxpayers. For example, auditing taxpayers in the 70–80th percentiles requires an average of 27.8 hours of auditor time while auditing taxpayers in the top 0.1 percent requires 64.6 hours of auditor time.[31]

[29] Appendix Figure A4 panels A and B presents results separately by income (TPI) bin. This allows for heterogeneity in the repayment patterns by income bin.

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

[31] Audits of high-income taxpayers are also conducted by slightly more experienced auditors. Those auditors receive higher pay, and so there is some variation in the average hourly costs of audits across the income distribution. For example, as shown in Appendix Figure A5, audits of taxpayers at the 70–80th income percentiles are conducted by auditors earning an average of $36.86 per hour. Audits of taxpayers at the top 0.1 percentile are conducted by auditors earning an average of $45.66 per hour.

While audit costs rise with taxpayer income, the same is true for audit revenue. In fact, the additional costs stemming from the complexity associated with auditing high-income taxpayers are more than offset by additional revenues. Auditing taxpayers in the bottom half of the income distribution produces an average of $4,984 in revenue. An audit of a taxpayer in the 70–80th percentiles produces $8,270 in revenue, while an audit of a taxpayer in the top 90–99th percentiles yields an average of around $14,973, and an audit of a taxpayer in the top 0.1% yields an average of approximately $95,491.

Figure 2B divides the average revenue by average cost to show the average returns to audits across the income distribution. The 2:1 average return across the full population varies considerably with income. In the bottom half of the income distribution each dollar of audit costs produces an average of $0.96 in revenue. Audits in the 70–80th percentiles return $1.58 for each dollar in auditing costs. Audits in the 90–99th percentiles return $2.18 for every $1, audits in the top 0.1% return $6.29.[32]

4 Marginal Audits

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

The revenue and costs associated with marginal audits may differ from the average revenues and average costs of audits for two key reasons. First, if the audit selection process seeks to maximize revenue per dollar of audit cost, then there may be diminishing marginal revenues associated with additional audits. Second, economies of scale mean that some costs – particularly costs other than direct labor costs – may not increase linearly in audit hours. This section explores the consequences of these potential forces and examines how they impact returns to marginal audit expansions.

(Lack of) Diminishing Marginal Revenue We study the return to marginal audits by exploiting the steep audit rate decline for returns filed between the 2010 and 2014[33] tax years. The IRS

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

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

audited 0.92% of all returns from the 2010 tax year, but that rate fell to 0.56% for 2014 tax returns. As shown in Figure 3A, this represents a 40% decline in the overall audit rate.[34] We use this reduction of audit rates to understand the revenue associated with marginal audits. We conceptualize marginal audits as those that would have been conducted at 2010 levels of audit intensity but not conducted at 2014 levels. This exercise helps us shed light on the potential returns to expanding audit expenditures back to their 2010 levels.[35]

The return to reversing recent audit declines depends, in part, on the audit selection process used by the IRS. If audit reductions between 2010 and 2014 prioritized cutting audits with low revenue per unit cost, then audit expansions would have lower returns than average audits.[36] By contrast, if the audit selection process prioritized other criteria – fulfilling statutory requirements, minimizing the fraction of audits that result in no change to tax liability, etc – then the returns to marginal audits may not differ from the returns to average audits.

Figure 3A displays the average revenue per audit and average direct wage cost of each audit between 2010 and 2014. Despite the sharp decline in audits in this period, the revenue and cost figures remain essentially unchanged. In 2010, revenues were 11.0 times direct wage costs. In 2014, revenues were 11.2 times direct wage costs.[37] This suggests that the types of audits cut between 2010 and 2014 had similar returns to the audits that were still conducted in 2014. In other words, the return to marginal audits was the same as the return to average audits.[38]

That basic pattern is further validated when examining changes in audit patterns across the income distribution. In particular, we find both that the reduction in audit expenditures was relatively consistent across the full income distribution and that the returns to audits did not

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

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

[36]In this context, the phrase "return" is meant to capture revenue per dollar of direct wage cost. We discuss the distinction between average versus marginal overhead in the next sub-section.

[37]These figures are reported in terms of direct wage costs rather than total costs. As noted previously, this is done to set aside any discussion about changes to overhead costs over time. We discuss marginal overhead costs in the next sub-section.

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

meaningfully change at any income level. Figure 4A reports the change in the total direct wage costs of audits at each 5-percent TPI bin (and several high-income TPI bins) between 2010 and 2014. Audit expenditures fell between 31% and 42% for the bottom half of the income distribution and fell slightly more in the top decile, with reductions between 46% and 61%.[39] This consistent decline in audits across the income distribution helps explain why the average return to audits did not rise between 2010 and 2014, even though the average return to audits varies across the income distribution. Audit reductions were not concentrated among income groups with low average audit returns. Figure 4B 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.[40] We find a slope close to 1 (1.05 se 0.03) and intercept close to 0 (0.21 se 0.34), showing that, across TPI bins, there is no systematic change in the returns to audits despite significant declines in audit rates.

One potential concern with this conclusion is that there may have been an underlying time trend in the returns to a typical audit between 2010 and 2014. For example, it could be the case that tax evasion fell substantially between 2010 and 2014. That pattern alone would produce a reduction in the returns to audits. This could mean that changes in the audit selection process increased the average return to audits between 2010-2014, but that this increase was masked by a compositional shift in the amount of tax evasion in the first place. We are, however, able to mitigate this concern. We do so by drawing upon audit data from the National Research Program (NRP). Each year, the NRP randomly selects a series of returns for review. As noted above, NRP audits are more intensive than operational audits because they are designed to measure tax compliance. If tax evasion declined over time, then NRP audits would detect less evasion in later years. Appendix Figure A8 panel A plots the cost of and revenue from NRP returns over time. There is no systematic decline in the return to NRP audits. Instead, the return to NRP audits rises slightly. Costs remain consistent over time and the revenue raised rises slightly between 2010 and 2014. Appendix Figure A9 repeats this analysis separately by income bin. It shows there is no systematic decline in NRP returns within any income group.

Taken together, this evidence suggests that marginal audits cut between 2010 and 2014 produced similar revenue (per dollar of auditor wages) to the audits that remained. While estimating the IRS'

[39] Here, we report declines in audit expenditures rather than declines in audit rates. This is meant 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.

[40] Appendix Figure A7 reports the analogous figure for 2010 and 2014 revenue per audit by TPI bin.

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

These results imply that expanding audit rates back to 2010 levels would not diminish the average revenue collected per audit. In other words, for substantial audit expansions, marginal revenue closely resembles average revenue.

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

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

As noted previously, the costs of an audit beyond the direct wage costs fall into three broad categories: labor/fringe, organization-wide costs, and central overhead/headquarters costs. 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.[41]

Those results are broadly consistent with, but slightly more conservative than, estimates from existing IRS budget requests. For example, the IRS has published budget estimates from audit expansions as part of the Program Integrity Allocation Adjustment. Total marginal non-labor

[41]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).

costs in those requests are approximately 23.6%[42] as large as total labor and fringe costs.[43] By comparison, our baseline estimates assume that total marginal non-labor costs are 25.6% as large as total labor and fringe costs.

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

Putting together these results on marginal revenues and marginal costs, Figure 5 reports our estimates for the marginal returns to expanding audits from their 2014 to 2010 levels. We start by combining the average costs and average revenue of audits across the income distribution from Figure 2. Figure 5 then adjusts the average costs downward because the marginal costs fall below the average costs. There is no adjustment made to marginal revenue because we find that restoring audits to 2010 levels would not diminish the marginal returns. The figure reports the returns to marginal audits in each income group. The return to a dollar of spending on marginal audits in the bottom half of the income distribution remains close to $1. The return rises rapidly with income: the return is $2.99 in the 90–99th percentiles, $4.35 in the 99–99.9th percentiles and $8.63 in the top 0.1%.

5 Individual Deterrence Effects of Audits

Thus far, our analysis focuses on the direct revenue obtained from assessments made during an audit. Audits may also raise revenue in an indirect manner by deterring future non-compliance. In this section, we examine one particular form of deterrence, known as individual (or specific) deterrence. Individual deterrence refers to a situation where auditing an individual in one year encourages greater compliance by that individual in future years.[44]

In order to estimate individual deterrence effects, we draw upon random audits conducted by the National Research Program (NRP) and build on the approach developed by DeBacker et al.

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

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

[44]In general, the literature on tax audits distinguishes between two potential sources of deterrence – specific deterrence and general deterrence. General deterrence refers to a situation where auditing one individual encourages other individuals to change their behavior. We do not quantify the role of general deterrence in this paper, and so our estimates may be considered a lower bound on the total deterrence effect of tax audits.

(2018) (DHTY). As in DHTY, we construct a treatment group of individuals audited via the NRP and compare their taxes paid to a group of control individuals not selected for random audit. In our analysis, we construct our control group using stratified matching. The NRP divides taxpayers into strata as determined by return characteristics such as total positive income (TPI), the presence of the EITC, and the presence of self-employment income. For each audited observation, we select a matched control observation from the same stratum and same tax year using a nearest-neighbor approach. In particular, we match each selected return to the non-selected return within the same NRP stratum that has the closest level of TPI. We then track income reported and taxes paid in each year after the initial audit.

Our approach builds on DHTY in two key ways: First, we use additional data to substantially increase the duration of the observed post-audit period. We use random audits from 2006-2014 and estimate the impact on behavior up to 14 years post-audit. By contrast, DHTY draw upon random audits from 2006-2009 and estimate the impact on behavior up to 6 years post-audit. Second, our event study framework incorporates year-since-audit fixed effects. As we discuss more below, the omission of these controls in DHTY, in combination with the use of a top income quintile threshold, results in a mismatched comparison between treated and control taxpayers. That comparison produces a spurious negative deterrence effect at the top of the income distribution.[45]

Figure 6A plots results from our primary event study for the full population of audited individuals. We find clear and persistent deterrence effects. The figure displays the difference in mean taxes paid between taxpayers selected for an NRP audit and those in the matched control group. We weight each observation by the inverse of the NRP sampling probability to ensure that our results measure the average deterrence effect across the full population of taxpayers (as opposed to the distribution of audited taxpayers). As one would expect, the difference in taxes paid between treatment and control is statistically indistinguishable in the years prior to the NRP audit. In the years following the audit, a clear gap emerges. Starting in Year 2, we find a statistically significant deterrence effect.[46] The yearly impact on taxes paid is around $300 per audited taxpayer, which is approximately 30% of the $1,026 in revenue collected via the initial audit.[47] The results also

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

[46]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.

[47]The magnitude of this effect is broadly consistent with the previous literature, which finds annual deterrence effects ranging from 20% to 35%.

demonstrate that 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 sample window. There is no clear change in the magnitude of the effect over the full 14-year window.[48] 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.[49] This effect is roughly 3.2 times the revenue collected from the initial NRP audit.[50]

Next, we explore how these deterrence effects vary across the income distribution. Table 1 reports the results of the treatment-control difference-in-difference specification separately by bins of TPI.[51] We display both raw estimates and, to limit the influence of outliers, estimates where collected tax revenue is winsorized at the 99th percentile of the population distribution.[52] Broadly, we find significant deterrence effects across the income distribution. Figure 6B summarizes these results by plotting the ratio of the estimated deterrence effects to upfront revenue collected from the NRP audit.[53] 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.[54] While we lack the power to estimate precise effects in the top 1%, the point estimates remain similar at the 99–99.9th percentiles and in the top 0.1%.[55]

[48]Appendix Figure A10 shows that these results are also consistent across years of NRP audits. Panel A shows the deterrence event study separately for audited tax return years 2006-2009 and 2010-2014. Panel B then reports the same results for audited tax return years 2006-07, 2008-09, 2010-11, and 2012-14. The magnitude and trajectory of the treatment effects remain similar across treatment years.

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

[50]Appendix Figure A11 explores how deterrence effects differ for taxpayers with and without business income (as measured by income on Schedule C, E and F). For taxpayers without business income the future revenue collected is roughly 3.81 [1.30, 6.39] times initial NRP revenue. For taxpayers with business income, the ratio is 2.81 [1.47, 4.15]. The event study graphs show that, consistent with the patterns observed in DHTY, the deterrence effects for business owners are less persistent than the deterrence effects for other types of filers. Appendix Figure A12 explores deterrence effects among a sample of individuals chosen for typical in-person audits, rather than those chosen for NRP audit. These audits are non-random, but we nonetheless match audited individuals to non-audited individuals on the basis of income and return characteristics. The difference in differences estimates suggest a multiplier of approximately 2.5, consistent with the results in our NRP sample. These alleviates any concerns that the results from the NRP sample do not generalize to the case where individuals are faced with typical in-person audits.

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

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

[53]We express both of these using their present discounted value.

[54]In our baseline specification we winsorize collected tax revenue at the 1st and 99th percentiles of the distribution of taxes collected. Table 1 shows that these estimates remain consistent across the income distribution without winsorization. Appendix Figure A13 repeats the exercise found in Figure 6B to present a side-by-side comparison of the winsorized and unwinsorized deterrence ratios. While the unwinsorized estimates are noisier at the top of the income distribution, the broad trajectory of our results remains the same.

[55]For those in the top 1%, we cannot reject the hypothesis that audits generate 5x additional revenue from deter-

It is worth noting that our finding of positive deterrence at the top of the income distribution lies in contrast with previous estimates. Notably, DTHY find a negative deterrence effect in the top quintile of income. In analyzing their results, we find that this effect is driven by the omission of year-relative-to-audit fixed effects when restricting their sample to audits of high-income returns. The omission of these fixed effects means that treated audits in one year are compared against a set of control audits selected across multiple different years. The concern here is primarily one of mean reversion: filers classified as high-income in 2006 may have regressed farther to the mean during 2010-2014 than filers classified as high-income in 2009.[56] Without including year-since-audit fixed effects, the time path of treatment interacted with year since audit partially captures this mean reversion.[57] This produces a spurious negative effect for filers classified as high-income. The inclusion of years-relative-to-audit fixed effects in their initial specification recovers estimates broadly consistent with the findings in this paper.[58]

Having assessed deterrence effects overall and across the income distribution, Figure 7 multiplies these ratios of revenue from deterrence to revenue from the initial NRP audit by the revenue raised from operational audits. Panel A takes the baseline R/C figures reported in Figure 5 and applies the overall deterrence multiplier of 3.2.[59] This produces a new estimate of the total revenue raised per marginal dollar spent on audits.[60] 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

rence, nor can we reject that they have large negative effects on deterrence.

[56]It may also be the case that the treatment and control groups differ across years in terms of their income trajectories. For example, individuals found in the highest income quintile during the onset of the Great Recession may have different income trajectories than those found in the highest income quintile before the recession in 2006. Such a pattern could also contribute to the misalignment between treatment and control.

[57]This also helps to explain why the deterrence effects in DHTY become more negative over time, as the misalignment between treatment and control becomes more substantial the more years after the initial audit.

[58]Including DHTY's tax year fixed in our specification in addition to existing year-since-audit fixed effects does not alter our results because treatment and control are balanced in each tax year.

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

[60]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 here is that future behavioral changes across NRP and in-person audits are a constant multiple of the size of the audit adjustment. This is consistent with a model where taxpayers learn from an audit that certain filing 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, there is a lot of commonality between NRP and in-person audits. Both are conducted by the same staff, and while NRP audits are more intensive than traditional in-person audits, that is primarily because they examine more line-items than in-person audits. That said, the use of NRP audits to measure deterrence effects may even be a slightly conservative assumption. Individuals subject to NRP audits are told they were randomly selected, while those chosen for traditional in-person audits have to reason to believe that their audit was triggered by content of their previous tax filings. Those subject to in-personal audits may have larger behavioral responses because they infer that similar filing practices will trigger additional audits in the future.

spent on auditing an individual in the 70–80th percentiles produces a return of $9.06. Each dollar spent auditing an individual in the 90–99th percentiles produces a return of $12.48. We use these figures below to conduct a welfare analysis of marginal audits.

Panel B shows how these estimates vary when using deterrence multipliers calculated within TPI bins. The confidence intervals increase as the deterrence effects are estimated with more uncertainty, but the trajectory of the point estimates remain the same. For example, when using the TPI-bin specific deterrence multiplier of 6.29x[61] for taxpayers at the 90–99th percentiles, we get a total return of $21.79:1. When following the same approach we get a point estimate $24.40:1 for taxpayers at the 99–99.9th percentiles and a point estimate of 44.05:1 in the top 0.1%.[62]

6 Welfare Analysis of Marginal Audits

Our results suggest that audits, particularly audits of high-income tax payers, have the potential to produce revenue that far exceeds their costs. What does that imply about the welfare consequences of tax audits? How should we think about the tradeoff between greater tax enforcement and alternate policies to 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) for a change in the audit rate at each point of the income distribution.

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

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

As is the case for any revenue raising policy, a lower MVPF means the policy is more efficient at raising funds. It imposes a lower welfare cost for each $1 of revenue raised. So, if the MVPF of a tax audit were 1.1, that would imply that the audit imposes $1.10 in welfare costs for each $1 in government revenue.[63]

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

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

[63] For point of reference, a simple non-distortionary tax would have an MVPF of 1. The revenue raised by the government would be exactly equal to the individual beneficiary's willingness to pay to avoid the tax.

Appendix C provides a formal model derivation of the MVPF in a standard class of dynamic models with audits and evasion; we focus the main text on the intuition of the MVPF components that we derive from this model. Let us first consider the numerator. We incorporate two key terms into the willingness to pay to avoid the audit. First, we assume that audited individuals experience a welfare loss equal to the amount of additional money they are required to pay the IRS. Put another way, it is essentially tautological that an individual should be willing to pay \$100 to avoid writing a \$100 check to the IRS. We denote this amount by R, as it constitutes the revenue raised by the audit. It is important to note 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.[64]

Second, the costs of audit include not only the revenue paid to the IRS but also other costs, which we call taxpayer burden and denote with B. Taxpayer burden captures an individual's willingness to pay to avoid the cost of representation by specialists, time costs, and other hassles associated with an audit. Combining the revenue paid with the taxpayer burden yields the willingness to pay to avoid an audit, $R + B$.

Next, we turn to the denominator of the MVPF, which captures the net cost to the government of additional audits. This is equal to the marginal revenue raised by the audit, R, minus the marginal cost the government pays to conduct the audit, C. When we combine the willingness to pay, $R + B$, and net cost to the government, $R - C$, this yields the formula for the MVPF of:

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

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

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

Here, R/C is the revenue raised per dollar of marginal spending and B/C is the taxpayer burden per dollar of government cost. Intuitively, a higher value of R/C pushes the MVPF of audits closer to 1. If the cost to conduct the audit is small compared to the revenue collected, the policy is closer to a non-distortionary tax (which would have an MVPF of 1). By contrast, if R/C is low, then

[64]The inclusion of the full deterrence effect in the willingness-to-pay may be a conservative assumption in this case. It could be the case that individuals choosing to evade large quantities of taxes are relatively close to indifferent between their options. (The value of the additional income just barely exceeds the monetary and psychological costs of evasion.) In that case, if the NRP audit induces a large behavioral response, the application of the logic of envelope theorem suggests that individuals may have a relatively small willingness to pay. In such a case, the initial revenue from the audit should be included in the numerator of the MVPF but the deterrence revenue should be omitted.

there are large distortionary costs associated with the audit, and so the MVPF is higher.[65]

In estimating the MVPF of tax audits across the income distribution we use the novel empirical estimates of R/C developed in this paper. While we do not conduct our own analysis of B/C, we use previously conducted IRS surveys that determine the total taxpayer burden required to comply with the audits. We follow the approach developed by Guyton and Hodge (2014), who monetize the taxpayer burden of audits reported in a survey. They estimate that the average dollar value of the taxpayer burden of an audit is \$3,198 for in-person audits.[66] We divide this figure by the average cost to conduct an in-person audit (\$6,418) which yields \$0.50 of taxpayer burden per dollar of government cost.

When calculating the magnitude of the taxpayer burden across the income distribution, we assume that the ratio B/C, remains constant. This assumption captures the idea that the burden on the part of taxpayers will vary with the degree of engagement by the auditors themselves. The results imply that the taxpayer burden ranges from \$2,600 per audit in the bottom half of the income distribution to \$5,018 in the 99-99.9th percentile.

With these inputs, we can calculate the MVPF of audits across the income distribution. Figure 8A reports the construction of the MVPF of expanding audits among those with TPI in the 20-30th percentiles. Recall, we estimate that at the 20–30th percentiles every \$1 in audit spending generates \$1.14 in upfront revenue and \$3.63 in future revenue from deterrence effects. This generates total revenue of \$4.78 per dollar of marginal spending. The willingness to pay to avoid the audit is the sum of this \$4.78 in revenue and the \$0.50 in taxpayer burden, for a total of \$5.27. Each dollar of government expenditure raises \$3.78 in net revenue (\$4.78 in upfront and deterrence revenue minus \$1 of initial spending). We take the ratio of these two terms to get an MVPF of 1.40. Figure 8B repeats this calculation for expanding audits among those in the 90–99th percentiles of the TPI distribution. Here, we find an MVPF of 1.13. This MVPF is meaningfully closer to 1 because the policy has a higher return on expenditure, R/C.

So, would it be welfare enhancing to audit more high-income individuals and fewer low-income individuals? That depends on the welfare weights placed on the audited 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

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

[66] This figure is a weighted average of the estimated taxpayer burden for field and office examinations.

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, as in Saez and Stantcheva (2016). 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. Given our MVPF estimates, this means that if a social planner places equal welfare weights on the willingness to pay of low- and high-income audited individuals, it would be welfare enhancing to increase audits on the 90–99th percentiles rather than increasing audits on the 20–30th percentiles. In fact, this calculation suggests that such a policy is welfare enhancing as long as the welfare weight on audited high-income individuals is less than 1.23 times the welfare weight placed on audited low-income individuals.

Comparison to tax and transfer policies A key benefit of the MVPF approach to welfare analysis is that it can be used to compare the welfare consequences of policies both within and across policy domains. Here, we consider how the MVPF of tax audits differs from the MVPF of other tax and transfer policies.

Hendren and Sprung-Keyser (2020) show that for a change in tax rates, the MVPF is given by

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

where FE is the impact of the behavioral response to the policy on the government budget. The MVPF of most tax increases lie above 1 as the behavioral response to those policies often induces a negative fiscal externality.[67]

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

[67] Intuitively, the deadweight loss associated with tax audits is driven by the cost of the audits; the deadweight loss from changes to the tax schedule is driven by the impact of the behavioral response to the tax change on tax revenue. In this sense, from a welfare perspective, the cost of the audit is analogous to the deadweight loss of a tax increase, though the cost of conducting an audit weighs more heavily because the funds to conduct an audit must themselves be raised.

the MVPF of a small change in the tax schedule at each point of the income distribution.[68] Finally, 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 closer to 1.5-2. 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 marginal social welfare weights placed on those two groups. Put another way, if, at an optimum, the MVPF of a high-income tax increase is 2 and the MVPF of a low income tax increase is 1, that implies the social planner values \$1 in the hands of a low-income person as much as it values \$2 in the hands of a high-income person.

While the MVPFs of tax and transfer policies increase with income, the MVPFs of marginal tax audit expansions display the exact opposite pattern. The MVPF of tax audits of high-income individuals is meaningfully lower than the MVPF of audits on lower income individuals. We find MVPFs between 1.29-1.39 for those with below-median income, but MVPFs that converge close to 1 for those with income above the 90th percentile.

These two sets of MVPFs, those of tax and transfer policies and those of tax audits, are inconsistent with canonical social preferences. We can see this in two ways – comparing MVPFs across income levels and within income levels.

When comparing across income levels, we see that the slope of the MVPF of tax changes is positive, while the slope of the MVPF of tax audits is negative. This means that the MVPFs for tax rate changes imply a set of social welfare weights that fall with income, while the MVPF of tax audits imply a set of weights that contradict this pattern because they rise with income.[69]

Focusing within high-income taxpayers, we can further explore why current audit rates are not consistent with canonical social preferences. In particular, it is helpful to consider two benchmark cases. First, suppose that society places equal weight on dollars in the hands of all high-income

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

[69] Formally, one could rationalize these patterns if tax rates and audit rates were set by a social planner that A) places more weight on earned income in the hands of low-income taxpayers over the hands of high-income taxpayers and B) prefers to collect audit revenue from low-income taxpayers rather than high-income taxpayers. It is hard to envision a set of preferences that result in such an outcome.

individuals. In other words, society does not care whether a dollar is collected via taxes or via audit. In that case, we should expect the MVPF of tax audits to be equal to the MVPF of high-income tax increases. Any deviation from that benchmark is socially inefficient. So, the fact that the MVPF of tax audits lies above the MVPF of tax changes is suggestive of such an inefficiency.[70]

Second, suppose that the social planner places little value on allowing taxpayers to keep unpaid tax dollars. In that case, the planner would place a relatively low welfare weight on the revenue collected via tax audits. (If the money collected via audit were already owed to the government, then a social planner might give little weight to a taxpayer's willingness to pay to avoid an audit.) One should therefore expect the MVPF of tax audits to lie above the MVPF of tax changes at each point in the income distribution. While we see such a pattern when comparing tax changes and tax audits for low-income individuals, we see no such pattern for high-income individuals. The MVPF of tax changes for high-income taxpayers typically lie well above the MVPF of tax audits. This ordering of MVPFs could only be rationalized by a set of social welfare weights that prioritize individuals who paid less in taxes than they owe.

7 Conclusion

In this paper, we conduct a detailed analysis of the returns to tax audits across the income distribution. We estimate that the average IRS audit for tax returns filed between 2010 and 2014 produced \$2.17 in revenue for each dollar spent on the audit. We find those returns varied substantially across the income distribution. Audits of higher income taxpayers are more time intensive and more costly than audits of lower income taxpayers, but the revenue obtained is also significantly higher. We find that the average return to IRS audits rises from \$0.96 in the bottom half of the distribution to \$2.18 in the 90–99th percentiles and \$6.29 in the top 1%.

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

We then use random audits conducted as part of the IRS National Research Program to estimate individual deterrence effects of audits. We find that auditing individuals in a given year results in a persistent increase in tax revenue collected from those individuals. Measured over 14 years, the

[70] It is worth noting that this finding will hold even if hassle costs on top earners far exceed the estimates used here. For example, among tax payers in the top 0.1%, the the MVPF of auditing top income individuals would remain below the MVPF of tax rate changes, even if estimated hassle costs were increased by a factor of 50.

future revenue collected is 3.2 times the return to the initial audit. We find that these deterrence effects are observed 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. We also consider the welfare implications of these results, interpreting our findings through the lens of the MVPF framework. We estimate that the MVPF of auditing taxpayers in the 90–99th percentiles of income is 1.13. By contrast, the MVPF of auditing a taxpayer at the 20–30th percentiles is 1.40. These results suggest that current audit levels across the income distribution are inconsistent with a wide range of canonical social preferences.

There is ample room for future work to build upon our findings. For example, our analysis measures the specific 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 could substantially increase the returns to audits. Moreover, our measures of average and marginal costs capture the steady state that existed for filings between tax years 2010 to 2014. Hiring and training new auditors requires a substantial upfront cost before those auditors begin to yield revenue. Future work should explore how the time path of hiring costs and the accumulation of auditor expertise shape the returns to audits. Finally, the analysis here is restricted to IRS in-person audits of individuals. While our work captures audits of individual income generated by businesses, future work should further examine the returns to auditing businesses themselves.

References

Advani, A., W. Elming, and J. Shaw (2021). The dynamic effects of tax audits. *The Review of Economics and Statistics*, 1–45.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

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

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

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

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

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

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

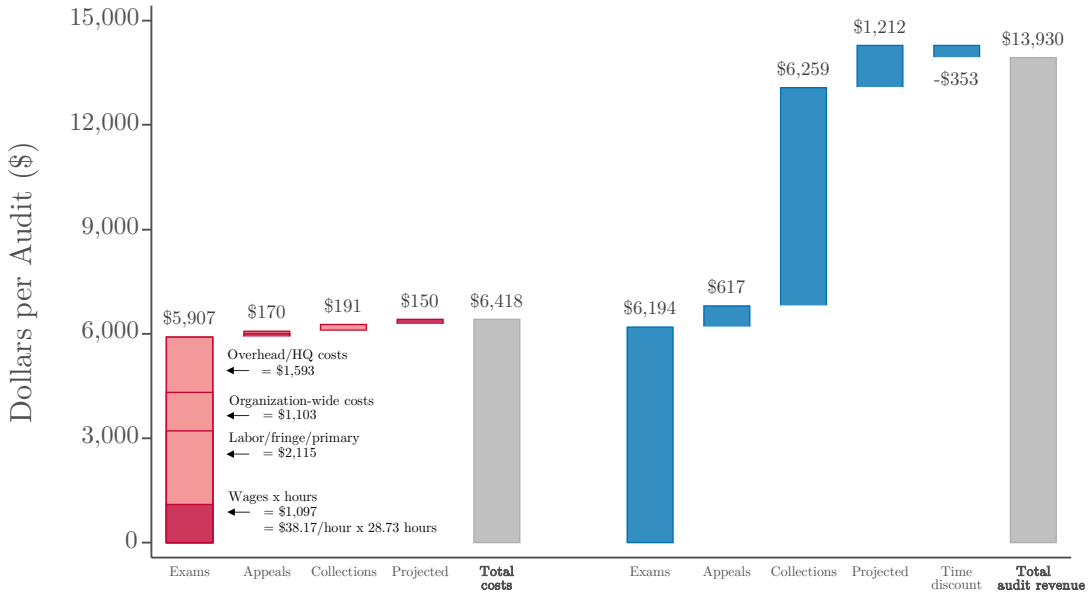
Sarin, N. and L. H. Summers (2019). Shrinking the tax gap: approaches and revenue potential. Technical report, National Bureau of Economic Research.

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

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

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

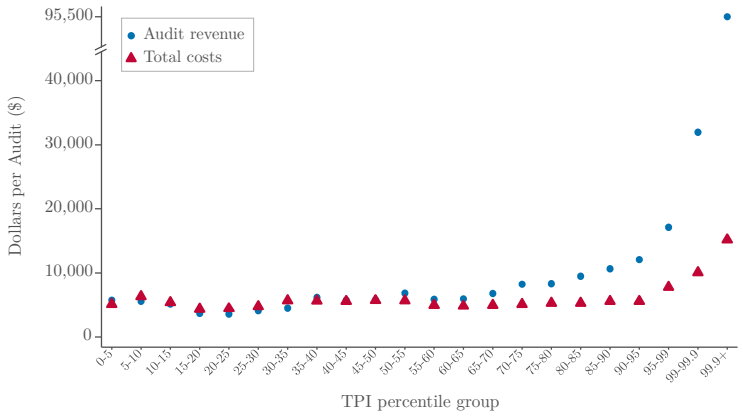
FIGURE 1: Average Costs and Revenue Raised per In-Person Audit



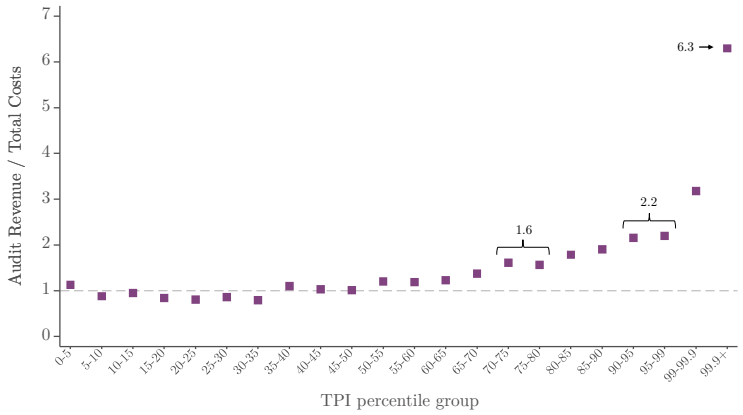
Notes: This figure presents the average total costs and revenue raised per in-person audit conducted between 2010-2014. On the cost side, total wage costs (calculated as auditor's wages x hours spent on exam) are shown in dark red, and additional costs are shown stacked on top in lighter red. Additional costs include labor/fringe/primary, organization-wide, and overhead/HQ costs. Together, these additional costs are 4.39, 0.57, and 4.15 times total wage costs at the exam, appeals, and collections stage respectively using average multiplier values from 2011-2015. Revenue raised at each stage of the audit process is shown in blue and is inclusive of revenue raised from additional tax liability, penalties and interest. Details on each stage of the audit process is outlined in Appendix Figure A1. Average costs and revenues are projected to include estimates of revenue collected and costs accrued outside the observed 7-11 year post-audit sample window. Revenues are discounted using a 3% discount rate since average revenues lag average costs by approximately one year. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

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

A. Average Total Costs and Audit Revenue



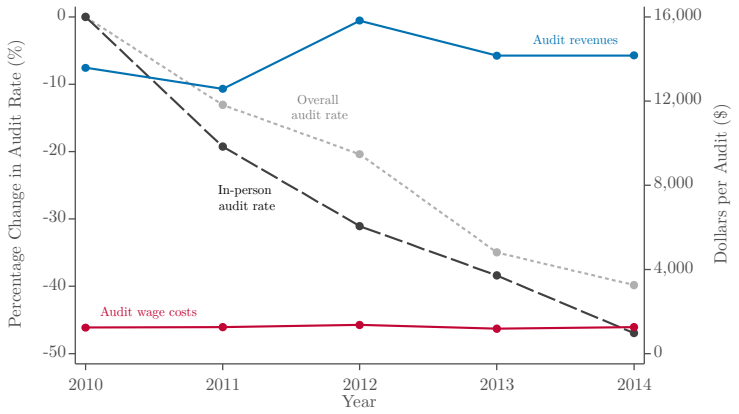
B. Average Audit Revenue over Total Costs



Notes: Panel A presents the average total costs and revenue raised per in-person audit conducted between 2010-2014 by 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 five percentile bins and splits out the top bin into the 95-99th and 99-99.9th percentile and the top 0.1%. Total costs are calculated as the sum of labor costs (auditor's wages x hours spent on exam) and additional costs (labor/fringe/primary, organization-wide, and overhead/HQ costs) which are allocated in proportion to direct labor costs across the income distribution. Total revenue is the sum of revenue raised from additional tax liability, penalties and interest. Average costs and revenues are projected to include estimates of revenue collected and costs accrued outside the observed 7-11 year post-audit sample window. Revenues are discounted using a 3% discount rate since average revenues lag average costs by approximately one year. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

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

A. Audit Probability, Audit Revenue Collected and Wage Costs

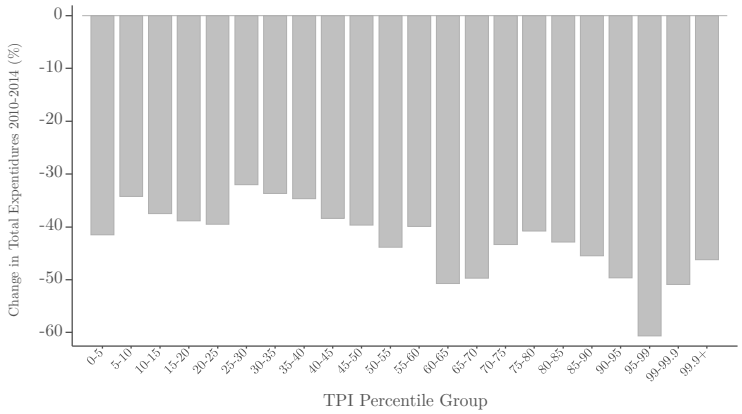


B. Components of Wage Costs

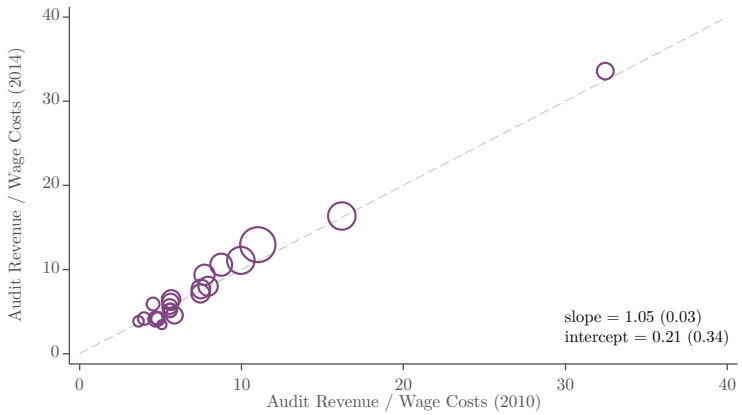
Notes: Panel A presents the percentage change in overall and in-person audit rates, total revenues raised, and direct labor costs (auditor's wages x hours spent on exam) per in-person audit each year in our sample frame (2010-2014). Panel B shows each component of labor costs (auditor's wages and hours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

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

A. Change in Audit Expenditures Between 2010-2014

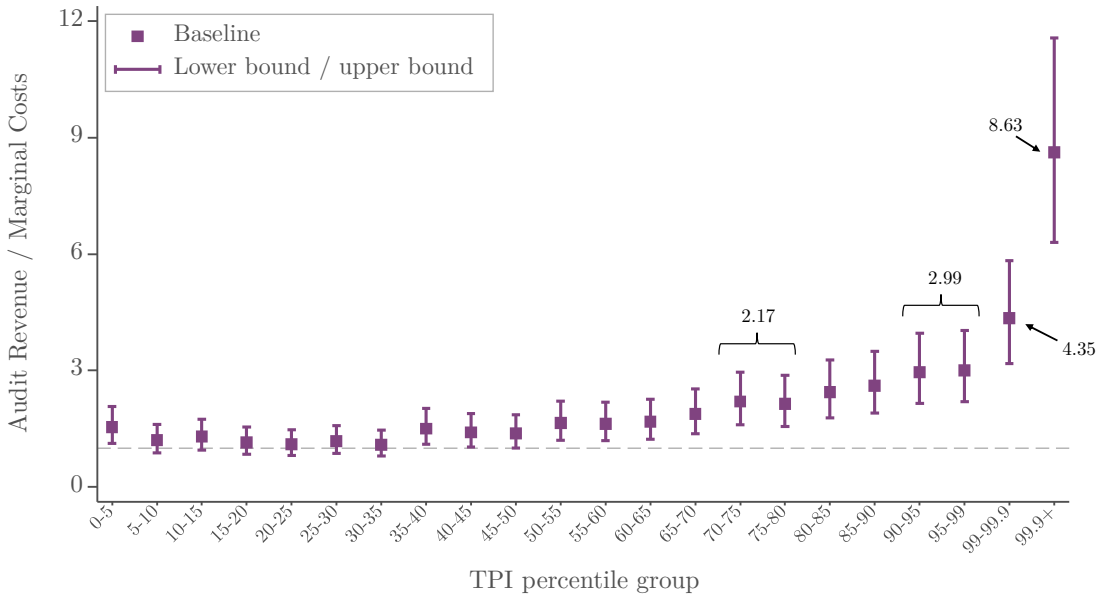


B. Change in Audit Revenue Over Wage Costs between 2010-2014



Notes: Panel A shows the percentage change in total direct wage costs of audits between the first year (2010) and the last year (2014) in our sample frame, by taxpayer's total positive income (TPI). Total direct wage costs are calculated as the product of labor costs per audit (auditor's wages x hours spent on exam) and the number of audits in each TPI bin. TPI is grouped into five percentile bins 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 to audits in each TPI bins (measured as the revenue per dollar of direct wage costs) in 2014 against the values in 2010. Average costs and revenues are projected to include estimates of revenue collected and costs accrued outside the observed 7 and 11 year post-audit sample window for 2014 and 2010 audits respectively. Revenues are discounted using a 3% discount rate since average revenues lag average costs by approximately one year. In particular, we use data from the 2003 tax year to separately discount the revenues raised and costs accrued each year post-audit back to the tax year. We then use the ratio of the discounted series (net present value of revenues over costs) to adjust revenues downwards to align the two paths.

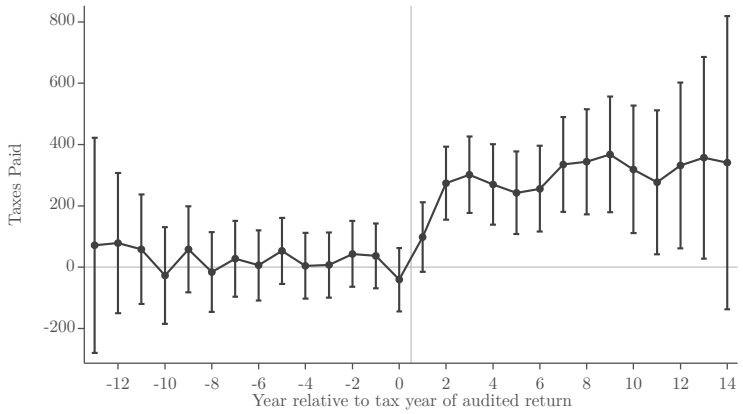
FIGURE 5: Average Audit Revenue over Marginal Costs



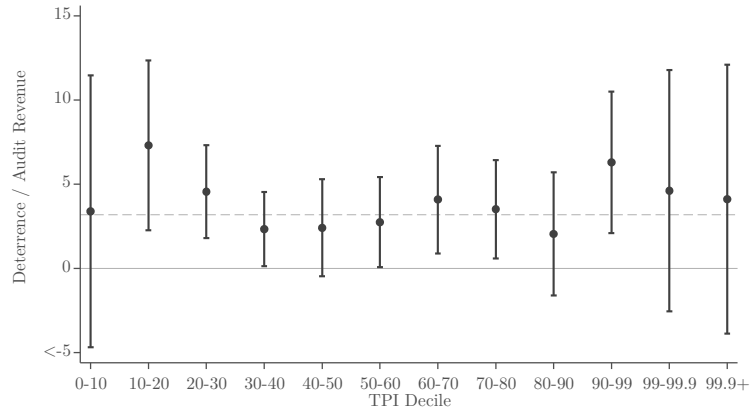
Notes: This figure shows the ratio of average revenue raised and total marginal costs (calculated using marginal overhead estimates) per in-person audit conducted between 2010-2014 by taxpayer's total positive income (TPI). Our baseline estimates use a marginal cost multiplier of $2.93 per dollar of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

FIGURE 6: Estimated Deterrence Effects

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



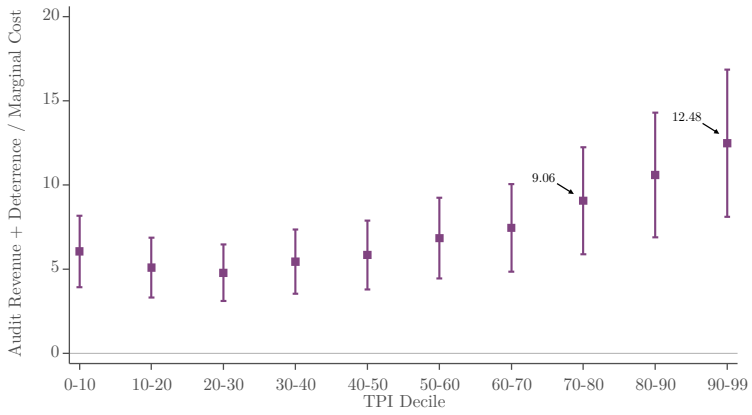
B. Deterrence effect over initial audit revenue, by income



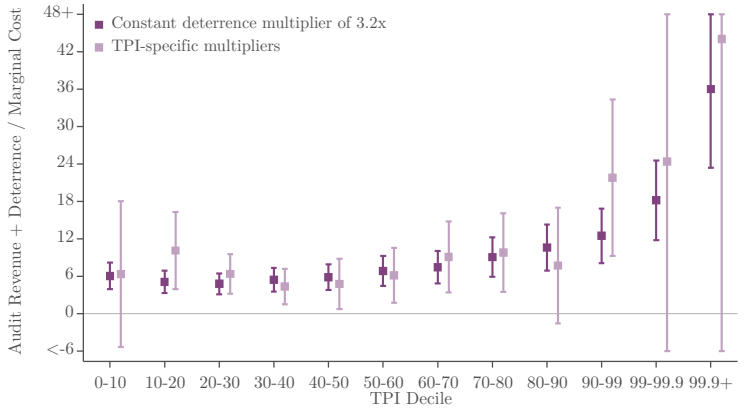
Notes: Panel A presents estimates of the change in taxes paid each year post-audit for the full population individuals selected for random audit via the National Research Program (NRP). The control group is a matched sample of individuals not selected for random audit. Collected tax revenue is winsorized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

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

A. Incorporating Overall Deterrence Effect

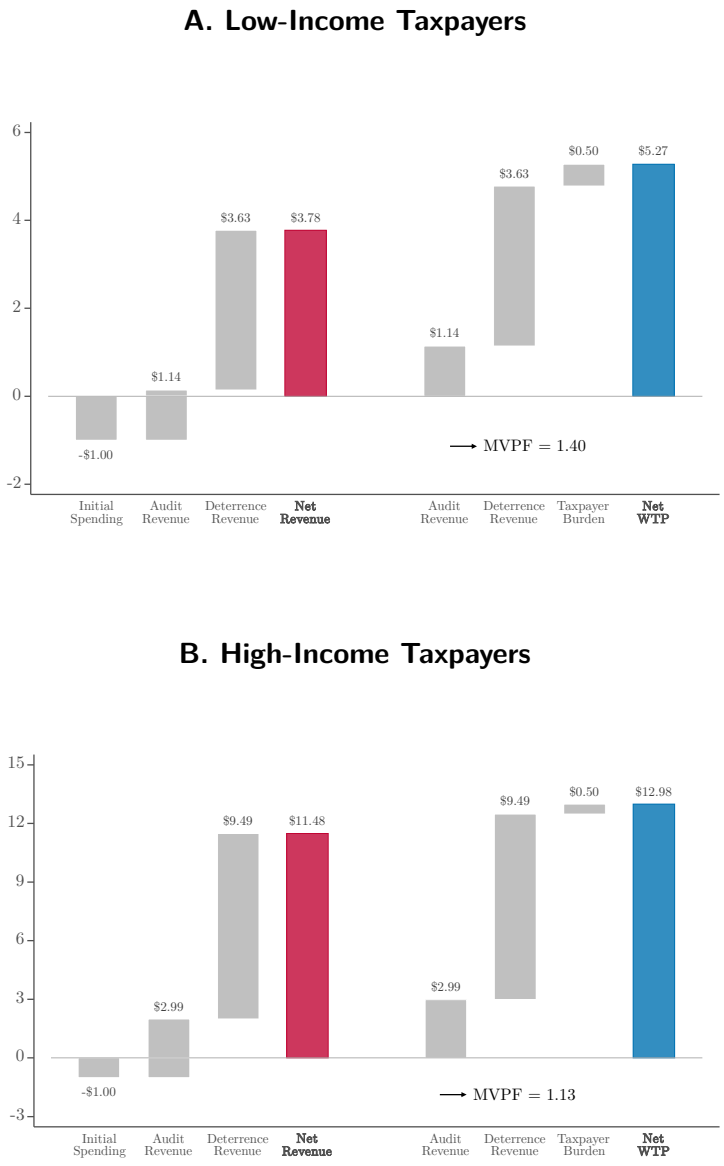


B. Incorporating TPI-Bin Specific Deterrence Effects



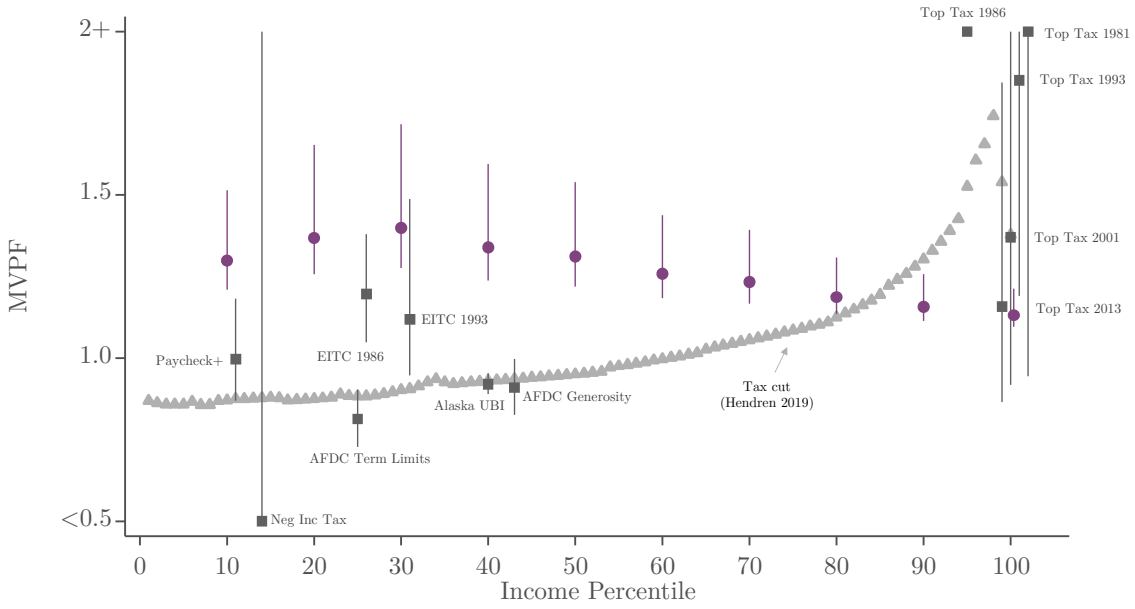
Notes: This figure shows the robustness of total revenue raised (deterrence and upfront audit revenue) per marginal dollar spent on audits across the income distribution using different estimates of deterrence effects. These estimates are calculated as the product of one plus the deterrence multiplier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

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



Notes: This figure presents the construction of the MVPF of expanding audits among those with TPI in the 20–30th percentiles (Panel A) and 90–99th percentiles (Panel B). The MVPF is calculated as the ratio of the taxpayer's willingness to pay to avoid audit (shown in blue) over the net government revenue raised by the audit (shown in red). Net government revenue includes upfront and deterrence revenue minus the cost of conducting the audit. Upfront revenue is measured as the revenue raised per dollar of marginal spending (reported in figure 7). The deterrence effect is calculated by multiplying direct revenue by the relevant deterrence multiplier shown in Figure 7 Panel B. Net willingness to pay to avoid audit includes the financial and non-financial costs of the audit. The financial costs include the upfront taxes paid as a result of the audit and the downstream additional taxes paid due to deterrence effects. Non-financial costs are measured by the total taxpayer time required to comply with the audits and monetized using average wages.

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



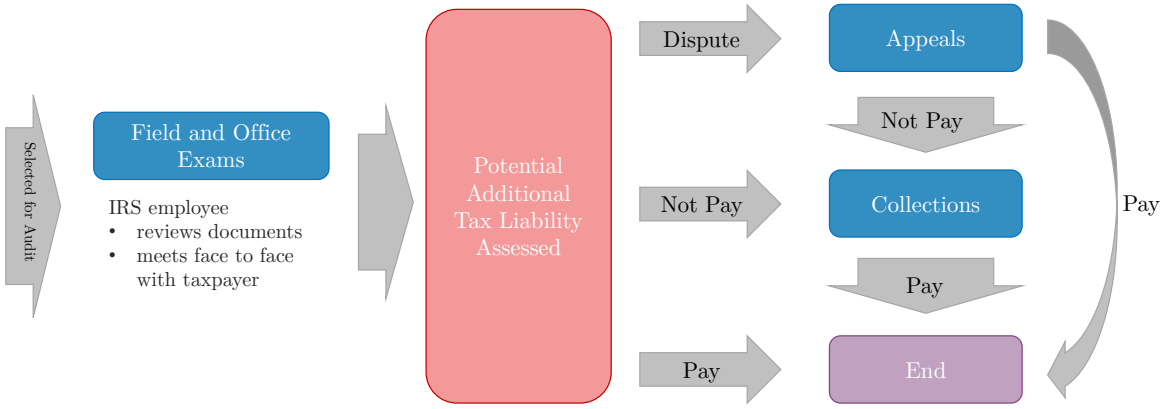
Notes: This figure compares the MVPF of tax audits to a wide range of tax and transfer policies within the United States. The horizontal axis reports the quantiles of the income distribution, displaying how these MVPFs vary with income. The gray triangles report MVPF estimates of tax and transfer policies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

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

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

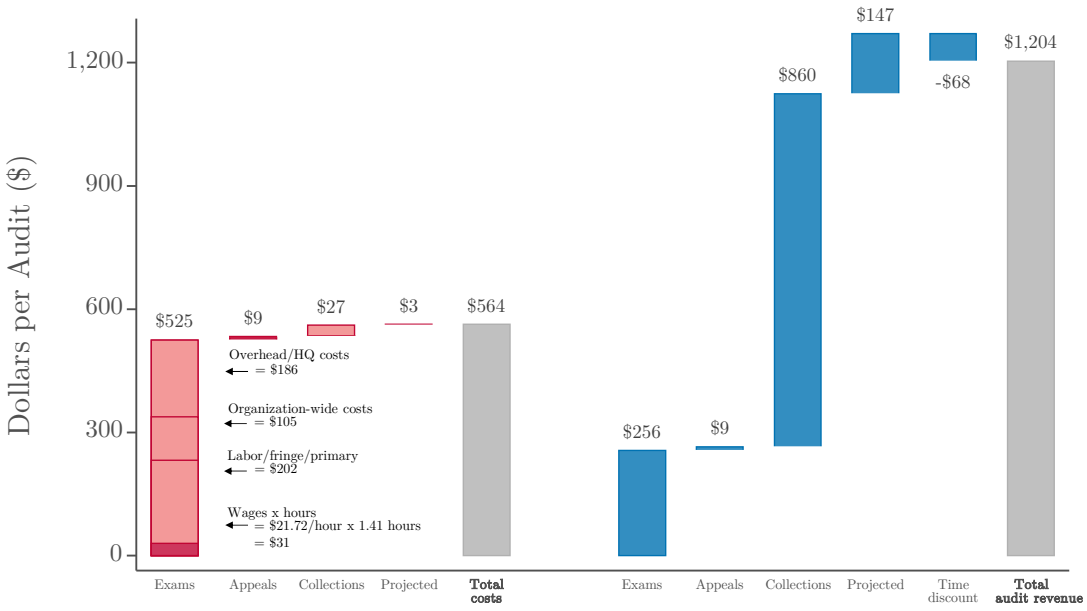
Notes: This table presents individual deterrence effects versus mechanical audit revenue by deciles of taxpayer's total positive income (TPI). Deterrence effects are estimated from a difference in differences regression that compares total taxes paid up to 13 years post-audit for individuals selected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

APPENDIX FIGURE A1: Overview of the Audit Process



Notes: This flow chart provides an approximation of what happens to a return after selection for a field or office exam. The exam begins when an IRS employee reviews the taxpayer's relevant documents and meets face-to-face with the taxpayer to assess whether they have additional tax liability. If there is no additional tax liability assessed, the exam ends, and the return exits the post-filing enforcement process. If there is additional tax liability assessed, the next stage of the audit process depends on whether the taxpayer agrees with the assessment. If they agree, then the outstanding amount is paid in full, and the exam ends. If they disagree, any disputed amount of the assessed liability goes to the appeals stage where a new assessment is performed. If the liability assessed during either the exam or the appeals stage is not collected in full during the audit process, the individual's case is sent to collections and is considered TDA (Tax Delinquent Account). In practice, very few exams ever end up in the appeals stage or the collections stage. Our estimates of RC consider revenue raised and costs accrued at the exam, appeals, and collections stages of the audit process.

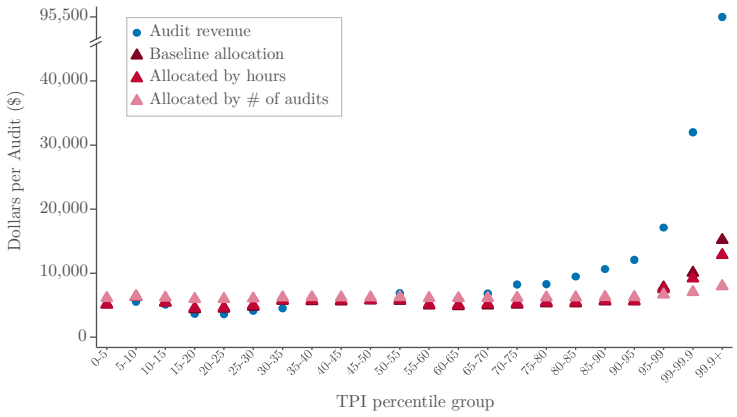
APPENDIX FIGURE A2: Average Costs and Revenue Raised per Correspondence Audit



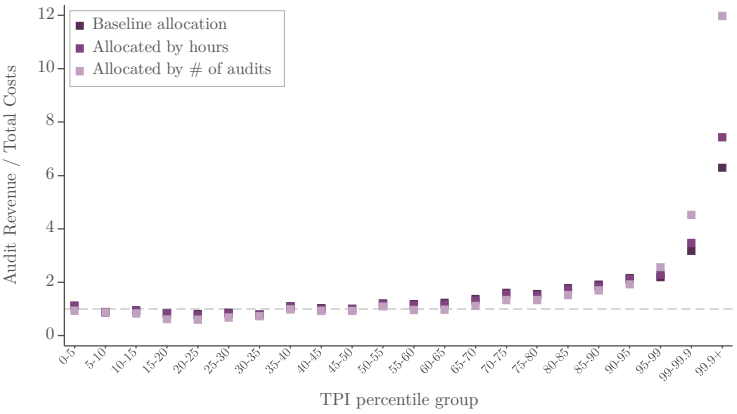
Notes: This figure presents the average total costs and revenue raised per correspondence audit conducted between 2010-2014. On the cost side, total wage costs (calculated as auditor's wages x hours spent on exam) are shown in dark red, and additional costs are shown stacked on top in lighter red. Additional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

APPENDIX FIGURE A3: Average Costs, Revenue and Revenue over Costs per In-Person Audit with Alternative Non-Direct Labor Cost Allocations, by Income Group

A. Average Costs and Revenue



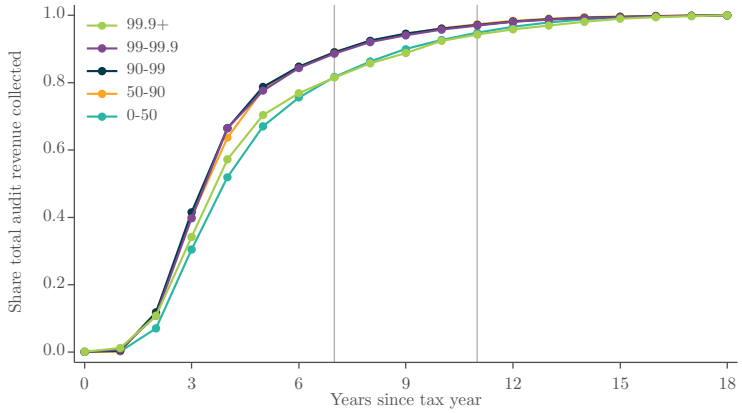
B. Average Revenue over Costs



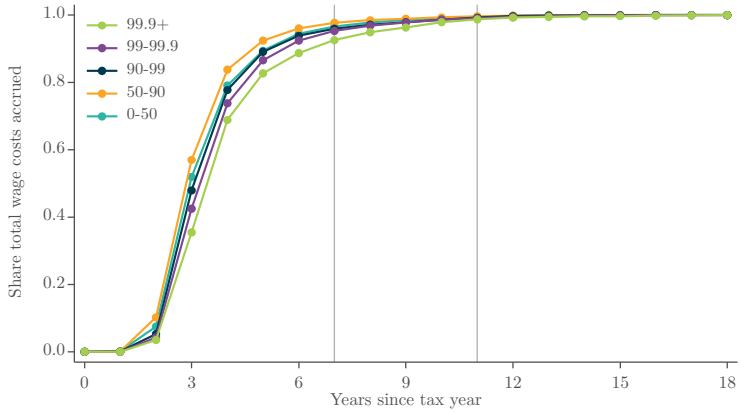
Notes: This figure presents the average total costs accrued and revenue raised per in-person audit (Panel A) and the ratio of average revenue and costs (Panel B) by taxpayer's total positive income (TPI) using different methods to allocate overhead (non-direct labor-related costs, organization-wide costs,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

APPENDIX FIGURE A4: Cumulative Share of Revenue Collected and Labor Costs Accrued, by Years Post-Tax Year and Income

A. Audit Revenues

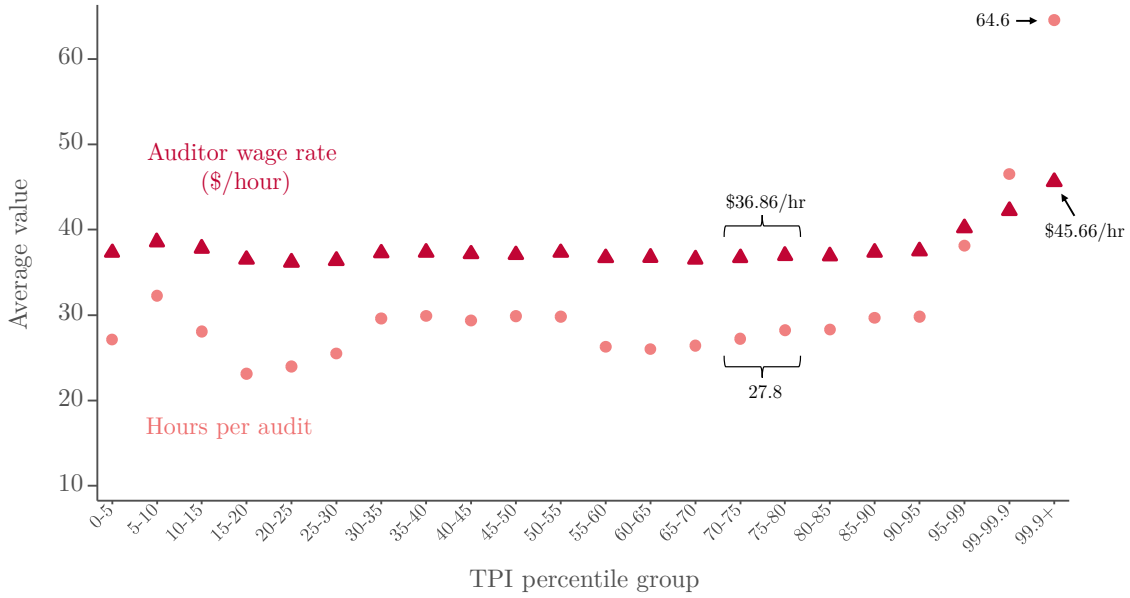


B. Labor Costs



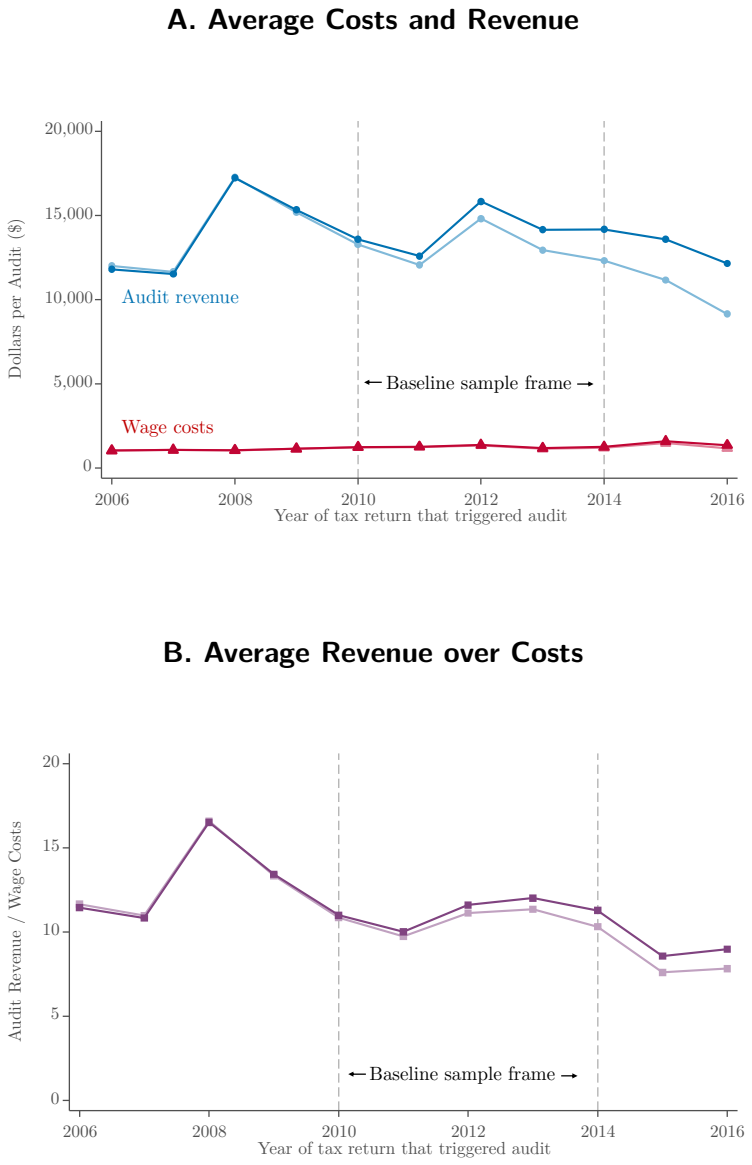
Notes: This figure presents the trajectory of revenues raised (Panel A) and costs accrued (Panel B) each year post-audit for returns filed in the 2003 tax year. The 2003 tax year lies before our primary sample window, but shows 18 years of follow-up data. The y-axis shows the cumulative share of revenues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

APPENDIX FIGURE A5: Average Hours per In-Person Audit and Auditor Wage Rate, by Income Group



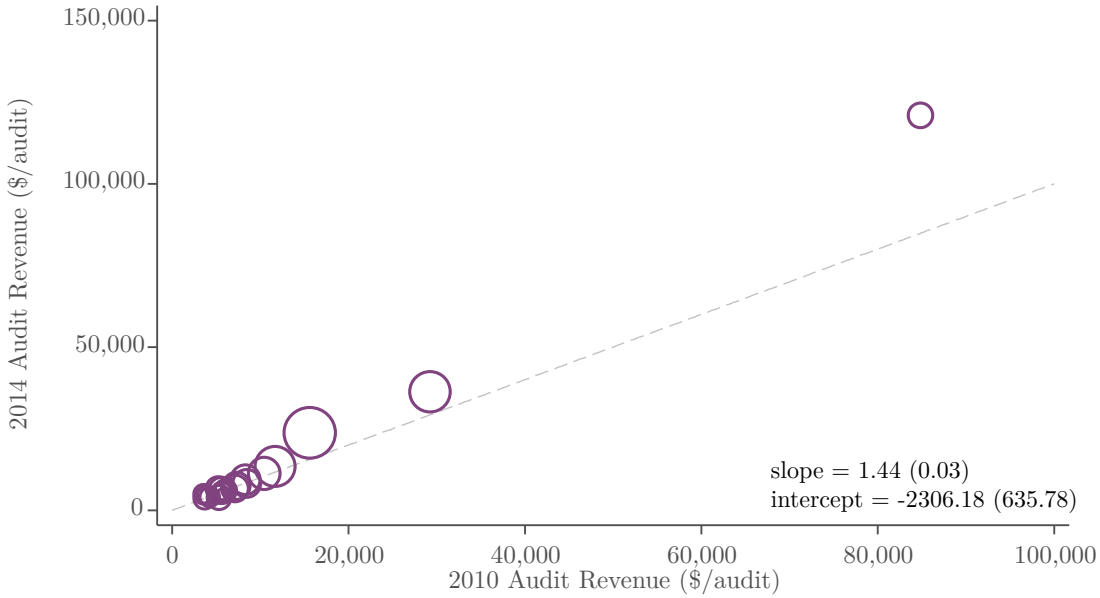
Notes: This figure presents the average value of the components of labor costs (auditor's wages x hours spent on exam) for in-person audits conducted between 2010-2014 by taxpayer's total positive income (TPI). The x-axis groups TPI into five percentile bins and splits out the top bin into the 95-99th and 99-99.9th percentile and the top 0.1%. Average hours per audit are projected to include estimates of labor hours accrued outside the observed 7-11 year post-audit sample window.

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



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

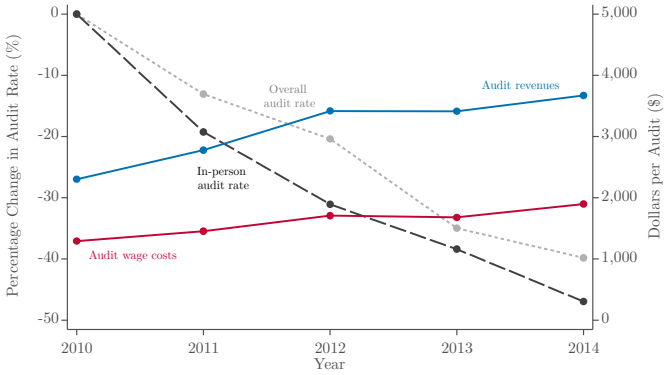
APPENDIX FIGURE A7: Marginal Revenue



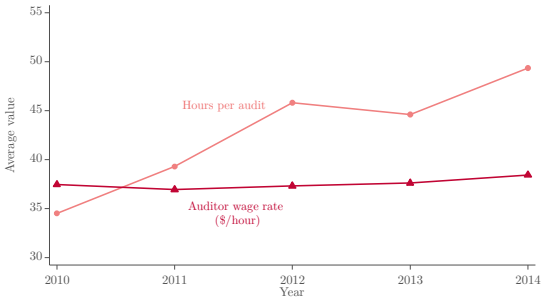
Notes: This figure shows average revenue raised per in-person audit in each 5% TPI bin in 2014 against the values in 2010. Average revenues are projected to include estimates of revenue collected and costs accrued outside the observed 7 and 11 year post-audit sample window for 2014 and 2010 audits respectively.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

APPENDIX FIGURE A8: Audit Probability, Revenue Collected and Wage Costs per NRP Audit, by Year

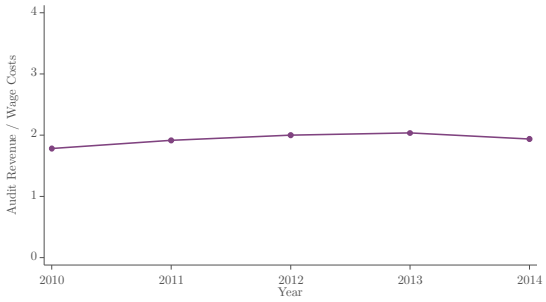
A. Audit Probability, Revenue Collected and Wage Costs



B. Components of Wage Costs

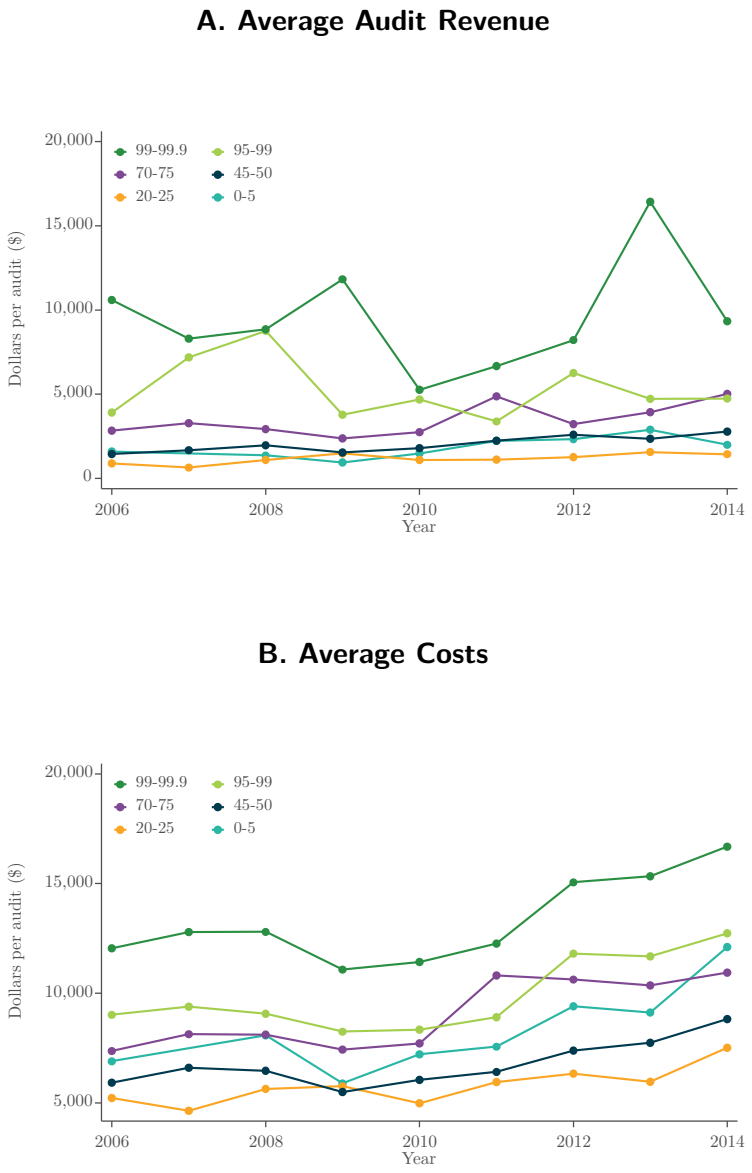


C. Audit revenue over Wage Costs



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

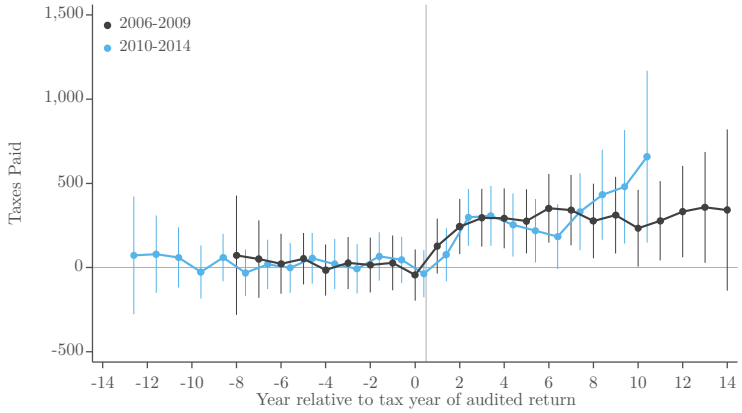
APPENDIX FIGURE A9: Average Revenue and Costs per NRP Audit, by Income and Year



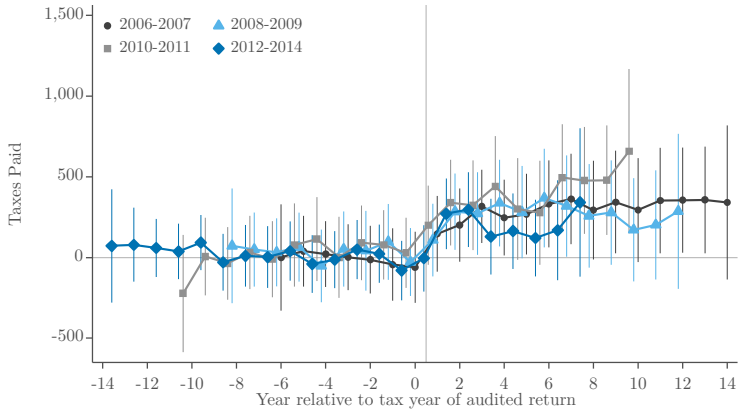
Notes: This figure presents total audit revenues raised (Panel A) and total costs accrued (Panel B) per random audit in each National Research Program (NRP) study year for select bins of Taxpayer's total positive income (TPI) across the income distribution. Total revenue is the sum of revenue raised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

APPENDIX FIGURE A10: Within-Taxpayer Impact of Audits on Future Tax Payments, by NRP Study Year

A. 4 year groupings

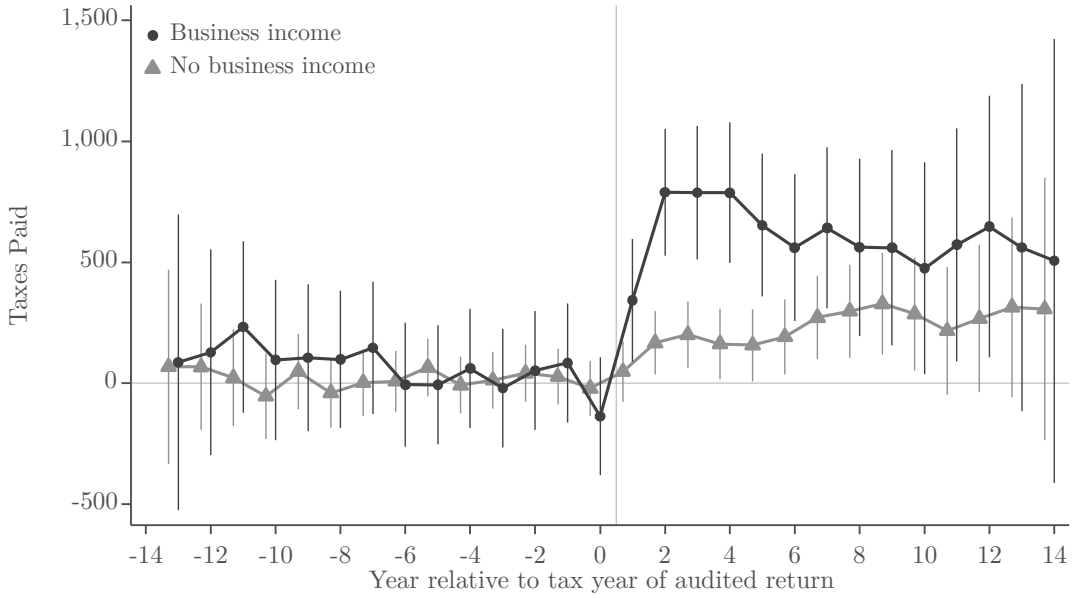


B. 2-3 year groupings



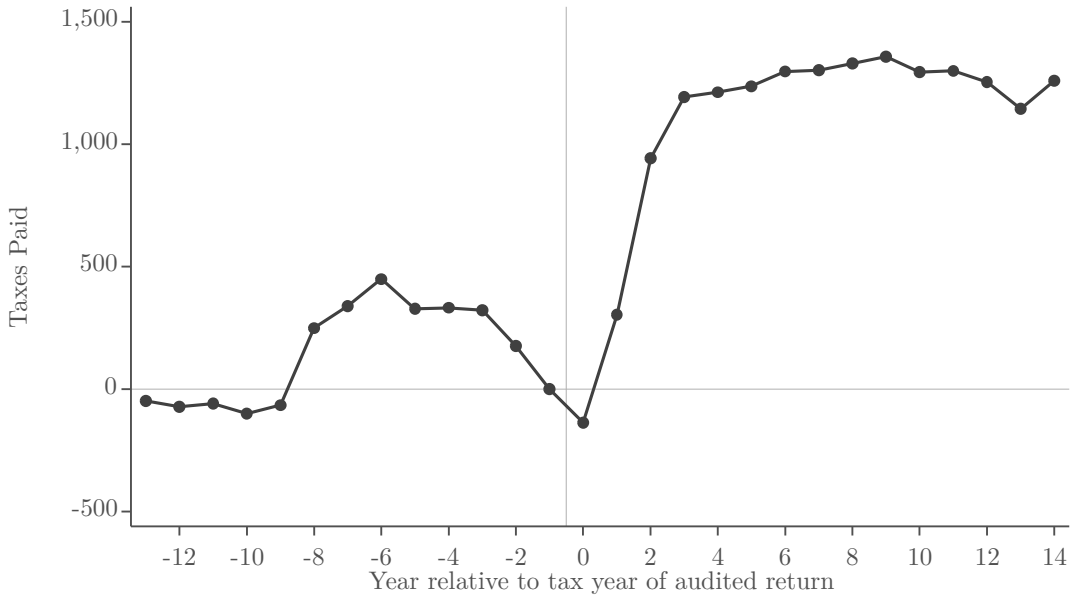
Notes: This figure presents estimates of the change in taxes paid each year post-audit for the full population individuals selected for random audit via the National Research Program (NRP), by groups of NRP study years. The control group is a matched sample of individuals not selected for random audit.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

APPENDIX FIGURE A11: Within-Taxpayer Impact of Audits on Future Tax Payments, by Presence of Income



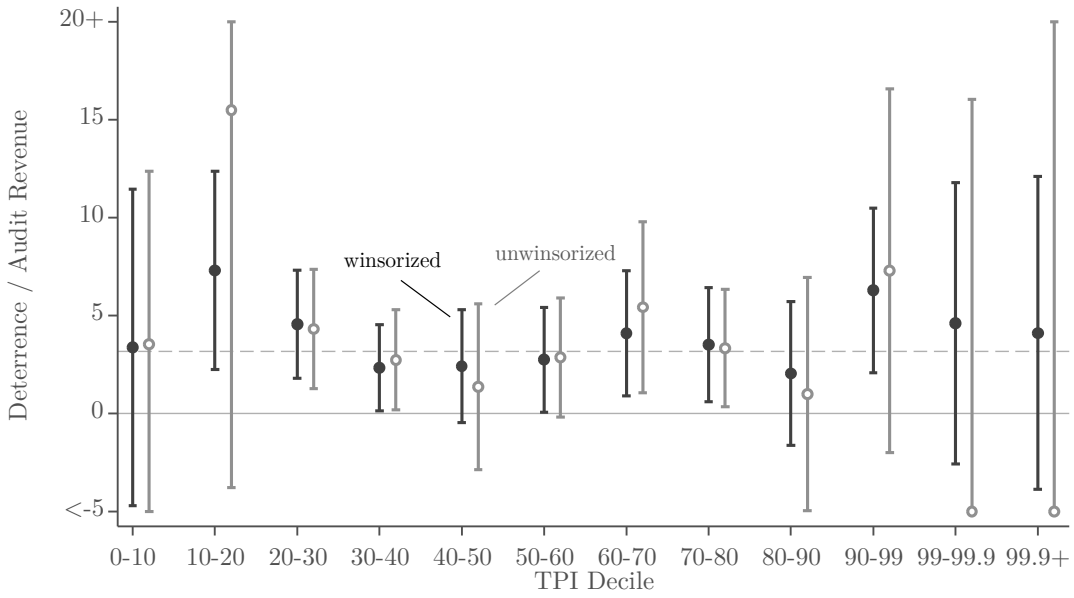
Notes: This figure presents estimates of the change in taxes paid each year post-audit for the full population individuals selected for random audit via the National Research Program (NRP) separately for individuals with and without business income (as measured by income on Schedule C, E and F). The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

APPENDIX FIGURE A12: Within-Taxpayer Impact of In-Person Audits on Future Tax Payments



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

APPENDIX FIGURE A13: Deterrence Effect over Initial Audit Revenue (Winsorized and Unwinsorized), by Income



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

Online Appendix

A Data Appendix

This appendix discusses the data used in our analysis and the methods used for computing the costs and revenues associated with each audit. As noted in the main text, we use two types of internal IRS data. In what follows, we detail how we handle each type of data to estimate revenues and costs.

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

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

We begin with all rows in the enforcement database where the taxpayer is an individual. [71] This includes both operational exams (in-person and correspondence) as well as random NRP audits.

Identifying a single audit. Processing these data first involved determining which audit enforcement activities correspond to the same audit. To account for this, we pool together all examinations that start within 90 days of when a previous exam ended. For example, if we observe that a taxpayers 2016 1040 return is examined, and within 90 days of the end of activity on that exam we observe an exam on the individuals 2015 or 2014 returns, we include these earlier returns in both the revenue and costs associated with the audit that initiated with the 2016 return. By this definition, 81.9% audits are associated with one tax return; 12.5% are associated with 2 tax returns, and 5.6% are associated with 3 or more tax returns.

Date variables. As noted above, we estimate the date the audit started based on the transactions associated with a given audit. In addition, we use five variables included in the raw enforcement data to estimate four year variables that help us track each audit over time, appropriate adjust for inflation, etc.. We define the audit start year as the year in which an audit began. We define the primary year to be the tax year of the return that triggered the enforcement process by our definition above.[72] For example, suppose in 2017 the IRS audited a 2014 tax return (which would

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

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

have been filed around April 2015) . The audit start year would be 2017 and the primary year would be 2014. The labor year is the fiscal year for which the hours are recorded.[73] We use this year to assign the pay rate that is matched to the data. Finally, the revenue year is the fiscal year in which payments are recorded in the transaction data.

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

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

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

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

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

Estimating direct labor costs (exam and appeals stages). Each row of data notes the hours spent by the IRS employee as well as their GS pay grade. We estimate direct labor costs for the exam and appeals stages by multiplying the hours in a given transaction by the matched hourly GS pay rate (described below). We note that hours and costs associated with legal counsel are listed

[73] When available, labor year is equal to assessment fiscal year. If assessment fiscal year is missing, we use the year after the exam start year. When both assessment fiscal year and exam start year are populated, assessment fiscal year is one greater than exam start year in 68.8% of rows (and equal in 22.6% of rows). Exam start year may differ from assessment fiscal year in situations where the return was examined after the audit was initiated. This leaves very few cases with a missing value for labor year. In those few cases, there are zero associated hours.

as a separate category of transactions. Because these expenses are generally incurred during the appeals phase, we add the direct costs from any "counsel" hours to appeals to get a total direct labor cost estimate for appeals. We use the GS pay scale hourly rate for Step 5 for these estimates. We then use the year in which the activities and expenses were incurred (i.e. "labor year" defined above) to adjust these costs for inflation.

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

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

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

We create a mapping of zip code to FIPS codes, and then FIPS codes to the localities provided in the historical GS pay scale data. After creating this mapping, we merge location-specific hourly pay rates to the transactions the enforcement data by year, location, and GS grade. If we had a matched zip code in the enforcement data but no associated GS pay rate for that zip code, we applied the Rest of US (RUS) rate for that GS grade and year. [74]

We then took an average of the difference between the assigned payrates and the base GS

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

payrates to determine our average location adjustment.[75] We use this average location adjustment for cases where there was no matched zip code.

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

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

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

SB/SE is the operating division responsible for audits of individual tax returns and therefore the operating division relevant for this project.[76] There are three organizations within SB/SE: collection, examination, and operations support (Internal Revenue Manual 1.4.40.2). The examination organization is responsible for both field examination and correspondence examination (Internal Revenue Manual 1.1.16.5).

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

In-person exams. To calculate the cost multipliers for in-person exams, we use internal IRS accounting data for the Field Exam organization from 2011-2020. These data include line-item level cost information for all Field Exam operations as well as total costs for Field Exams conducted by SB/SE. We create our in-person exam multipliers by calculating the appropriate ratios of these

[75] This average was weighted by number of hours.

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

line-items to total costs using the information available for all Field Exam costs and applying those ratios to the total Field Exam costs for SB/SE. We describe this process in detail below.

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

After determining the total costs associated with audits for the SB/SE operating division, we split these costs into three broad buckets: (1) non-direct labor-related costs, (2) organization-wide costs, and (3) other overhead. We use the line-item level data for all field exams to determine which costs go in which category. Non-direct labor-related costs include time spent by auditors on things beyond conducting audits (e.g. training), fringe benefits, management labor costs, training costs, and other primary labor-related costs. Organization-wide costs include things like space/rent and information technology costs incurred by the auditors. Finally, overhead costs are allocations of accounting costs from the central IRS management that overseas the audit programs.[77] We only observe these line-item costs totaled across all of field exam, including LB&I and TE/GE field exam in addition to SB/SE field exam. We again assume that these line-items are a constant fraction of the total budget for each operating division within field exam. In Table A3, we calculate what percentage of total relevant costs comes from each of these three categories and apply these

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

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

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

percentages to the total relevant costs estimated for SB/SE to estimate the value of each of these categories of costs for SB/SE. This gives us the costs for SB/SE that will be used in the numerator of our cost multipliers for field exam.

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

To calculate total direct labor costs for SB/SE, we proceed analogously to our measure of direct labor costs for individual audits measured above, but we now pull all transactions from the administrative enforcement data associated with in-person exams and with the SB/SE operating division (i.e. including businesses). We calculate total direct labor costs as described above for our audit-level dataset. For transactions related to businesses instead of individuals, we apply an average location adjustment rather than a zip-code specific hourly pay rate.[78] The total direct labor costs from SB/SE in-person exams is given in the last column of Table A1.

We calculate three multipliers for in-person exams using the three broad cost categories described above: (1) non-direct labor-related costs per-dollar of direct labor costs,[79] (2) organization-wide costs per dollar of direct labor costs, and (3) other overhead costs per dollar of direct labor costs (see Table A4). We sum these three values for an overall multiplier.

Because we focus on 2010-2014 tax year audits (which are filed in early 2011-2015), we use the average of the 2011-2015 fiscal year values of these multipliers. The overall value is 4.3854. That is, for every dollar of direct labor costs spent on in-person exams, we include $4.39 of non-direct labor costs, organization-wide costs, and other overhead costs. Non-direct labor costs account for nearly 50% of the additional costs.

Correspondence exams. To calculate the costs for correspondence exams, we use internal IRS accounting data for the Correspondence Exam organization from 2011-2020. We obtain data on costs for all correspondence exams within SB/SE, not just those of individuals. We create our

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

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

Appendix Table A2: Categories for Field Exam Costs

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

Appendix Table A3: Cost Breakdown for Field Exam

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

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

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

correspondence exam multipliers by calculating the appropriate ratios with the detailed data for all correspondence exams and applying those ratios to the total costs for SB/SE. We describe this process in detail below.

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

Table A8 also shows how we apply the relevant ratios to the total costs for SB/SE to estimate the division of costs between our three categories.[80]

We calculate three multipliers for correspondence exams using the three broad cost categories de-

[80]Before we break these relevant total costs down into our three cost subcategories, we need to subtract the component of the Correspondence Exam efforts that were conducted in support of field exam audits. These costs were included in the costs of Field Exam, because they are costs incurred as a result of the Field Exam program, and therefore should be excluded from the total costs for Correspondence Exam. We assume these costs in support of field exam efforts are incurred proportionally across the line items and subtract them from our cost categories accordingly.

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

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

Appendix Table A6: Cost Breakdown for Correspondence Exam

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

scribed above: (1) non-direct labor-related costs per-dollar of direct labor costs,[81] (2) organization-wide costs per dollar of direct labor costs, and (3) other overhead costs per dollar of direct labor costs (see Table A7). We sum these three values for an overall multiplier.

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

Appeals. Both field and correspondence exams recommendations can be appealed, which sends the case to the IRS independent office of appeals. To calculate the cost of the appeals stage of audits, we take the line-item costs for Appeals from the Field Exam data and divide by direct labor costs for appeals and counsel from the administrative enforcement data.[82] We do not have these cost numbers split out by operating division for either the internal accounting data or the administrative enforcement data. Consequently, we use the internal accounting data for all of Field Exam and all transactions associated with appeals and counsel from the administrative enforcement

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

[82] There is not a corresponding Appeals line-item from the Correspondence exam data.

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

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

data to estimate the direct labor costs. We use the average from 2011-2014, which is 0.5698.[83]

Collections. Not everyone pays their assessed tax revenue. 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 notification, the case will be handled by the Automated Collection System (ACS) or by a local field office (Field Collections). If the case is sent to ACS, ACS personnel will try to contact the taxpayer by correspondence and by phone to work with the taxpayer to find a payment solution. If ACS is unsuccessful at resolving the unpaid balance, the case is sent directly to a local IRS field office in which a Revenue Officer will work with directly with the taxpayer to attempt to resolve the unpaid tax liability.

To calculate the costs of these functions, we use data from 2016 from the internal IRS accounting data for Notices, ACS, Field Collections. The available data span 2016-2020; we do not have internal IRS accounting data for Collections prior to 2016. We use 2016 because it is the closest to the tax years we focus on in our analysis (2010-2014). Given that much of the costs of the various collections functions will be incurred well beyond the tax year of the audit, and data from 2016 is arguably the more relevant for measuring the cost of audits of tax years 2010-2014.

Per-collections notice. The cost multiplier we use for Notices is a cost-per-notice multiplier. We apply the per-notice average cost estimates from 2016 because this is the closest year to the 2011-2015 window we use for our other cost multipliers. These values are highlighted in Table A8. The average cost is $10.97 for sending a first notice, $9.13 for a second notice, and $17.70 for a "final" notice.

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

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

Appendix Table A8: Average Cost per Notice ($)

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

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

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

ACS. We estimate the cost multiplier from an audit going through ACS as a "cost-per-dollar raised." As with notices, we only have data from the 2016-2020 time period. We use the 2016 value of cost-per-dollar raised ($0.0513), as shown in Table A10.

Field collections. We apply our estimated cost multiplier from in-person exams to estimate the

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

Year	Cost per Revenue Dollar
2016	0.0513
2017	0.0551
2018	0.0567
2019	0.0529
2020	0.0518

total costs associated with field collections. This is because we are not able to separately calculate direct labor costs that result from time in field collections (as opposed to direct labor costs associated with the entire collections process). This assumes that, on average, the ratio of non-direct labor costs to direct labor costs is the same between in-person exams and in-person efforts to collect unpaid tax liability. This assumption is simpler and, in our view, more plausible than the various assumptions needed for our best attempt to calculate a separate, field collections-specific cost multiplier (which yielded a cost multiplier of 2.19).

B Relation to Existing Estimates

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

Holtzblatt and McGuire (2020) (referred to as HM2020 hereafter) estimate the revenues and costs associated with IRS operational audits in the US. They estimate that, for in-person audits, revenue collected divided by costs were 3.3 and 2.8 in 2010 and 2017, respectively. It is not possible

to conduct an exact apples to apples comparison because the samples differ across the two papers. For example, HM2020 include not only individual audits, but also audits of corporations. They examine audits from tax years 2010 and 2017, as long as they were completed before March 31st, 2012 and 2019. They exclude returns with EITC income and also remove any "outlier" returns, defined as those in the top 0.5% of taxes collected.

All of that said, it is worth highlighting one key difference between our two papers that leads to higher average returns in HM2020 as compared to our work. (We estimate the average return to spending across all audit is 2.2:1.) The cost estimates in HM2020 only include the direct costs of hours spent on enforcement activity, rather than additional labor and non-labor costs. In particular, HM2020 calculate the cost of direct enforcement activity using hours and wage information in the ERIS database. They then incorporate an estimate of additional employee benefit costs on top of those hourly wage costs. As they note, this does not include labor costs for management and support staff. It appears to omit the labor costs of enforcement personnel in non-audit hours, and it does not include non-labor costs such as building and IT costs. In our cost estimates, which are based on internal IRS accounting information, we find that non-labor costs contribute a substantial amount to total costs. We estimate that, on average, total costs are 4.39 times larger than the direct labor cost of enforcement activities, and we estimate that more than half of that 4.39 figure is the result of non-labor costs.[84] The inclusion of these costs in our average cost estimates can help to explain the divergence between our estimates and HM2020.[85]

Sarin and Summers (2019) discuss the returns to auditing very high income taxpayers and argue that auditing taxpayers with more than $5 million in earnings can produce a return of 18:1. In particular, they conduct a back-of-the-envelope calculation, drawing upon hourly audit adjustment estimates from George (2019) and dividing by estimates of average auditor costs. While the broad trajectory of our results are consistent with their findings, the approach used in our paper differs from the approach in theirs. First, George's estimates quantify recommended audit adjustments rather than audit revenue collected. A meaningful portion of assessed tax obligations are never collected, and so these adjustments often far exceed audit revenue. This means that the observed return on

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

[85]It is also worth noting that our marginal cost estimates in Section 4 fall closer to the estimates in HM2020 because, while they still include the labor costs associated with support staff and non-auditing hours, they don't include the portion of non-labor costs we estimate to be fixed.

audit expenditures should fall below the ratio of audit assessments to auditor costs. Second, if the calculation in George (2019) is focused on the hours necessary to produce audit adjustments, these calculations may omit the cost of auditor hours once an initial assessment is made. While there is no formal confirmation of this hypothesis in George (2019), we find the average hours associated with an audit of the top 0.1% are meaningfully in excess of the average hours reported in that work.[86] The inclusion of hours past the initial assessment stage could explain this discrepancy. Finally, our analysis of very high income taxpayers focuses on those in the top 0.1%. This threshold falls below the $5 million threshold in Sarin and Summers (2019) and so we should expect a lower return to audits. Those differences help to explain the discrepancy between 18:1 figure in Sarin and Summers (2019) and the 6:1 average return we find for the top 0.1% when deterrence effects are not included.[87]

C MVPF of Tax Evasion

In this Appendix, we provide a class of structural models that motivate our MVPF formula in equation (1). Our modeling approach builds on the large literature on tax evasion (e.g. Allingham et al. (1972); Keen and Slemrod (2017)), but extends to a dynamic context that allows for audits today to change tax payments and evasion behavior in the future. In order to incorporate dynamics while still keeping the model relatively tractable, we economize on other features of the model. For example, we assume quasilinear utility, we do not allow the probability of audits to depend on past behavior, and we do not allow for strategic interactions between evasion levels and probability of audits. We show that, with these assumptions, we can derive our exact formula for the MVPF in equation (1). We stress that each of the assumptions we make could be relaxed (and should be relaxed in future work). Ultimately, the MVPF of audits will continue to be the ratio of the WTP to avoid the audit relative to the net revenue collected by the audit.

C.1 Setup

Individuals have utility functions over consumption, c_t, and earnings, y_t, in each period indexed by t. Earnings are taxed at $T(y_t)$ so that in the absence of any evasion consumption would be equal to $y_t - T(y_t)$. However, individuals have the opportunity to evade e_t dollars of their tax liability.

[86] Our audit definitions are not identical so we should not expect the exact same number of ours, but the total revenue estimates in George (2019) are relatively similar to the revenue estimates in our work.

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

This increases consumption by e_t in the event they are not audited. We let a_t denote an indicator for being audited in period t, and $\alpha_t = (a_1, ..., a_{t-1})$ to denote the individuals' audit history up through period t. We assume that, when audited, individuals must pay back their evaded amount, e_t, plus a penalty $\phi^{\alpha_t}(e_t)$ that depends not only on the evaded amount but also on the individual's audit history, α_t.

Utility in each period is given by $u(c_t, y_t) = c_t - \psi(y_t) - \mathbb{I}\{a_t = 1\} B$, where $\psi(\circ)$ measures the disutility of earning income and B measures the disutility or taxpayer burden of being audited. As noted above, we assume for simplicity that evasion has no psychic cost. This means the expected PDV of utility is given by

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

where the expectation is taken with respect to the probability that $a_t = 1$ in each period, which we denote $p_t = \Pr\{a_t = 1\}$. We assume for simplicity this probability is exogenous to evasion choices and income, and that the probability of future audit does not depend on past audits. These assumptions are easily relaxed.

In each period, the budget constraint is given by

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

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

Individuals make choices of earnings and evasion. The additive separability in the model implies that the choice of earnings y_t is independent of audits and/or evasion: The optimal choice of income, y_t^*, satisfies $\psi'(y_t^*) = 1 - T'(y_t^*)$ in each period. After plugging the budget constraints into the objective function, we see that the choice of evasion in each period solves:

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

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

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

Note that this equation shows how audits can impact future evasion behavior: if being audited increases the marginal penalty from future evasion (e.g. because plausible deniability about tax rules no longer exists), individuals may choose to reduce their evasion behavior in the future.

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

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

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

C.2 WTP for Expanded Audits

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

$$V = p_1\left[y_1^*-T\left(y_1^*\right)-\psi\left(y_1^*\right)-\phi^{\alpha_1}\left(e_{1,\alpha_1}^*\right)-B+\beta V_1\right]+(1-p_1)\left[y_1^*-T\left(y_1^*\right)-\psi\left(y_1^*\right)+e_{1,\alpha_1}^*+\beta V_0\right]$$

where $\phi^{\alpha_1}(\circ)$ is the penalty in the first period (before any audits have occurred) and e_{1,α_1}^* is the choice of evasion in the first period. The first term in brackets is the utility if audited and the second term is the utility if not audited. The term V_1 is the PDV of future utility in subsequent periods if $a_1=1$ and V_0 is the PDV of future utility in subsequent periods if $a_1=0$. The envelope theorem implies that the impact of increasing p_1 affects utility through both the first period utility and the impact on future utility:

$$\begin{aligned} -\frac{dV}{dp_1} &= e_1^{a_t}+\phi^{\alpha_1}\left(e_{1,\alpha_1}^*\right)+B+(V_0-V_1) \\ &= R^{mech}+B+(V_0-V_1) \end{aligned}$$

The first period utility impact of the additional audits is given by the level of evasion, plus the penalty, and the hassle cost. To calculate the impact of the audit on future periods, note that we

can write $V_0 - V_1$ as the present discounted future revenue collected by the government from reduced evasion:

$$\begin{aligned} V_0 - V_1 &= \sum_{t=2}^{\infty} \beta^{t-1} \left(E\left[e^*_{t,\alpha_t} - p_t \left(e^*_{t,\alpha_t} + \phi^{\alpha_t} \left(e^*_{t,\alpha_t} \right) \right) | a_1 = 0 \right] - E\left[e^*_{t,\alpha_t} - p_t \left(e^*_{t,\alpha_t} + \phi^{\alpha_t} \left(e^*_{t,\alpha_t} \right) \right) | a_1 = 1 \right] \right) \\ &\equiv R^{future} \end{aligned}$$

so that $V_0 - V_1$ is the causal effect of the audit in period 1 on the PDV of evaded tax revenue in the future (inclusive of revenue from audits). Combining, the willingness to pay to avoid an expansion of audits in period 1 is given by:

$$-\frac{dV}{dp_1} = R^{mech} + R^{future} + B \tag{3}$$

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

The formula assumes quasilinear utility so that changes in evasion don't affect labor earnings, y_t. They also assume that the probability p_t is independent of the choice of income that an individual has. This latter assumption can easily be relaxed by assuming that income choices, y_t, are affected by the probability of the audit. The envelope theorem implies that these will not enter the willingness to pay to avoid the expanded audits; however, they could generate an additional revenue to the government from the audit. We turn to government revenue now.

C.3 Government Revenue and MVPF of Expanded Audits

Let G denote the PDV of government revenue:

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

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

It is straightforward to see that the impact of expanding audits in period 1 is given by the sum of the revenue collected in the first period, R^{mech}, and the revenue collected in future periods, R^{future}, minus the marginal cost of the audits in period 1:

$$\frac{dG}{dp_1} = R^{mech} + R^{future} - C \tag{4}$$

One term that is not in this expression that one might think should be included is a term corresponding to the deterrence effect of the increased probability, p_t, on evasion in the first period, e^*_{1,α_1}.

Note that the net impact on the government budget from changes in e^*_{1,α_t} is given by:

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

Because individuals are maximizing their expected income from evasion, this means they are also minimizing the government revenue from evasion. Recall from the individuals' optimization problem above that $\frac{\partial\phi^{\alpha_t}}{\partial e}\left(e^*_{t,\alpha_t}\right)=\frac{1-p_t}{p_t}$. This means that small changes in evasion behavior in response to changes in the audit probability do not affect government revenue. This could be generalized if there were other motivations that individuals had that determined their evasion. For example, if individuals had a psychic cost of evasion, then reductions in evasion would have a first-order impact on government revenue, and thus $\frac{dG}{dp_1}$ would include this "general deterrence" term arising from increased tax revenue paid by those in period 1 who were not audited. This would in turn further push the MVPF towards 1, the implications of which we discuss below.

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

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

which is precisely our formula in equation (1).

C.4 Income Heterogeneity and the MVPF of Tax Changes

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

We can also compute the MVPF using the formula above now conditioning on a particular point of the income distribution. In doing so, it is important to note that changes in the audit probabilities

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

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

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

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

We can now compare the MVPF of tax audits to the MVPF of changes in the income tax schedule across the income distribution. In the environment above, Hendren (2020) shows that the MVPF of a tax change targeted to a particular region of the income distribution is given by

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

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

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

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